Original Article

Single-cell DNA methylome analysis of circulating tumor cells

Hengyu Chen[1*], Zhe Su[1*], Ruoyan Li[1], Ning Zhang[2,3], Hua Guo[2], Fan Bai[1,3]

[1]Biomedical Pioneering Innovation Center (BIOPIC), School of Life Sciences, Peking University, Beijing 100871, China; [2]Laboratory of Cancer Cell Biology, Tianjin Medical University Cancer Institute and Hospital, Tianjin 300060, China; [3]Translational Cancer Research Center, First Hospital, Peking University, Beijing 100871, China

*These authors contributed equally to this work.

Correspondence to: Hua Guo. Laboratory of Cancer Cell Biology, Tianjin Medical University Cancer Institute and Hospital, Tianjin 300060, China. Email: guohua@tjmuch.com; Fan Bai. Biomedical Pioneering Innovation Center (BIOPIC), School of Life Sciences, Peking University, Beijing 100871, China. Email: fbai@pku.edu.cn.

Abstract

Objective: Previous investigations of circulating tumor cells (CTCs) have mainly focused on their genomic or transcriptomic features, leaving their epigenetic landscape relatively uncharacterized. Here, we investigated the genome-wide DNA methylome of CTCs with a view to understanding the epigenetic regulatory mechanisms underlying cancer metastasis.

Methods: We evaluated single-cell DNA methylome and copy number alteration (CNA) in 196 single cells, including 107 CTCs collected from 17 cancer patients covering six different cancer types. Our single-cell bisulfite sequencing (scBS-seq) covered on average 11.78% of all CpG dinucleotides and accurately deduced the CNA patterns at 500 kb resolution.

Results: We report distinct subclonal structures and different evolutionary histories of CTCs inferred from CNA and DNA methylation profiles. Furthermore, we demonstrate potential tumor origin classification based on the tissue-specific DNA methylation profiles of CTCs.

Conclusions: Our work provides a comprehensive survey of genome-wide DNA methylome in single CTCs and reveals 5-methylcytosine (5-mC) heterogeneity in CTCs, addressing the potential epigenetic regulatory mechanisms underlying cancer metastasis and facilitating the future clinical application of CTCs.

Keywords: Circulating tumor cells; methylome; copy number alteration

Submitted Mar 19, 2021. Accepted for publication May 10, 2021.
doi: 10.21147/j.issn.1000-9604.2021.03.10
View this article at: https://doi.org/10.21147/j.issn.1000-9604.2021.03.10

Introduction

Circulating tumor cells (CTCs) are cancer cells that have entered the peripheral blood from the primary tumor and possess the potential for seeding metastases (1). The molecular characterization of CTCs can help to elucidate the dynamic process of cancer metastasis and may serve as a non-invasive biomarker for the early detection, real-time monitoring, and prognosis prediction of cancer. The rapid development of single-cell sequencing has paved the way for the molecular characterization of CTCs; however, despite a myriad of attempts on the molecular analyses of CTCs (1-5), their epigenetic characterization remains largely unexplored.

Over the past decade, there have been a few research attempts to depict the methylation state of CTCs. Chimonidou and colleagues contributed a series of pioneering studies using methylation-specific PCR (MSP), which added a new dimension to the molecular characterization of CTCs. These studies: 1) demonstrated that promoter methylation of tumor suppressor and metastasis suppressor genes is a hallmark of CTCs and may be heterogeneous (6); 2) indicated a direct connection between CTCs and cell-free circulating DNA (cfDNA) (7);

and 3) examined the relationship between the epigenetic silencing of *BRMS1* and clinical outcome in breast cancer (8). An exploratory study provided DNA methylation profiling of CTCs using a methylation microarray covering 27,000 CpGs, which suggested that CTCs epigenetically resemble the primary tumor tissue in castration-resistant prostate cancer (CRPC) and that DNA methylation is likely to be crucial in CTC survival and regulation of metastatic potential (9). In addition, an interesting approach using a syngeneic murine hepatocellular carcinoma (HCC) model, high resolution melt (HRM) analysis, and pyrosequencing was employed by Ogunwobi and colleagues. This study demonstrated that increased hepatocyte growth factor (HGF) expression, possibly in conjunction with the upregulation of c-Met, induces epithelial-mesenchymal transition (EMT) of CTCs during dissemination in HCC (10). Taken together, these pioneering investigations represent landmarks in the interrogation of epigenetic features in CTCs but unfortunately failed in assessment at the single-cell level due to early technical limitations. Pixberg *et al*. (11) provided the first DNA methylation profiling in single CTCs by implementing a creative and effective protocol based on agarose-embedded bisulfite treatment in combination with multiplex PCR (multiplexed-scAEBS); nonetheless, many features associated with the epigenetic regulation of cancer metastasis remained unexplored considering its limited throughput.

Recent advances have made it possible to detect DNA methylation using ultra-low DNA input. Here, to gain a better understanding of the metastatic cascade as well as its clinical relevance, we applied the single-cell bisulfite sequencing (scBS-seq) method (12), which enabled a genome-wide assessment of the methylation levels in individual CTCs. Our study provides new insights into the epigenetic regulation of cancer metastasis.

Materials and methods

Patient recruitment and clinical information

We enrolled 17 patients in the present study. Patient information is summarized in *Supplementary Table S1*. This study was approved by the Institutional Ethics Committee at Tianjin Cancer Hospital. All participants provided written informed consent.

CTC capture and isolation

CTCs in 7.5 mL whole blood from each patient were first captured with CELLSEARCH® CTC Kit (Epithelial) (Veridex, LLC) using magnetic beads conjugated to EpCAM antibodies. Captured CTCs were stained with 4',6-Diamidino-2-Phenylindole (DAPI) and anti-cytokeratin (CK)-phycoerythrin and anti-CD45-allophycocyanin antibodies to distinguish cancer cells from carried-over white blood cells (WBCs). Subsequently, individual CTCs (DAPI+, anti-CK+, anti-CD45−) and WBCs (DAPI+, anti-CK−, anti-CD45+) were isolated by mouth-manipulation under a fluorescence microscopy. An additional FITC channel was added to confirm that the fluorescence signal in the anti-CK-phycoerythrin channel was not due to cross-color noise. In addition to the immunofluorescence criteria described above, a morphometric and copy number alteration (CNA) analysis of candidate cells was performed for validation.

Single-nucleus suspensions were prepared from frozen tumor tissues of Patient gastric cancer (GC)1 with NST-DAPI solution using a previously described method (13). All cells and nuclei were imaged under an inverted fluorescence optical microscope (Nikon, ECLIPSE Ti).

scBS-seq library construction

The scBS-seq library was prepared according to a previously published protocol, with minor modifications (12). The isolated cell or nucleus underwent lysis and release of double-stranded DNA, which was denatured into small fragments of single-stranded DNA during bisulfite conversion and purification (all volumes were halved) (ZYMO RESEARCH). Subsequently, the converted DNA was subjected to five rounds of random priming and extension using oligo1 (CTACACGACGCTCTTCCG ATCTNNNNNN) and Klenow polymerase (3' to 5' exo$^-$, Enzymatics). Excess random primers were digested by the addition of Exonuclease I. The product was purified and tagged with oligo2 (AGACGTGTGCTCTTCCGATC TNNNNNN, which was specially designed as an Illumina sequencing adaptor). Libraries were amplified after 13−14 cycles of PCR with the Illumina universal primer and Illumina indexed primer. Purification and quality control were the last steps of library preparation. The final quality-ensured libraries were pooled and sequenced on the Illumina HiSeq XTen platform to yield 150 bp paired-end reads (Novogene).

Sequencing read quality control and alignment

All sequencing data in this study were generated on the

Illumina HiSeq XTen platform. We first performed quality control on raw paired-end reads to remove the first 6 bp random priming segment, adaptor contamination, and poor-quality reads using Trim Galore (V0.3.3; https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/), with the major parameter settings as follows: -clip_r1 6 -clip_r2 6. The paired-end reads after the trimming process were mapped to the hg19 Refseq reference genome using Bismark (V0.16.3; parameters: --phred33-quals, --bowtie1, --non_directional, --unmapped), resulting in 22.65% of mapping efficiency (range, 6.30%−38.34%). To recover additional sequencing reads, remaining unmapped reads were realigned to the same reference genome in single-end alignment mode (Bismark parameters: --phred33-quals, --bowtie1, --non_directional). "Samtools merge" was used to merge the BAM files from the above two rounds of alignment. Duplicated reads were removed from each sample using "Samtools rmdup". Methylation calls were extracted using Bismark_methylation_extractor.

DNA methylation level estimation

For each CpG site in the reference genome sequence, the DNA methylation level was determined by counting the methylated and unmethylated reads. A DNA methylation level ≥0.9 was considered methylated, whereas that ≤0.1 was considered unmethylated. The read coverage threshold used to call the DNA methylation level for any cytosine was 1× for single cells or nuclei and 3× for pseudo-bulk samples.

An in-house R script was used to estimate the methylation level of gene promoter regions. Briefly, we categorized genes into those on the positive strand of DNA and those on the negative strand. For each gene, we considered the region 1,000 bp upstream of the transcription start site (TSS) to 500 bp downstream of the TSS as the promoter region. Again, a DNA methylation level ≥0.9 was considered methylated, whereas ≤0.1 was considered unmethylated. We used this binary data to calculate the methylation level of gene promoter regions.

The gene body was defined as the region from TSS to the transcription end site (TES) of each gene. Similarly, the methylation level of the gene body region (from TSS to TES) and 5 kb upstream/downstream flanking regions was estimated using an in-house R script. Briefly, the gene body region was divided into 40 bins of equal size, whereas the 5 kb upstream/downstream region was divided into 10 bins of equal size. We then calculated the methylation level of each bin using the binary call for CpG sites as methylated or unmethylated. The average methylation levels of different genes in different bins were calculated.

