# Circulating tumor cell methylation profiles reveal the classification and evolution of non-small cell lung cancer

**Jia-Hao Jiang**
Zhongshan Hospital Fudan University
**Chang-Yue Chen**
Shanghai zhiyi biological technology company
**Jian Gao**
Zhongshan Hospital Fudan University
**Jing Li**
shanghai zhiyi biomedical technology company
**Yuan Lu**
shanghai zhiyi biomedical technology company
**Wang Fang**
Jinkang an pharmaceutical company
**Hai-Kun Wang**
Institut Pasteur of Shanghai Chinese Academy of Sciences
**Yun-Feng Yuan**
Zhongshan Hospital Fudan University
**Jian-Yong Ding** ( ding.jianyong@zs-hospital.sh.cn )
Zhongshan Hospital Fudan University

**Research**

**Keywords:** non-small cell lung cancer, circulating tumor cells, DNA methylation

**Posted Date:** August 31st, 2021

**DOI:** https://doi.org/10.21203/rs.3.rs-842155/v1

# Abstract

Backgrounds

The ability of circulating tumor cells (CTCs) to identify lung adenocarcinoma (LUAD) and squamous cell carcinoma (LUSC) holds great promise for improving pathological diagnosis and selecting treatment in non-small cell lung cancer (NSCLC). In addition, previous studies have shown that DNA methylation exhibits cell and tissue specificity. Thus, we aimed to explore the methylation status of CTCs in LUAD and LUSC and identify the potential biomarkers.

Methods

we first analyzed Infinium 450K methylation profiles obtained from The Cancer Genome Atlas and Gene Expression Omnibus and further identified the results by performing whole-genome sequencing of CTCs in tumor and matched normal lung tissues and white blood cells from 6 NSCLC patients.

Results

the bioinformatic analysis revealed that an NSCLC-specific DNA methylation marker panel which could accurately distinguish between LUAD and LUSC with high diagnostic accuracy. The whole-genome sequencing of CTCs in NSCLC patients also showed 100% accuracy for distinguishing between LUAD and LUSC based on CTC methylation profiles. To investigate the function of CTCs, we further analyzed similar and different methylation profiles between CTCs and their primary tumors and found very high similarities between CTCs and their primary tumor tissues, indicating that these cells inherit information from primary tumors. CTCs also showed some characteristics that were different from those of the primary tumor tissues, suggesting that CTCs evolve some unique characteristics after migrating from the primary tumor; these characteristics may be one of the reasons for the ability of tumor cells to evade immune surveillance.

Conclusion

thus, our study provides insight into the potential use of CTCs in pathological classification of NSCLC patients as well as CTC primary tumor inheritance and CTC evolution influence metastasis and immune escape.


# Background

Lung cancer is one of the world's most common and deadliest forms of cancer [1]. Approximately 85% of lung cancers are non-small cell lung cancers (NSCLCs), which include lung adenocarcinoma (LUAD) and lung squamous cell carcinoma (LUSC). As the main treatment options are determined according to histologic features, a pathological diagnosis is key for NSCLC treatment. Current pathological diagnosis is based on the morphologic features or on immune-cytochemical and immune-histochemical analyses of NSCLC tissue or cytologic samples obtained via surgical biopsy, bronchoscopy or bronchial brushing

[2]. However, it is difficult to make a pathological diagnosis in cases in which tumor biopsy or cytology material is not available.

Circulating tumor cells (CTCs) are cancer cells that have disseminated from primary or metastatic sites into the peripheral blood and present great potential as diagnostic and prognostic biomarkers for guiding individualized treatment in lung cancer. In the lung cancer diagnostic field, the CTC detection rate is approximately 87% among those who have more than 3 CTCs per 3.2 ml of blood [3]. Furthermore, CTCs carry information on the primary tumor cells and are considered an alternative means of tumor subtype classification [4].

In metastasis, the methylation status of the primary tumor is inherited by the metastatic tumor. Previous studies have shown that DNA methylation exhibits cell and tissue specificity [5], and changes in DNA methylation play an important regulatory role in the development of cancer [6, 7]. Indeed, both genome-wide hypomethylation and hypermethylation modifications have the ability to alter the expression of neighboring genes and contribute to cancer phenotypes [7, 8]. Although DNA methylation has been extensively investigated in primary tumors [9], the events that shape the DNA methylome during metastatic dissemination are largely uncharacterized [10, 11]. Overall, the methylation profile of CTC DNA may broaden our understanding of tumor cell origin and evolution.