Inferring CNAs based on single-cell methylome data

Single-cell methylome data were used to infer whole genome-wide CNAs in CTCs. Briefly, BAM files were collected from the alignment and duplicated reads were removed using "Samtools rmdup". Subsequently, the depth of each covered base was calculated using "Samtools depth". We segmented the whole genome into small bins of 500 kb in size, in which the total depth was calculated. Following normalization of the total depth of each bin according to sequencing data volume, we applied Lowess regression normalization to the bin counts in order to reduce the GC bias (14), and again to calculate the depth ratio and correct the bias caused by PCR amplification using the bin counts of WBCs as a reference. Segmented copy numbers were generated using the circular binary segmentation (CBS) algorithm implemented in R package "DNAcopy" (http://bioconductor.org/packages/release/bioc/html/DNAcopy.html). Finally, small adjacent segments with non-significant differences were merged into large segments using Merge Levels (15).

Phylogeny analysis using both CNA and methylation data

To construct phylogenetic trees based on CNA dataset, integer copy numbers were calculated for each single cell by multiplying the segmented depth ratio by 2 and rounding to the closest integer value. Subsequently, the integer copy number dataset was categorized into three states: normal (copy number=2), loss (copy number <2), and gain (copy number >2). We constructed phylogenetic trees using R package "phangorn" (V2.5.5). Briefly, the categorized CN data were format-transformed using the "phyDat" function. We calculated the pairwise Hamming distance across all CTCs and WBCs (WBCs were used as the root or outgroup in phylogenetic tree construction) using the "dist.hamming" function. Hierarchical clustering analysis was performed based on the R "hclust" function using the clustering method "ward. D". Similarly, we constructed phyloepigenetic trees based on the methylation dataset by first rounding methylation levels to three values, 0, 0.5, and 1, which corresponded to three methylation states, unmethylated, semi-methylated, and fully methylated, respectively. Phyloepigenetic trees were also constructed using R package "phangorn" (V2.5.5).

Differentially methylated region (DMR) analysis

All single cells from the primary tumors, CTCs, and metastases were merged as pseudo-bulk samples. Methylation calls for each pseudo-bulk sample were extracted using Bismark_methylation_extractor. We applied a tile-based method to bin consecutive genomic windows with a fixed length (1,000 bp) for comparison across samples. The average methylation level in each window was calculated for the three pseudo-bulk samples. Similarly, the window-based methylation levels were calculated among single cells including those from primary tumors, CTCs, and metastases. Subsequently, the pseudo-bulk samples were used to search for initial DMR candidates. We considered the windows with a methylation level <25% in one pseudo-bulk sample and >75% in another sample as DMR candidates. For example, a genomic region with a methylation level <25% in the primary tumor and >75% in CTCs was considered a DMR candidate with increasing methylation level. Following identification of these DMR candidates, we used the single-cell methylation level of the same genomic windows to evaluate whether these DMR candidates were statistically significant [multiple *t*-test, false discovery rate (FDR) <0.05]. Accordingly, we finally identified DMRs that were significantly different between any two states (primary, CTCs, and metastases).

Methylation analysis of CpG islands

For each cell, cytosines that were not in CpG positions were filtered out, and one position was retained as a methylated cytosine in CpG if 90% of the covered cytosines were methylated and one position was retained as an unmethylated cytosine in CpG if 90% of the covered cytosines were unmethylated. The CpG positions were set as methylated if there was at least one cytosine that was methylated, and all the other covered CpG positions were marked as unmethylated. In each CpG island, the ratio of methylated CpG positions was calculated. We merged all cells from different patients and added the missing values using R package "FactoMineR". The *t*-Distributed Stochastic Neighbor Embedding (*t*-SNE) analysis was performed using R package "tsne".

Using the imputed autosome CpG island matrix, the Pearson correlation coefficient between each two cells was calculated and transformed as the distance, then the hierarchical clustering with "ward. D2" method was performed.

Results

scBS-seq of 196 individual cells from 17 cancer patients

A total of 17 patients with six different types of cancer [breast cancer, colon cancer, GC, prostate cancer, lung adenocarcinoma (ADC), and small cell lung cancer (SCLC)] were recruited to the present study (*Supplementary Table S1*). For DNA methylome analysis, we performed scBS-seq on 196 single cells, including 107 CTCs captured by the CellSearch® platform (*Figure 1A*,*B*), 16 WBCs, and 73 cell nuclei from paired primary/metastatic tissues (*Supplementary Table S2*).

Firstly, we examined the DNA methylation pattern across gene body regions in all samples. As observed in previous studies on both development and disease (16,17), the valleys around TSS were hypomethylated at the single-cell level, with hypermethylation patterns along gene bodies in the present study (*Figure 1C*). Tumor cells have been shown to exhibit genome-wide DNA hypomethylation compared with paired normal cells in colorectal cancer (18). Similarly, tumor nuclei and CTCs showed lower methylation levels than those in paired normal nuclei and WBCs, respectively (*Figure 1D*).

CNA is a major form of genetic variation in cancer and is widely reported in human malignancies (1,3,4). During bisulfite conversion, unmethylated cytosines were converted to uracil, while methylated cytosines and other nucleobases remained unchanged; therefore, the information retrieved from methylation sequencing could be used to infer the CNA pattern of a single tumor cell. From our scBS-seq data, we can infer CNAs of individual CTCs at 500 kb resolution. Two representative examples derived from Patient SC6 with a WBC as the control are shown in *Figure 1E*.

Inter- and intra-patient epigenetic heterogeneity of CTCs in lung cancer

To evaluate the inter- and intra-patient heterogeneity of CTCs, we focused on the lung cancer patients from whom we collected the largest number of CTCs, in particular those with SCLC, which is an aggressive subtype of lung cancer associated with an extremely poor prognosis. We analyzed the methylome of individual CTCs and WBCs derived from Patients SC6 and SC7. Interestingly, these cells exhibited similar DNA methylation patterns in gene promoter regions, which could be divided into three main clusters in a patient-dependent manner, with WBCs

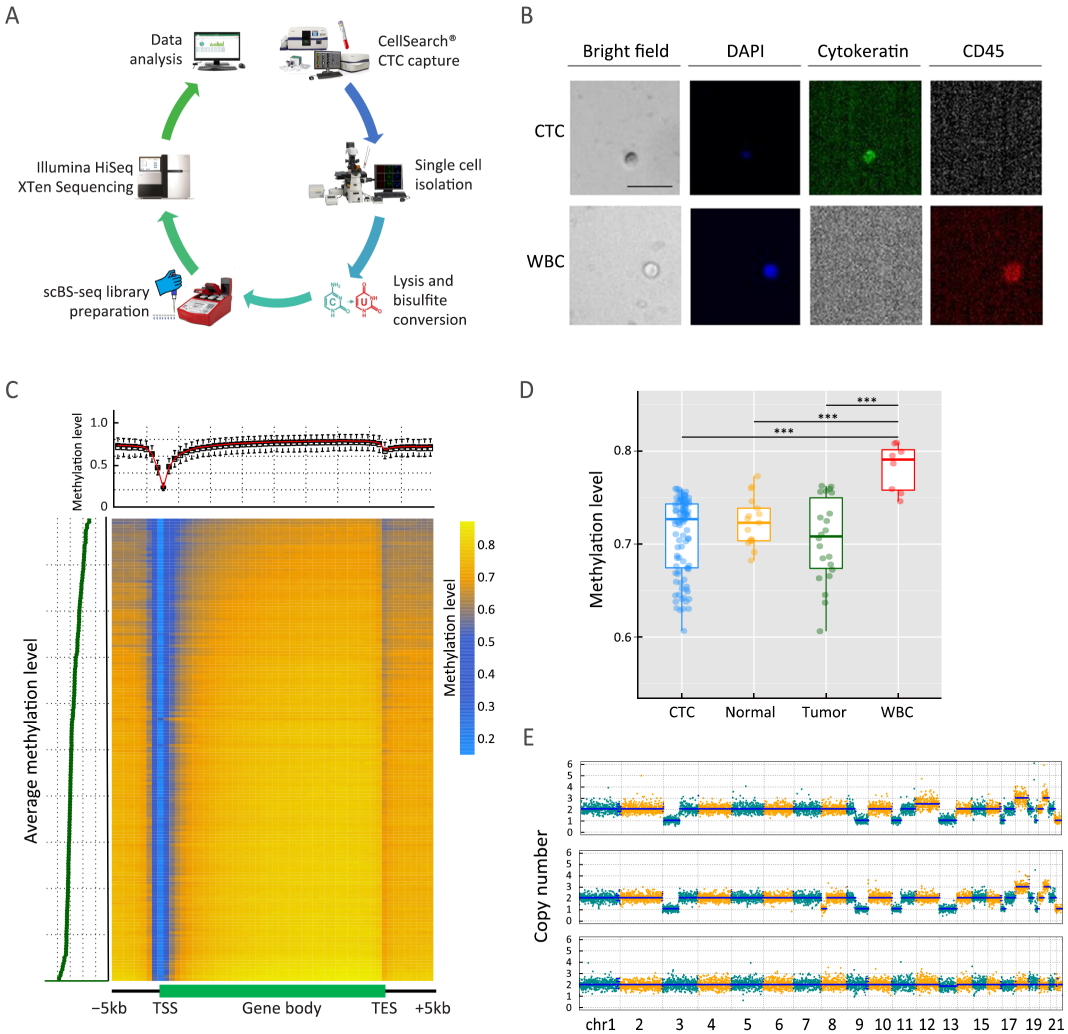

Figure 1 Single-cell genome-wide bisulfite sequencing of cancer patient CTCs and tumor tissues. (A) Workflow illustrating an overview of the experimental steps. Standard blood samples obtained from cancer patients were used for CTC enrichment via the CellSearch® system, authentication from experience, and isolation by micromanipulation; (B) Representative immunofluorescence images of a CTC (top row) and a WBC (bottom row) co-stained with DAPI, anti-cytokeratin (CK), and anti-CD45. A CTC was identified as DAPI+, CK+, and CD45−, whereas a WBC was defined as DAPI+, CK−, and CD45+. Scale bar, 20 μm; (C) DNA methylation pattern in gene body regions as determined from all samples' scBS-seq data. A total of 123 single cells and 73 nuclei were sequenced in the present study. The averaged DNA methylation level of CpG sites was calculated from all annotated RefSeq genes in gene body and their 5 kb flanking regions; (D) Box plot of the average methylation level of CTCs (blue), normal nuclei (orange), tumor nuclei (green), and WBCs (red); (E) CNA profiles from Patient SC6 (shown as an example) at 500 kb resolution. The top 3 panels show the CNA pattern of CTCs, whereas the bottom panel shows the normal copy number pattern of the WBC control. CTC, circulating tumor cell; WBC, white blood cell; DAPI, 4',6-diamidino-2-phenylindole; TSS, transcription start site; TES, transcription end site. ***, P<0.001.

separated (*Figure 2A*). The clustering of the intra-patient CTCs displayed a consecutive methylation pattern with small differences (*Supplementary Figure S1*). These results demonstrate inter- and intra-patient heterogeneity in gene promoter regions.

Subsequently, we interrogated epigenetic changes in promoter regions of tumor-associated genes. Epigenetic alterations in lung cancer can serve to inform us of the carcinogenic process and provide clinically relevant biomarkers (19). We selected 20 known SCLC-associated genes with a view to examining their methylation states in promoter regions. The promoters of *RASSF1* (41/55,

74.5%), *ERBB2* (36/55, 65.5%), and *APC* (16/55, 29.1%) were highly methylated, which emphasizes the importance of epigenetic changes mediating the loss of gene expression and function in SCLC (*Figure 2B*). In addition, massive DNA hypomethylation was observed in the promoter regions of *KRAS* and *MYC*. Interestingly, DNA methylation

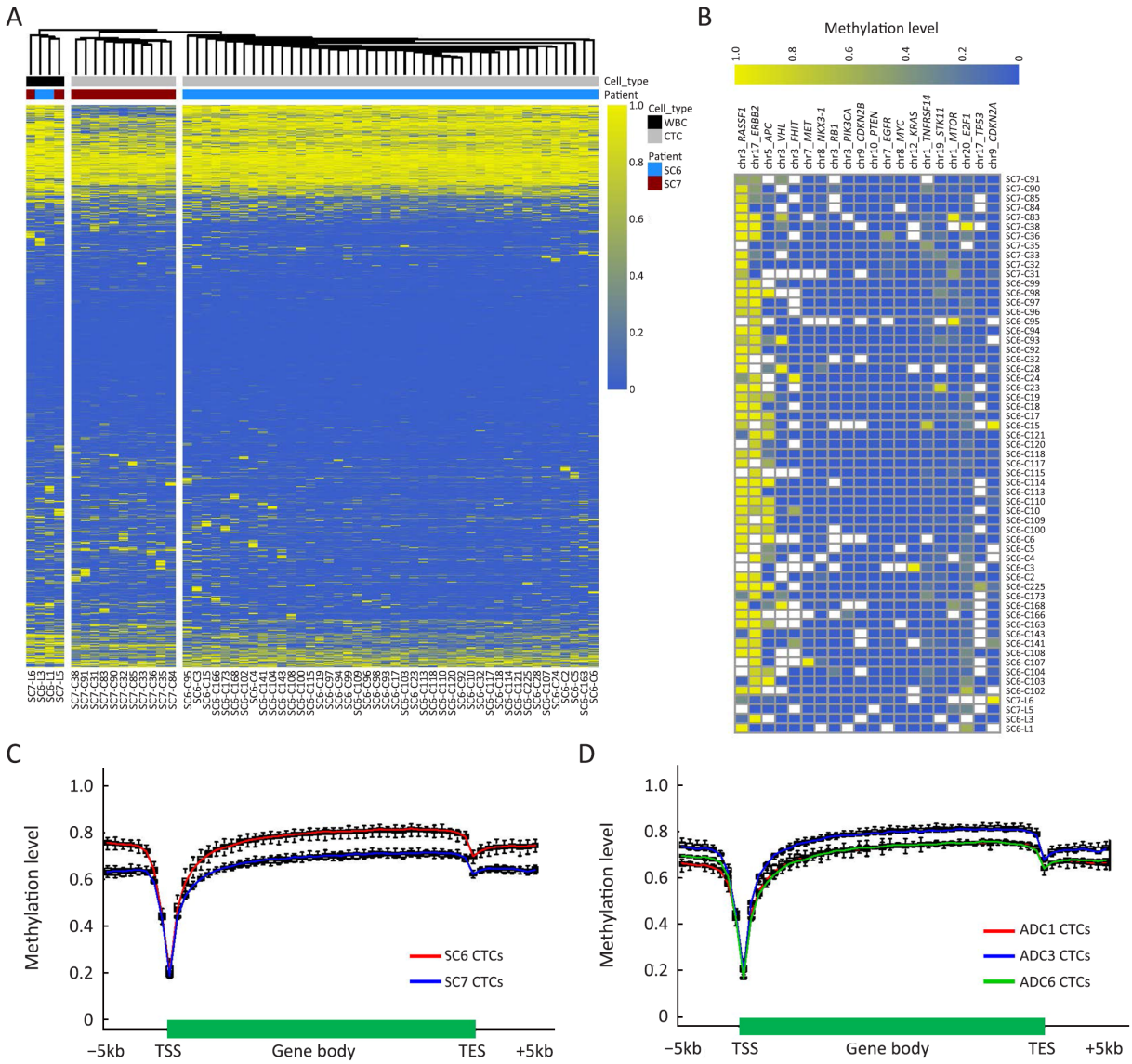

Figure 2 Inter- and intra-patient epigenetic heterogeneity of CTCs in SCLC. (A) Unsupervised hierarchical clustering of the methylome of individual CTCs (grey) and WBCs (black) derived from Patient SC6 (blue) and Patient SC7 (red). All RefSeq gene promoters were used for the analysis. The color key from blue to yellow indicates low to high methylation level; (B) DNA methylation heatmap of individual CTCs and WBCs derived from Patients SC6 and SC7. A total of 20 known SCLC-associated gene promoters were used for the analysis. The color key from blue to yellow indicates low to high methylation level, whereas white checks indicate unavailable data; (C,D) DNA methylation pattern in gene body regions of CTCs from Patients SC6 and SC7 with SCLC (C), and Patients ADC1, ADC3, and ADC6 with ADC (D). The averaged DNA methylation level of CpG sites was calculated from all annotated RefSeq genes in gene body and their 5 kb flanking regions. The line represents the mean value for CTCs from each patient: Patient SC6 (left, red), Patient SC7 (left, blue), Patient ADC1 (right, red), Patient ADC3 (right, blue), and Patient ADC6 (right, green). CTC, circulating tumor cell; SCLC, small-cell lung cancer; ADC, adenocarcinoma; TSS, transcription start site; TES, transcription end site.

levels in the promoter regions of several genes showed discrepancies among different CTCs. For instance, hypermethylation in the *APC* promoter was only observed in CTCs from Patient SC6 but not SC7; massive DNA hypermethylation in the *RASSF1* promoter was detected in most CTCs (30/44, 68.2%) from Patient SC6, while a few of CTCs (3/44, 6.8%) exhibited hypomethylation; and hypomethylation in *ERBB2* promoter was only observed in 45% of CTCs (5/11) from Patient SC7, with 27% of CTCs (3/11) being hypermethylated. These results show that DNA methylation in promoter regions of tumor-associated genes in CTCs can be highly heterogeneous between patients and even among cells from the same patient.

In addition to promoter regions, inter-patient DNA methylation heterogeneity also existed in gene body and their 5 kb flanking regions. Patient SC7 showed a lower DNA methylation level across gene body regions in their CTCs as compared with Patient SC6 (*Figure 2C*). In contrast, among three patients with lung ADC, there was an overlap of the average CpG methylation levels along the scaled gene bodies and 5 kb downstream of TES for all RefSeq genes between Patients ADC1 and ADC6, with Patient ADC3 exhibiting a higher DNA methylation level across gene body regions in their CTCs (*Figure 2D*). These results show inter-patient DNA methylation heterogeneity in the same cancer type.

Distinct evolutionary patterns inferred from CNA and methylation profiles of CTCs in SCLC

We performed CNA analysis for individual CTCs derived from Patient SC6. As expected, these CTCs exhibited a comparatively reproducible CNA pattern, with two obvious cell subpopulations (*Figure 3A*). Several known genes frequently reported in SCLC, such as *VHL*, *FHIT*, *RASSF1*, *RB1*, and *TP53*, were mostly inactivated through copy number deletion, with *CCNE1* and *KRAS* (partly) being amplified among CTCs. Intriguingly, frequent promoter hypermethylation of *RASSF1* was observed (*Figure 2B*) in one allele, reciprocally interacting with its loss from the other allele (*Figure 3A*) and converging to a double-hit effect on gene inactivation. Since RASSF1 protein has been shown to interact with the DNA repair protein XPA and induce cell cycle arrest, the double-hit effect supports the view of the convergence of genetic and epigenetic changes driving aberrant cell cycle function (20,21).

We then performed unsupervised clustering analysis for these cells based on the methylation rate in 10 kb windows across the genome; nevertheless, the global methylation pattern in these cells seemed indistinguishable along all chromosomes, rendering further subclonal classification difficult (*Supplementary Figure S1*). Taken together, cell subclonal structures differ in their genetic and epigenetic profiles.

To catalog the differences between genetic alterations and epigenetic changes, we compared the evolutionary patterns inferred from the CNA and DNA methylation profiles of CTCs from Patient SC6 (*Figure 3B*,*C*). The phylogenetic tree shows that aneuploid copy number changes may occur early in CTC evolution in punctuated bursts, indicating stable subclonal expansion (*Figure 3B*), which supports the view of several recent reports (3,13,22). In contrast, the phyloepigenetic tree exhibits a gradual acquisition of methylation variance (*Figure 3C*). Considering that DNA methylation changes are heritable stochastic features (23), the rate of epigenetic change has been estimated to be orders of magnitude higher than that of genetic change (24). Taken together, these data indicate distinct evolutionary patterns for copy number and DNA methylation among CTCs in SCLC at the single-cell level.