In this study, we combined immunostaining-FISH-based CTC identification, laser capture microdissection-based CTC capture and single-cell resolution DNA methylation to explore CTC methylation signatures in the origin, classification and evolution of these cells in NSCLC. Our study provides a genome-wide DNA methylation landscape of primary tumor tissues, CTCs, matched normal lung tissues and white blood cells (WBCs) in 6 NSCLC patients. The results demonstrate that CTCs can be used as an effective blood-based method for the classification of LUAD versus LUSC. Results also evidenced that both CTC primary tumor inheritance and CTC evolution affect metastasis and immune escape.

## Methods

# Characteristics of patients and samples

Clinical characteristics and molecular profiling, including methylation data for a training cohort of 553 tumor samples and 60 matched adjacent normal tissue samples, as well as a validation cohort of 288 tumor and 14 matched normal samples, were obtained from The Cancer Genome Atlas (TCGA). A separate validation cohort of 37 tumor samples and 74 normal samples was obtained from Gene Expression Omnibus (GEO). Another separate validation cohort of 6 tumor and 6 matched normal samples was obtained from Zhongshan Hospital of Fudan University, Shanghai, China. Matched adjacent normal tissue samples and WBCs were collected at the same time as tumor tissue from a patient and verified by histology to display no evidence of cancer. The clinical characteristics of all patients are summarized in Supplementary Table 1.

# Experimental method details

# Subtraction enrichment of CTCs and identification of aneuploid CTCs

Enrichment and identification of CTCs were performed according to the CTCseq™ kit instructions (Majorbio). Samples were fixed on slides, followed by counting and photographing. Each suspicious tumor cell coordinate was recorded to facilitate subsequent target cell identification. The identification principle of CTCs is that (i) they are negative for CD45 and (ii) positive for chromosome 8 heteroploidy. The slides were stored at -20°C.

# Laser capture microdissection (LCM)

Samples were loaded onto the stage of a Zeiss PALM MicroBeam (Zeiss) under a ×40 objective. After microdissection with a 355-nm laser beam, target cells were collected onto an AdhesiveCap 200 opaque cover (Zeiss). A 10-μl reaction volume contained 5 μl M-Digestion Buffer (2X), 0.5 μl proteinase K (EZ DNA Methylation-Direct™ Kit, Zymo) and 4.5 μl nuclease-free water (Ambion). The reagents were mixed and placed on the cover of AdhesiveCap 200 opaque; the tube was briefly microcentrifuged, and the reaction was incubated in a thermal cycler for 20 min at 50°C, with a 4°C hold. The samples were stored at -20°C.

# Whole-genome bisulfite sequencing analysis

# Tumor DNA extraction

Genomic DNA extraction from freshly frozen normal or cancer tissues or WBCs was performed with a QIAamp DNA Mini Kit (Qiagen) according to the manufacturer's recommendations. DNA was extracted from approximately 0.5 mg of tissue and stored at -20°C; the samples were analyzed within one week of preparation.

# Bisulfite conversion of genomic DNA and whole-genome bisulfite sequencing

Bisulfite conversion was performed with EZ DNA Methylation-Lightning™ Kit (Zymo Research). WGBS was performed using KAPA Hyper Prep Kit (Roche) with several modifications, as previously described [12]. The WGBS libraries of tissue and WBCs were sequenced with paired-end flow cell lanes in the HiSeq4000 system (Illumina) for 150 cycles.

# Capture and sequencing

Capture was performed using SeqCap Epi CpGiant Enrichment Kit (Roche) as directed by the manufacturer. Briefly, 4–6 bisulfite-treated libraries (200 ng/sample) were hybridized to the SeqCap Epi probe pool; the beads were captured, washed, amplified, quantified and qualified as directed in the protocol.

The captured pooled library (tissue 2N) was sequenced using the Illumina HiSeq X Ten system with a 150-bp paired-end model.

# Bisulfite conversion and single-cell whole-genome bisulfite library preparation

The library was produced according to a previously published protocol [12]. In brief, after cell lysis for 20 min, CTC samples were subjected to bisulfite conversion using EZ DNA Methylation-Direct™ Kit (Zymo) according to the manufacturer's instructions. The bisulfite-converted DNA was then synthesized using Klenow exo- (Enzymatics) with a truncated Illumina P5 adapter (5'- CTACACGACGCT CTT CC GATCTNNNNNN-3') followed by a random hexamer at the 3' end. This step was repeated four additional times for preamplification. The excess primers were removed using exonuclease I (New England Biolabs). Following purification, second strands were synthesized similarly but using a truncated P7 Illumina adapter (5'-AGACGTGTGCTCTTCCGATCTNNNNNN-3'). The final library was amplified using KAPA HiFi HotStart ReadyMix (Kapabiosystems) with NEB primers (universal primer and index primer). The amplified libraries were purified twice with 0.9 × AMPure XP beads (Beckman Coulter) and quantified using Qubit ds HS dye and a 2100 Bioanalyzer (Agilent Technologies). The final quality-ensured libraries were sequenced with a HiSeq4000 system (Illumina) for 150 cycles.