Tracing dynamic DNA methylation changes from primary tumors to CTCs and matched metastatic samples

To investigate the dynamic methylome changes that occur during cancer metastasis, we collected frozen primary and metastatic tumor tissues, in addition to 4 CTCs and 4 WBCs derived from Patient GC1 (*Figure 4A*). We examined the copy number profiles and observed that these CTCs, as well as tumor nuclei from paired primary and metastatic tumor tissue biopsies, exhibited a relatively uniform CNA pattern with small discrepancies in some chromosomes (*Figure 4A*), in accordance with our previous finding that CTCs showed reproducible CNA patterns (3). Moreover, several CNA regions containing known GC-associated genes were identified (*Figure 4B*). *FHIT* and *CDKN2A* were mostly inactivated through deletion among CTCs and primary and metastatic tumor nuclei, with *CDK6*, *MYC*, *GATA4*, and *ZNF217* being amplified, highlighting the importance of genetic changes in mediating cell cycle dysregulation during cancer metastasis. Copy number gain of *ERBB2* (7/15) and *EGFR* (2/15) existed in primary tumor nuclei and CTCs (*Figure 4B*). With respect to metastatic tumor nuclei, copy number gain of *KRAS* and *MDM2* and loss of *PRKN* were observed (*Figure 4B*).

Subsequently, we interrogated epigenetic changes in promoter regions of 20 known tumor-associated genes.

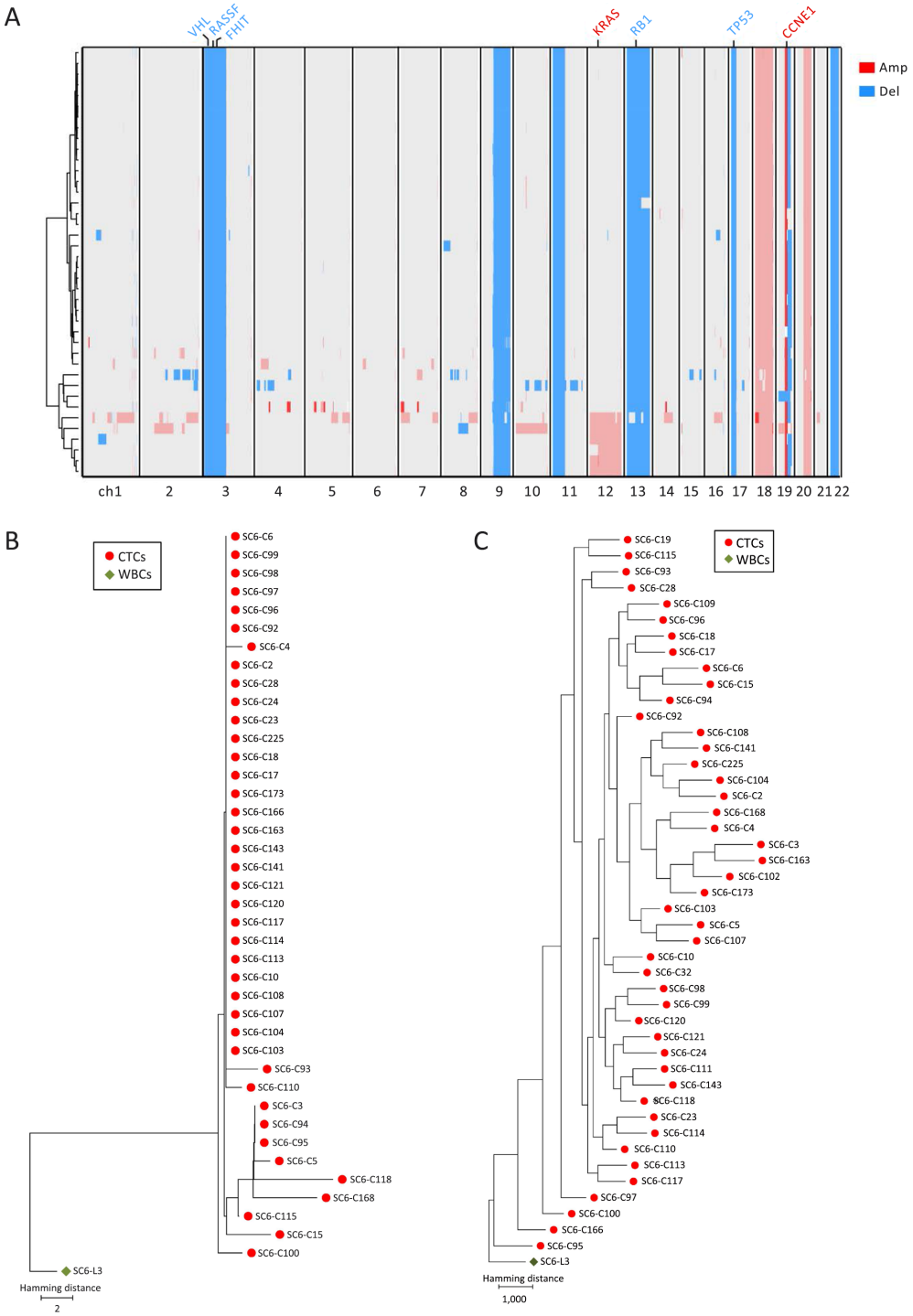

Figure 3 Distinct subclonal structures and evolutionary patterns inferred from the copy number and methylation profiles of CTCs from Patient SC6. (A) CNA profiling of individual CTCs from Patient SC6 at 500 kb resolution. Deletions (blue) and amplifications (red) for chromosomes 1–22 were shown; (B,C) Phylogenetic (B) and phyloepigenetic (C) reconstruction showing clonal relationships in Patient SC6 inferred from the chromosomal breakpoint profiles (B), and DNA methylation distance matrix (C). Cell types were color coded: green (WBCs) and red (CTCs). The Hamming distance metrics in (B) and (C) are individual and non-comparable. CTC, circulating tumor cell; CNA, copy number alteration; WBC, white blood cell.

Interestingly, *APC* inactivation through frequent promoter hypermethylation also occurred in Patient GC1 (*Supplementary Figure S2*), as previously observed in Patients SC6 and SC7 (*Figure 2B*). Moreover, infrequent promoter hypermethylation of *RASFF1* occurred in both primary and metastatic tumor nuclei. Scarce promoter hypermethylation of *CDKN2A* and *E2F1* was also shown in a primary tumor cell and a CTC, respectively (*Supplementary Figure S2*).

We also examined the DNA methylation pattern across gene body regions in Patient GC1. Genome-wide DNA hypomethylation was detected in cancer nuclei as compared with paired normal nuclei (*Figure 4C*). Moreover, the mean value of methylation level for primary tumor nuclei was relatively lower than that for metastatic tumor nuclei. Furthermore, there was an overlap among the three CTCs, with GC1-C12 exhibiting the lowest DNA methylation level (*Figure 4C*), reflecting intra-patient epigenetic heterogeneity.

To evaluate representative expression patterns during cancer progression, we identified DMRs among primary tumor nuclei, CTCs, and metastatic tumor nuclei. We calculated the methylation level and variance with a 1 kb sliding window across all the individual tumor cells and identified a total of 16,796 DMRs. Subsequently, we classified these DMRs into stage-specific categories and chose the six categories with the most significant variations, which helped to depict the dynamics of epigenetic remodeling during metastasis (*Figure 4D*).

The category C2 represents up-regulated methylation level with cancer progression. Often reported to be down-regulated in tumors (25), *LTF* was included in category C2. Several genes expressed in the ovaries, such as *C1orf35*, *DENND6A*, and *ZNF285*, included in category C3, were transiently down-regulated methylated. Furthermore, *FBP2*, *HIVEP3*, *PTPN21*, and *CEACAM5* in category C3, indicate that oncogenic transformation- and cell adhesion-associated pathways may play a role in tumor progression. Up-regulated methylation of *MYEOV*, *ERCC1*, *MUM1L1*, and *SELENBP1*, in addition to down-regulated methylation of *LHX8*, *GABRP*, *POU6F2*, *VWC2*, *TJP2*, *HYDIN*, *SNORD22*, *SLC25A30*, and several ovary-specific genes were detected in metastases, suggesting that pathways associated with DNA repair, ciliary motion, and cell adhesion may be involved in this process. Moreover, several genes associated with the immune response were included in categories C4 and C6.

To address the question of whether primary or metastatic tumors epigenetically resemble CTCs in GC, we constructed phylogenetic and phyloepigenetic trees from the DNA methylation and copy number profiles with a view to determining the tumor evolutionary history in Patient GC1 (*Figure 4E*). Intriguingly, GC1-C29 and GC1-C30 appeared closer to metastatic tumor nuclei, whereas GC1-C12 and GC1-C13 had variable positions in different trees. Specifically, GC1-C12 was relatively closely related to primary tumor nuclei in the phyloepigenetic tree, in agreement with the previous finding of its lower DNA methylation level in gene body regions (*Figure 4C*). This analysis yielded different evolutionary histories than those previously reported by bulk-sample studies that support the concept of co-dependency of aberrant DNA methylation and genetic alterations, including either CNAs or somatic mutations, producing a remarkably similar tumor evolutionary history (21,26). Importantly, the signal from bulk tumor samples is a population average, in which dominant major subclones often render rare subclones undetectable. Moreover, the occult non-tumor cells can also confound and mask signals from the tumor, causing an underestimation of heterogeneity (27). By employing the ability of scBS-seq to simultaneously determine the CNA and DNA methylation profiles of a single cell, we present different evolutionary histories traced from copy number and methylation profiles in Patient GC1, paying meticulous attention to rare cells within the tumor and providing a new perspective to understand the nature and dynamics of human cancer evolution.

Clustering of tumor tissue-of-origin from CTCs

DNA methylation patterns exhibit tissue-type specific signatures which can be used to infer the tumor tissue-of-origin from circulating DNA (ctDNA) or CTCs. Such applications in liquid biopsy have been successfully demonstrated by implementing the methylation fingerprint of ctDNA in a series of recent studies (28-30). In comparison with ctDNA, CTCs are representative of intact tumor cells and believed to provide more comprehensive and accurate information regarding the tumor tissue-of-origin.

To evaluate the feasibility of utilizing DNA methylation profiles of CTCs to trace the tissue-of-origin and diagnose different cancer types, we first examined how effective the classification could be based on *a priori* knowledge. We performed *t*-SNE analysis (*Figure 5A*) and clustering (*Figure 5B*) according to the methylation level of CpG islands in all samples from our 17 patients. Our analytical

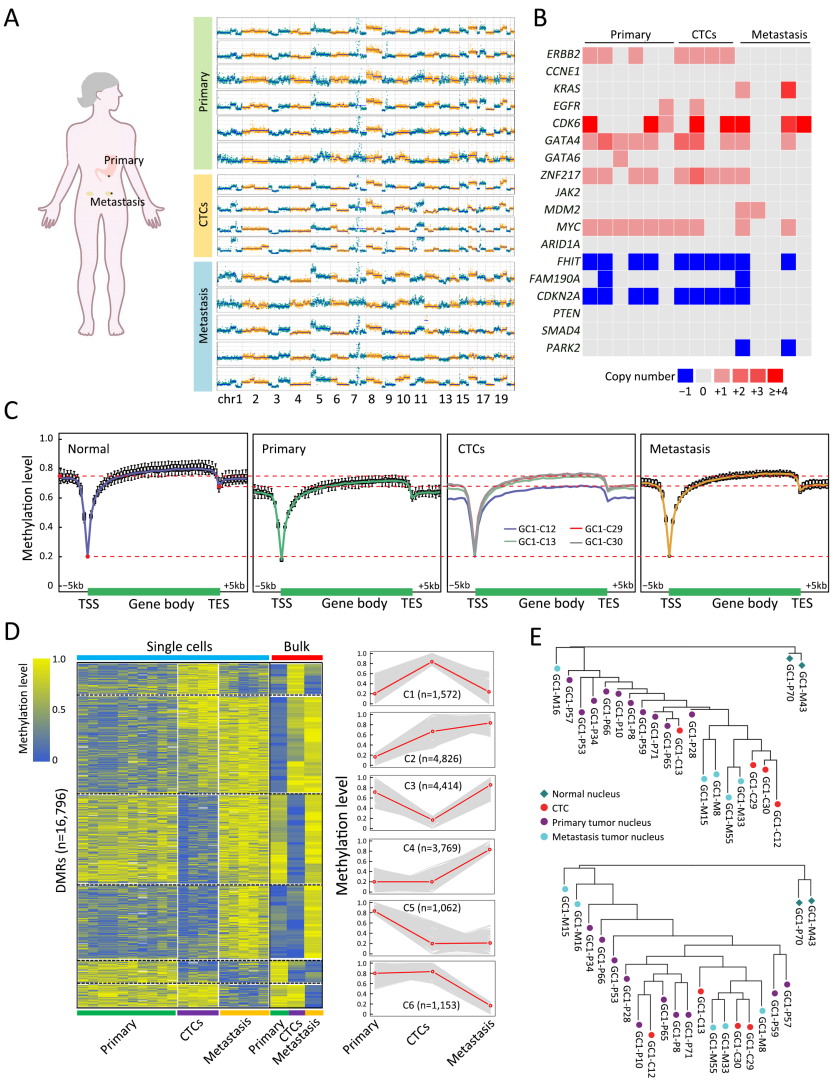

Figure 4 Different evolutionary histories traced from copy number and methylation profiles in Patient GC1. (A) Left: a diagram of tumor progression in Patient GC1 who was diagnosed with GC accompanied by abdominal ovarian metastasis; right: CNA deduction results of scBS-seq datasets from Patient GC1. The copy numbers (blue and red dots) were plotted along the genome at a bin size of 500 kb. The ordinate coordinate represents copy numbers ranging from 0 to 6 (a copy number >6 was set to 6); (B) CNA regions containing known GC-associated genes. Deletions (blue) and amplifications (red) are shown; (C) DNA methylation in gene body regions of normal nuclei (purple), primary tumor nuclei (green), CTCs (multi-colored), and metastatic tumor nuclei (orange) derived from Patient GC1. Average CpG methylation levels along the scaled gene bodies, 5 kb upstream of TSS, and 5 kb downstream of TES for all RefSeq genes were used for the analysis; (D) Methylation heatmap of DMRs among different samples from Patient GC1 (blue, individual cells; red, pseudo-bulk sample merged with individual cells of the same cell type). Cell types were color coded: green (primary tumor nuclei), purple (CTCs), and orange (metastatic tumor nuclei). The color key from blue to yellow indicates low to high methylation level. The line charts show representative DNA methylation patterns during tumor metastasis; (E) Phylogenetic (top) and phyloepigenetic (bottom) reconstruction showing evolutionary histories in Patient GC1 inferred from the chromosomal breakpoint profiles (top) and DNA methylation distance matrix (bottom). Cell types were color coded: peacock green (primary normal cells), red (CTCs), purple (primary tumor cells), and turquoise (metastatic tumor cells). The Hamming distance metrics are individual and non-comparable. GC, gastric cancer; CNA, copy number alteration; CTC, circulating tumor cell; TSS, transcription start site; TES, transcription end site; DMR, differentially methylated region.

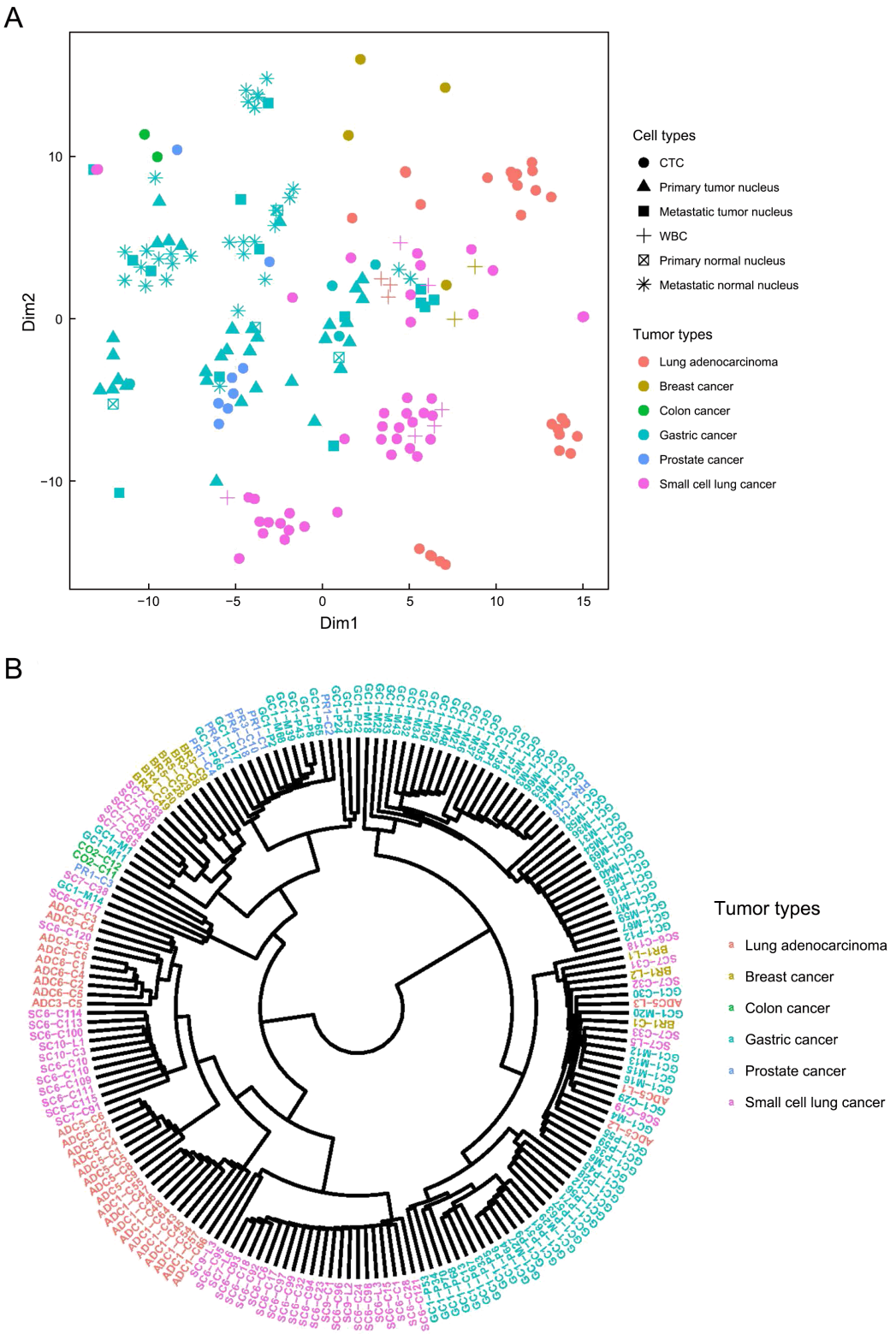

Figure 5 Classification of CTCs from different cancer types based on the tissue-specific DNA methylation signature. (A) The *t*-SNE plot grouped by cancer type and cell type; (B) Clustering deduction by cancer type. CTC, circulating tumor cell; WBC, white blood cell.

framework mainly segregated CTCs based on cancer type, such as lung ADC, SCLC, colon cancer, and breast cancer (*Figure 5A*,*B*); however, cells and nuclei from prostate cancer and GC were diversified and we failed to distinguish them effectively. Moreover, samples from the same patient tended to be clustered together (*Figure 5B*). Specifically, in the case of Patient GC1, the CTCs and primary and metastatic nuclei were clustered together, irrespective of tumor or normal origin, indicating that the tissue-specific DNA methylation signature is more predominant than differences in copy numbers.

Discussion

For comparison with bulk measurements, the association between genetic and epigenetic evolutionary histories was revisited at the single-cell level. Of interest is elucidating the association between intra-tumoral genomic heterogeneity (ITGH) and intra-tumoral methylation heterogeneity (ITMH).

For a better understanding of ITH, we acquired both CNA and DNA methylation information from individual CTCs. The phylogenetic tree for Patient SC6 exhibited an abrupt burst of aneuploid copy number changes, supporting a model of punctuated copy number evolution (PCNE) (22,31), whereas the phyloepigenetic tree showed a gradual acquisition of methylation variance (32). Our findings differ from the results based on bulk samples that support the concept of co-dependency of aberrant DNA methylation and genetic alterations (21,26). This may reflect a distinction between clinical samples of different cancer types, differences in applied methods, or the power of single-cell sequencing to identify signals obscured in bulk sampling.

Tumor cells subvert both the genome and the epigenome to evolve mechanisms by which to escape growth control and host surveillance (33). Given that DNA methylation is reversible and more error-prone than DNA replication, in addition to the rate, timing, and location of exact DNA methylation changes being obscure, the extent of co-dependency between genetics and epigenetics remains uncertain (31). As some evidence supports, genetic and epigenetic mechanisms may not be separate events in cancer development, instead intertwining and taking advantage of each other during tumorigenesis (34). However, further studies using single-cell sequencing may shed light on this problem.

Cancer with an unknown primary site is occasionally seen in clinical practice, which brings diagnostic challenges and is usually associated with a poor prognosis (35). In the present study, we aimed to trace tumor tissue-of-origin and classify different cancer types according to the methylation profiles of CTCs. Unfortunately, our clustering results are not yet ideal. In future studies, a greater number of CTC samples from different cancer types will help to build a comprehensive methylation feature map, yielding a more precise classification.

Single-cell sequencing is of great significance for the in-depth and systematic understanding of CTCs. In recent years, rapid technological development has been seen in single-cell genome, transcriptome, epigenome, and non-coding RNA sequencing. Our work employed scBS-seq to reveal both genomic and methylation information from a single CTC; however, transcriptomic information is missing. In future studies, multi-omics information should be obtained from a single CTC. In particular, transcriptomic profiles are important for understanding the biological state of CTCs, which could also facilitate classification of tumor tissue-of-origin. To enable single-cell RNA sequencing of CTCs, a gentle and less invasive enrichment method is needed.

Conclusions

Our exploratory work provides an important survey of the single-cell DNA methylome in CTCs. We simultaneously analyzed the copy number and DNA methylation profiles of CTCs at the single-cell level. Moreover, we characterized the tumor heterogeneity and built an evolutionary history of the tumor cell methylome during cancer metastasis. Furthermore, we demonstrated the potential to classify tumor origin based on tissue-specific DNA methylation profiles from CTCs.

Acknowledgements

This study was financially supported by the Guangdong Province Key Research and Development Program (No. 2019B020226002) and the National Science and Technology Major Project (No. 2019YFC1315702).

Footnote

Conflicts of Interest: The authors have no conflicts of interest to declare.

References

1. Ni X, Zhuo M, Su Z, et al. Reproducible copy number variation patterns among single circulating tumor cells of lung cancer patients. Proc Natl Acad Sci U S A 2013;110:21083-8.
2. Aceto N, Bardia A, Miyamoto DT, et al. Circulating tumor cell clusters are oligoclonal precursors of breast cancer metastasis. Cell 2014;158:1110-22.
3. Gao Y, Ni X, Guo H, et al. Single-cell sequencing deciphers a convergent evolution of copy number alterations from primary to circulating tumor cells. Genome Res 2017;27:1312-22.
4. Su Z, Wang Z, Ni X, et al. Inferring the evolution and progression of small-cell lung cancer by single-cell sequencing of circulating tumor cells. Clin Cancer Res 2019;25:5049-60.
5. Ramsköld D, Luo S, Wang YC, et al. Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells. Nat Biotechnol 2012;30:777-82.
6. Chimonidou M, Strati A, Tzitzira A, et al. DNA methylation of tumor suppressor and metastasis suppressor genes in circulating tumor cells. Clin Chem 2011;57:1169-77.
7. Chimonidou M, Strati A, Malamos N, et al. SOX17 promoter methylation in circulating tumor cells and matched cell-free DNA isolated from plasma of patients with breast cancer. Clin Chem 2013;59: 270-9.
8. Chimonidou M, Kallergi G, Georgoulias V, et al. Breast cancer metastasis suppressor-1 promoter methylation in primary breast tumors and corresponding circulating tumor cells. Mol Cancer Res 2013;11:1248-57.
9. Friedlander TW, Ngo VT, Dong H, et al. Detection and characterization of invasive circulating tumor cells derived from men with metastatic castration-resistant prostate cancer. Int J Cancer 2014;134:2284-93.
10. Ogunwobi OO, Puszyk W, Dong HJ, et al. Epigenetic upregulation of HGF and c-Met drives metastasis in hepatocellular carcinoma. PLoS One 2013;8:e63765.
11. Pixberg CF, Raba K, Müller F, et al. Analysis of DNA methylation in single circulating tumor cells. Oncogene 2017;36:3223-31.
12. Clark SJ, Smallwood SA, Lee HJ, et al. Genome-wide base-resolution mapping of DNA methylation in single cells using single-cell bisulfite sequencing (scBS-seq). Nat Protoc 2017;12:534-47.
13. Wang Y, Waters J, Leung ML, et al. Clonal evolution in breast cancer revealed by single nucleus genome sequencing. Nature 2014;512:155-60.
14. Baslan T, Kendall J, Rodgers L, et al. Genome-wide copy number analysis of single cells. Nat Protoc 2012; 7:1024-41.
15. Willenbrock H, Fridlyand J. A comparison study: applying segmentation to array CGH data for downstream analyses. Bioinformatics 2005;21: 4084-91.
16. Guo H, Zhu P, Yan L, et al. The DNA methylation landscape of human early embryos. Nature 2014;511: 606-10.
17. 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.
18. Bian S, Hou Y, Zhou X, et al. Single-cell multiomics sequencing and analyses of human colorectal cancer. Science 2018;362:1060-3.
19. Nelson HH, Marsit CJ, Christensen BC, et al. Key epigenetic changes associated with lung cancer development: results from dense methylation array profiling. Epigenetics 2012;7:559-66.
20. Gkountela S, Castro-Giner F, Szczerba BM, et al. Circulating tumor cell clustering shapes DNA methylation to enable metastasis seeding. Cell 2019; 176:98-112.
21. Mazor T, Pankov A, Johnson BE, et al. DNA methylation and somatic mutations converge on the cell cycle and define similar evolutionary histories in brain tumors. Cancer Cell 2015;28:307-17.
22. Navin N, Kendall J, Troge J, et al. Tumour evolution inferred by single-cell sequencing. Nature 2011;472: 90-4.
23. Nam AS, Chaligne R, Landau DA. Integrating genetic and non-genetic determinants of cancer evolution by single-cell multi-omics. Nat Rev Genet 2021;22:3-18.
24. Greaves M, Maley CC. Clonal evolution in cancer. Nature 2012;481:306-13.
25. Zhang Y, Lima CF, Rodrigues LR. Anticancer effects

of lactoferrin: underlying mechanisms and future trends in cancer therapy. Nutr Rev 2014;72:763-73.
26. Brocks D, Assenov Y, Minner S, et al. Intratumor DNA methylation heterogeneity reflects clonal evolution in aggressive prostate cancer. Cell Rep 2014;8:798-806.
27. Mazor T, Pankov A, Song JS, et al. Intratumoral heterogeneity of the epigenome. Cancer Cell 2016;29:440-51.
28. Xu RH, Wei W, Krawczyk M, et al. Circulating tumour DNA methylation markers for diagnosis and prognosis of hepatocellular carcinoma. Nat Mater 2017;16:1155-61.
29. Guo S, Diep D, Plongthongkum N, et al. Identification of methylation haplotype blocks aids in deconvolution of heterogeneous tissue samples and tumor tissue-of-origin mapping from plasma DNA. Nat Genet 2017;49:635-42.
30. Luo H, Zhao Q, Wei W, et al. Circulating tumor DNA methylation profiles enable early diagnosis, prognosis prediction, and screening for colorectal cancer. Sci Transl Med 2020;12:eaax7533.
31. 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.
32. Davis A, Gao R, Navin N. Tumor evolution: Linear, branching, neutral or punctuated? Biochim Biophys Acta Rev Cancer 2017;1867:151-61.
33. Jones PA, Issa JP, Baylin S. Targeting the cancer epigenome for therapy. Nat Rev Genet 2016;17:630-41.
34. You JS, Jones PA. Cancer genetics and epigenetics: two sides of the same coin? Cancer Cell 2012;22:9-20.
35. Moran S, Martínez-Cardús A, Sayols S, et al. Epigenetic profiling to classify cancer of unknown primary: a multicentre, retrospective analysis. Lancet Oncol 2016;17:1386-95.

Cite this article as: Chen H, Su Z, Li R, Zhang N, Guo H, Bai F. Single-cell DNA methylome analysis of circulating tumor cells. Chin J Cancer Res 2021;33(3):391-404. doi: 10.21147/j.issn.1000-9604.2021.03.10

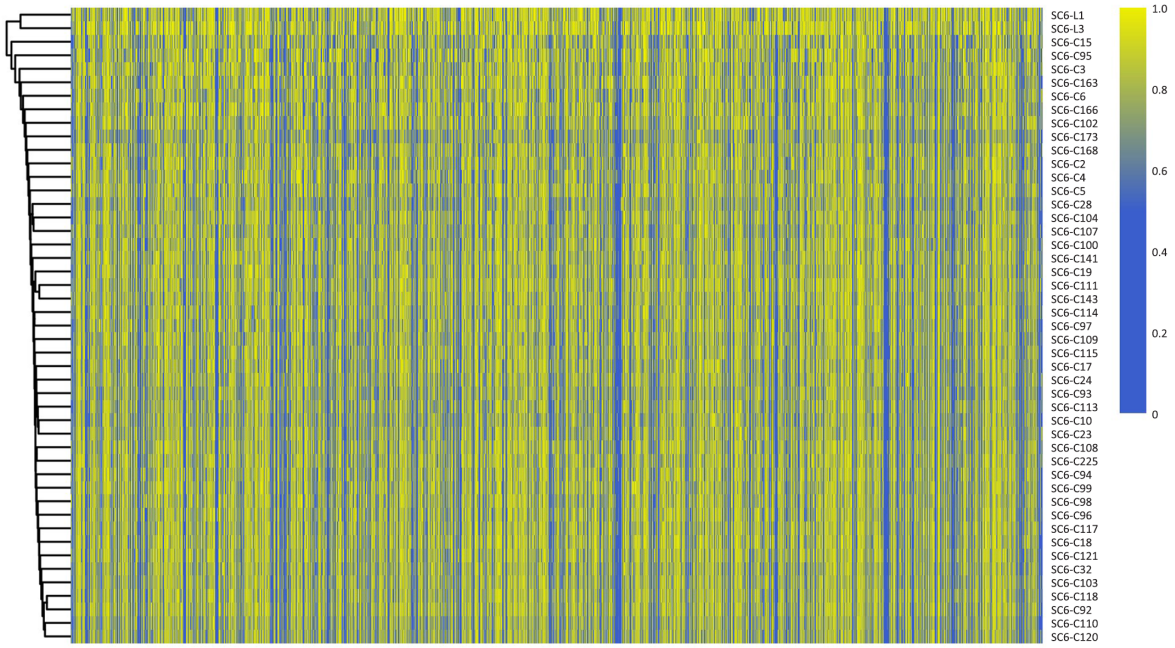

Figure S1 Unsupervised hierarchical clustering of methylome of individual CTCs and WBCs derived from Patient SC6. Rows of the heatmap display the methylation levels along chromosomal regions. The color key from blue to yellow indicates low to high methylation level. All detected CpG sites were used for the analysis. CTC, circulating tumor cell; WBC, white blood cell.

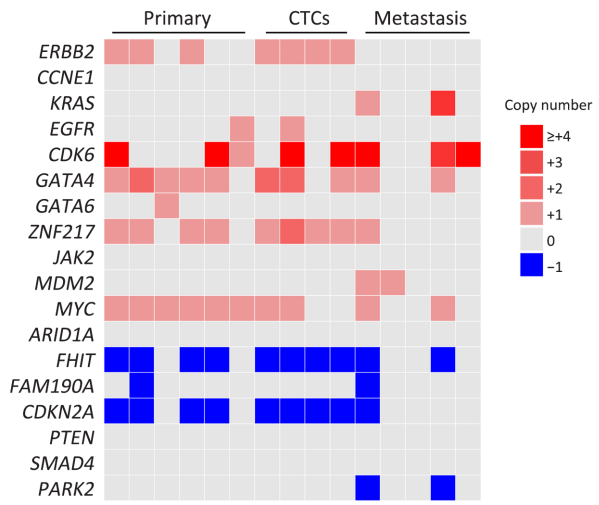

Figure S2 Methylation heatmap of individual CTCs and tumor nuclei from paired primary and metastatic tumor tissue derived from Patient GC1. A total of 20 known gastric cancer-associated gene promoters were used for the analysis. The color key from blue to yellow indicates low to high methylation level; the white checks indicate unavailable data. CTC, circulating tumor cell.

Table S1 Patient recruitment

Patient ID	Age (year)	Gender	Cancer type	Stage	Metastasis	CTCs	WBCs	Tumor nuclei	Normal nuclei
GC1	51	Female	Gastric cancer	IV	Ovary	4	4	39	34
BR1	NA	Female	Breast cancer	IV	NA	1	2	0	0
BR3	47	Female	Breast cancer	IV	Liver, lung	2	0	0	0
BR4	46	Female	Breast cancer	IIIA	NA	2	0	0	0
BR5	56	Female	Breast cancer	IIIA	NA	2	0	0	0
PR1	NA	Male	Prostate cancer	NA	NA	4	0	0	0
PR3	72	Male	Prostate cancer	IV	Bone	1	0	0	0
PR4	73	Male	Prostate cancer	IV	Bone	3	0	0	0
CO2	61	Female	Colon cancer	IV	Liver, lymph node	2	0	0	0
SC6	64	Male	Small cell lung cancer	IV	NA	46	2	0	0
SC7	63	Male	Small cell lung cancer	IV	NA	10	2	0	0
SC9	41	Female	Small cell lung cancer	IV	NA	1	2	0	0
SC10	53	Male	Small cell lung cancer	IV	NA	2	1	0	0
ADC1	53	Female	Lung adenocarcinoma	IV	Liver, bone	10	0	0	0
ADC3	64	Male	Lung adenocarcinoma	IV	Pleura	3	0	0	0
ADC5	NA	Female	Lung adenocarcinoma	NA	Mammary gland	9	3	0	0
ADC6	50	Male	Lung adenocarcinoma	IIIB	NA	5	0	0	0

CTC, circulating tumor cell; WBC, white blood cell; NA, not available.

Table S2 Sequencing information for all samples

Cell type	Sample ID	Pair reads	Mapping efficiency (2×) (%)	No. of covered CpG	Covered CpG (%)
CTC	GC1-C12	14,690,291	34.25	8,923,559	15.81
CTC	GC1-C13	15,127,453	36.45	10,254,064	18.17
CTC	GC1-C29	16,582,575	37.37	10,379,210	18.39
CTC	GC1-C30	16,377,413	38.34	10,940,083	19.39
WBC	GC1-L1	8,326,219	23.96	4,646,516	8.23
WBC	GC1-L2	13,096,379	19.91	7,078,053	12.54
WBC	GC1-L3	25,146,875	22.63	13,510,695	23.94
WBC	GC1-L4	17,596,259	22.59	10,207,329	18.09
Normal	GC1-P1	10,280,635	21.50	5,541,956	9.82
Tumor	GC1-P2	12,938,455	18.90	6,488,517	11.50
Tumor	GC1-P3	10,005,921	22.08	5,923,237	10.50
Tumor	GC1-P5	11,117,342	23.01	7,154,027	12.68
Normal	GC1-P6	13,180,088	20.75	8,149,880	14.44
Tumor	GC1-P8	11,607,560	22.93	7,663,399	13.58
Tumor	GC1-P10	15,931,839	20.25	8,815,803	15.62
Tumor	GC1-P11	9,155,050	18.64	3,445,136	6.10
Tumor	GC1-P16	12,765,868	21.12	6,746,019	11.95
Tumor	GC1-P17	19,326,842	6.30	3,803,176	6.74
Tumor	GC1-P28	59,970,421	23.81	19,740,723	34.98
Tumor	GC1-P33	10,055,129	21.90	5,824,825	10.32
Tumor	GC1-P34	16,381,627	26.79	11,113,395	19.69
Tumor	GC1-P43	10,668,761	22.24	6,349,919	11.25
Tumor	GC1-P45	8,389,850	18.65	4,630,085	8.20
Tumor	GC1-P51	20,102,532	23.06	11,925,607	21.13
Tumor	GC1-P53	13,804,687	23.65	8,918,748	15.80
Tumor	GC1-P55	9,403,480	20.88	6,003,802	10.64
Tumor	GC1-P56	14,791,169	22.28	9,935,776	17.61
Tumor	GC1-P57	12,449,958	22.55	8,227,586	14.58
Tumor	GC1-P59	8,044,470	23.71	5,675,620	10.06
Tumor	GC1-P61	7,626,478	17.35	4,153,380	7.36
Normal	GC1-P63	7,547,766	21.10	4,342,340	7.69
Tumor	GC1-P64	9,761,796	22.94	6,373,298	11.29
Tumor	GC1-P65	20,128,066	23.41	11,714,496	20.76
Tumor	GC1-P66	18,037,022	23.35	10,905,290	19.32
Tumor	GC1-P67	10,105,637	18.96	5,356,092	9.49
Tumor	GC1-P68	11,544,975	18.90	6,326,638	11.21
Normal	GC1-P70	7,843,436	21.04	4,924,141	8.73
Tumor	GC1-P71	15,240,399	15.36	7,071,459	12.53
Tumor	GC1-M1	12,270,201	22.67	6,596,035	11.69
Normal	GC1-M3	12,365,544	25.64	7,019,791	12.44
Normal	GC1-M4	11,488,117	23.34	6,429,573	11.39

Table S2 (*continued*)

Table S2 (*continued*)

Cell type	Sample ID	Pair reads	Mapping efficiency (2×) (%)	No. of covered CpG	Covered CpG (%)
Tumor	GC1-M6	10,110,542	23.18	5,207,793	9.23
Normal	GC1-M7	13,453,720	23.34	6,679,339	11.84
Tumor	GC1-M8	12,320,271	24.30	7,598,536	13.46
Tumor	GC1-M12	10,004,606	26.42	6,324,223	11.21
Tumor	GC1-M13	14,200,596	21.98	8,130,235	14.41
Tumor	GC1-M15	25,583,204	25.46	13,796,276	24.45
Tumor	GC1-M16	16,581,685	25.00	9,887,464	17.52
Normal	GC1-M18	16,102,509	22.04	8,642,190	15.31
Normal	GC1-M20	12,464,221	19.92	5,239,128	9.28
Tumor	GC1-M21	11,996,318	24.56	6,182,276	10.95
Tumor	GC1-M25	7,311,016	18.81	3,562,720	6.31
Normal	GC1-M27	8,720,905	21.71	5,120,014	9.07
Normal	GC1-M29	8,297,463	22.47	5,537,984	9.81
Normal	GC1-M30	7,950,082	17.79	4,154,749	7.36
Normal	GC1-M31	8,274,400	22.94	4,819,799	8.54
Normal	GC1-M32	13,520,043	21.03	6,729,491	11.92
Tumor	GC1-M33	19,107,941	22.34	10,171,375	18.02
Normal	GC1-M34	8,940,347	22.23	5,315,042	9.42
Normal	GC1-M35	13,434,690	13.63	5,084,402	9.01
Normal	GC1-M36	13,319,731	24.92	5,437,391	9.63
Normal	GC1-M37	24,712,849	24.07	12,520,168	22.19
Normal	GC1-M38	17,055,382	23.40	9,392,753	16.64
Tumor	GC1-M39	16,881,284	20.87	7,926,600	14.05
Normal	GC1-M40	10,572,990	25.10	6,968,290	12.35
Normal	GC1-M43	10,491,094	26.05	6,465,413	11.46
Tumor	GC1-M44	9,176,188	25.49	6,269,247	11.11
Normal	GC1-M45	8,642,946	25.02	5,624,778	9.97
Normal	GC1-M46	14,232,395	8.89	3,719,035	6.59
Normal	GC1-M48	13,359,213	24.42	7,104,204	12.59
Normal	GC1-M51	40,683,880	24.95	13,795,961	24.45
Normal	GC1-M52	15,495,617	24.96	9,389,496	16.64
Normal	GC1-M54	12,693,692	22.55	6,854,436	12.15
Tumor	GC1-M55	16,861,790	24.29	9,157,418	16.23
Normal	GC1-M58	18,296,800	10.03	5,102,083	9.04
Normal	GC1-M59	10,328,505	23.16	6,237,797	11.05
Normal	GC1-M63	13,945,728	11.70	4,755,874	8.43
Normal	GC1-M66	10,596,969	23.48	6,171,626	10.94
Normal	GC1-M67	13,371,112	21.55	6,514,642	11.54
Normal	GC1-M68	13,802,435	22.88	7,306,918	12.95
Normal	GC1-M69	12,006,900	22.70	6,581,664	11.66
CTC	SC6-C2	10,623,852	25.25	5,178,508	9.18
CTC	SC6-C3	14,139,172	25.13	4,466,900	7.92

Table S2 (*continued*)

Table S2 (*continued*)

Cell type	Sample ID	Pair reads	Mapping efficiency (2×) (%)	No. of covered CpG	Covered CpG (%)
CTC	SC6-C4	16,471,120	21.76	5,815,004	10.30
CTC	SC6-C5	15,555,308	25.21	6,280,963	11.13
CTC	SC6-C6	7,328,574	26.86	3,949,284	7.00
CTC	SC6-C10	16,207,418	21.82	7,292,541	12.92
CTC	SC6-C15	9,829,838	12.23	2,338,300	4.14
CTC	SC6-C17	11,403,314	24.12	6,185,321	10.96
CTC	SC6-C18	14,723,105	20.62	6,739,918	11.94
CTC	SC6-C19	11,085,731	21.42	5,771,081	10.23
CTC	SC6-C23	13,216,934	24.39	6,891,780	12.21
CTC	SC6-C24	12,448,121	23.21	6,473,552	11.47
CTC	SC6-C28	11,427,930	22.06	5,806,958	10.29
CTC	SC6-C32	26,061,739	20.63	8,873,496	15.72
CTC	SC6-C92	27,549,994	23.12	11,725,029	20.78
CTC	SC6-C93	12,951,246	22.87	6,481,704	11.49
CTC	SC6-C94	11,306,224	21.96	5,842,692	10.35
CTC	SC6-C95	11,219,917	8.68	2,088,008	3.70
CTC	SC6-C96	12,447,819	23.66	6,495,877	11.51
CTC	SC6-C97	12,950,804	18.91	5,473,335	9.70
CTC	SC6-C98	12,441,573	21.54	6,137,426	10.88
CTC	SC6-C99	9,691,483	23.79	5,675,860	10.06
CTC	SC6-C100	13,362,912	13.25	4,631,184	8.21
CTC	SC6-C102	10,588,711	23.86	4,928,700	8.73
CTC	SC6-C103	17,874,018	25.35	8,844,052	15.67
CTC	SC6-C104	12,576,264	25.17	6,316,454	11.19
CTC	SC6-C107	14,532,103	23.22	6,359,667	11.27
CTC	SC6-C108	15,757,246	25.65	7,309,406	12.95
CTC	SC6-C109	9,733,417	23.85	5,567,069	9.86
CTC	SC6-C110	13,359,702	25.10	8,418,570	14.92
CTC	SC6-C111	16,778,997	21.96	9,243,185	16.38
CTC	SC6-C113	13,248,561	21.98	7,754,688	13.74
CTC	SC6-C114	14,044,297	15.45	5,370,867	9.52
CTC	SC6-C115	12,979,007	18.89	5,589,438	9.90
CTC	SC6-C117	15,213,928	23.19	7,182,503	12.73
CTC	SC6-C118	18,805,362	21.89	9,103,813	16.13
CTC	SC6-C120	18,943,746	22.06	9,074,315	16.08
CTC	SC6-C121	15,220,940	21.56	7,225,036	12.80
CTC	SC6-C137	15,718,408	26.81	8,415,128	14.91
CTC	SC6-C141	14,108,998	21.71	6,040,371	10.70
CTC	SC6-C143	14,544,004	24.86	7,700,367	13.64
CTC	SC6-C163	10,357,457	17.91	3,682,022	6.52
CTC	SC6-C166	13,280,346	14.47	4,266,137	7.56

Table S2 (*continued*)

Table S2 (*continued*)

Cell type	Sample ID	Pair reads	Mapping efficiency (2×) (%)	No. of covered CpG	Covered CpG (%)
CTC	SC6-C168	10,891,939	25.72	5,333,041	9.45
CTC	SC6-C173	16,067,741	19.33	5,548,685	9.83
CTC	SC6-C225	12,732,117	25.19	6,765,224	11.99
WBC	SC6-L1	9,485,519	25.41	4,415,781	7.8
WBC	SC6-L3	8,961,054	26.05	3,981,417	7.05
CTC	SC7-C31	12,714,175	23.34	5,284,828	9.36
CTC	SC7-C32	12,993,932	24.39	6,246,898	11.07
CTC	SC7-C33	14,804,741	25.20	7,036,916	12.47
CTC	SC7-C36	11,566,135	22.49	5,715,660	10.13
CTC	SC7-C38	10,465,521	23.56	4,906,689	8.69
CTC	SC7-C83	11,673,811	24.43	5,511,189	9.77
CTC	SC7-C84	13,815,221	24.58	5,575,205	9.88
CTC	SC7-C85	14,226,786	24.79	6,223,694	11.03
CTC	SC7-C90	13,534,504	24.68	6,091,151	10.79
CTC	SC7-C91	13,753,956	23.70	5,477,690	9.71
WBC	SC7-L5	9,041,774	26.32	4,764,931	8.44
WBC	SC7-L6	4,227,353	26.30	2,865,506	5.08
CTC	SC9-C1	8,456,137	26.59	3,448,009	6.11
WBC	SC9-L2	5,547,824	24.04	2,923,312	5.18
WBC	SC9-L3	5,024,901	25.51	2,977,638	5.28
CTC	SC10-C1	9,560,646	27.61	5,608,956	9.94
CTC	SC10-C3	7,637,244	25.84	3,876,736	6.87
WBC	SC10-L1	9,034,180	25.96	4,062,800	7.20
CTC	ADC1-C43	15,780,502	25.36	7,391,389	13.10
CTC	ADC1-C45	13,319,854	26.02	6,049,663	10.72
CTC	ADC1-C46	14,309,217	25.11	7,797,003	13.82
CTC	ADC1-C47	15,748,063	26.43	9,239,604	16.37
CTC	ADC1-C48	17,802,407	25.89	7,481,921	13.26
CTC	ADC1-C54	14,698,210	28.74	5,650,975	10.01
CTC	ADC1-C55	17,232,293	27.43	8,451,985	14.98
CTC	ADC1-C57	13,614,737	28.48	5,084,288	9.01
CTC	ADC1-C64	15,642,736	20.67	5,902,550	10.46
CTC	ADC1-C66	18,338,219	19.44	4,422,499	7.84
CTC	ADC3-C3	4,286,961	27.69	3,698,495	6.55
CTC	ADC3-C4	2,623,531	30.67	2,368,737	4.20
CTC	ADC3-C5	6,317,688	23.32	4,158,014	7.37
CTC	ADC5-C1	12,919,564	22.10	6,490,620	11.50
CTC	ADC5-C2	18,819,378	22.95	8,377,833	14.85
CTC	ADC5-C3	24,658,619	10.71	4,320,597	7.66
CTC	ADC5-C4	17,873,495	20.25	7,403,731	13.12
CTC	ADC5-C5	13,082,685	19.90	5,781,666	10.24

Table S2 (*continued*)

Table S2 (*continued*)

Cell type	Sample ID	Pair reads	Mapping efficiency (2×) (%)	No. of covered CpG	Covered CpG (%)
CTC	ADC5-C6	38,024,868	19.89	11,032,124	19.55
CTC	ADC5-C7	18,696,400	21.31	7,728,075	13.69
CTC	ADC5-C8	9,541,521	21.24	5,218,519	9.25
CTC	ADC5-C9	7,706,715	23.13	4,742,244	8.40
WBC	ADC5-L1	15,797,286	21.98	7,058,024	12.51
WBC	ADC5-L2	12,445,122	19.82	5,610,195	9.94
WBC	ADC5-L3	12,333,753	16.90	4,823,956	8.55
CTC	ADC6-C1	14,126,121	24.08	8,750,735	15.51
CTC	ADC6-C2	13,074,137	19.19	6,233,262	11.05
CTC	ADC6-C4	16,387,703	23.39	8,830,325	15.65
CTC	ADC6-C5	11,933,664	20.37	5,094,588	9.03
CTC	ADC6-C6	18,688,676	24.04	9,880,752	17.51
CTC	BR1-C1	19,863,868	25.72	8,330,663	14.76
WBC	BR1-L1	11,353,494	26.30	5,432,662	9.63
WBC	BR1-L2	13,557,222	24.59	5,954,832	10.55
CTC	BR3-C8	13,417,989	24.79	6,475,217	11.47
CTC	BR3-C9	13,425,021	23.78	6,163,949	10.92
CTC	BR4-C49	21,240,841	18.52	7,777,848	13.78
CTC	BR4-C50	12,807,969	22.24	6,725,071	11.92
CTC	BR5-C28	22,035,079	16.67	6,843,138	12.13
CTC	BR5-C29	12,809,952	20.31	6,072,732	10.76
CTC	CO2-C11	19,338,547	17.01	7,218,625	12.79
CTC	CO2-C12	21,695,855	7.58	4,166,000	7.38
CTC	PR1-C1	10,034,150	23.30	4,721,387	8.37
CTC	PR1-C2	15,242,064	19.20	6,547,983	11.60
CTC	PR1-C3	12,426,412	14.70	4,272,652	7.57
CTC	PR1-C4	12,091,913	18.05	5,223,939	9.26
CTC	PR3-C10	10,673,823	25.22	5,923,919	10.50
CTC	PR4-C16	8,769,356	25.50	5,200,400	9.21
CTC	PR4-C17	12,138,753	19.81	5,205,920	9.22
CTC	PR4-C18	16,986,231	27.19	9,222,482	16.34

CTC, circulating tumor cell; WBC, white blood cell; The abbreviation “-C” means CTC and “-L” indicates WBC, whereas “P” and “M” represent the nucleus from the primary tumor and metastases, respectively.