# Quantification and statistical analysis

# Processing methylation microarray data

DNA methylation data were obtained from TCGA analysis of 485,000 sites generated using Infinium 450K Methylation Array and the following GSE datasets: GSE85845, GSE83842, GSE66087, GSE63704, and GSE53051. The microarray data (level 3 in TCGA and processed matrix files in GEO database) provide the methylation levels of individual CpG sites. Methylation levels for two cancer subtypes (LUAD and LUSC) and normal lung tissues were extracted. Six matched cohorts (cancer, normal, bulk WBCs and CTCs per patient) were obtained by WGBS and analyzed as described.

# Building the multiclass classifier

For each of the 3 subtypes, LUAD and LUSC cancer and corresponding normal tissue samples from TCGA, we randomly split the full TCGA 450K dataset into training and validation sets at a 2:1 ratio. We first performed prescreening to remove excessive noise from the training data using the Dunn test. (1) A CpG site methylation level was marked as 'not available' (NA) if methylation measurements were not available for more than half of the CpG sites. (2) Any samples that had missing methylation levels of more than 5 k CpG sites were marked as 'not available' (NA). (3) For each set of comparisons, one type of sample was compared against the other 2 types of samples. A list of markers with significant differences in methylation differences $\geq 0.2$ and P values $< 0.05$ between LUAD and LUSC and in methylation differences $\geq 0.1$ and P values $< 0.05$ between LUAD and normal tissues or LUSC and normal tissues were retained for future analysis. The Benjamini-Hochberg procedure was used to control the FDR at a significance level of 0.05. For multinomial classification, we used logistic regression with the L2 regularization model (Ridge), and the tuning parameter was determined by the expected generalization error estimated from 5-fold crossvalidation. A multiclass prediction system was constructed to predict a

cancer subtype or normal sample in the validation data using the selected features. A confusion matrix and ROC curves were also produced to evaluate sensitivity and specificity in addition to prediction accuracy.

All data analyses were conducted by custom-made bash and R and Python scripts (R version = 3.4.2, Python version = 3.7.2) with the packages dunn.test (R) and sklearn (Python).

# WGBS processing

After tissue and CTC WGBS sequencing, initial quality assessment of the data was performed using FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc). Adaptor sequences, low-quality ends, and 6 bp from both the 5' and 3' ends of reads were removed with Trim Galore (v0.4.2, http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/; parameters: –clip_R1 6 –clip_R2 6 –three_prime_clip_R1 6 –three_prime_clip_R2 6). Trimmed reads were aligned to the hg19/GRCh37 human genome using Bismark with the alignment tool Bowtie2 (v2.2.9) (main parameter: –score-min L,0, 0.2) [13, 14]. Finally, methylation calls were extracted after deduplication using Bismark. Only CpG sites with depth of coverage ≥ 3x were considered for methylation analysis.

# Differential methylation analysis and enrichment analysis

Differentially methylated CpGs were assessed using a methylKit (R package) [15] at a P value less than 0.05 using the SLIM method and were considered differentially methylated [16]. We selected DMCs detected in 2 of the 6 patients for future functional pathway analysis. TBFSs in DMCs were calculated using i-cisTarget (https://gbiomed.kuleuven.be/apps/lcb/i-cisTarget/) with full motif analysis [17]. KEGG analysis with a hypergeometric test implemented in the clusterProfiler R/bioconductor package was performed using ClueGO (Vocci and London, 1997). Annotation used for CpGIslands and RefSeq genes was performed using the genomation toolkit in the R/bioconductor package.

# KEGG pathway enrichment analysis based on genomic features

Individual CpGs were mapped to genes and their promoters using RefSeq gene annotation from the UCSC genome browser (https://genome.ucsc.edu; date: 10/10/2020). Promoters were defined as the ± 2 kb region around the transcription start site (TSS). Mapping to superenhancer regions was based on dbSUPER (http://asntech.org/dbsuper/), an integrated database of superenhancers that provides a list of genes associated with each region. Each genomic feature was interrogated for differential methylation in the same manner as for genomic tiles. Similar genes corresponding to genomic features with a normal P value > 0.05 and a methylation difference < 10% were considered for enrichment analysis between CTCs and matched tumor tissues. Differential genes corresponding to genomic features with a normal P value < 0.05 and an absolute methylation difference > 0.1 were considered for enrichment analysis in CTCs compared to matched tumor tissues. Gene set enrichment analysis was performed using a hypergeometric test implemented in the clusterProfiler R/bioconductor package. Gene sets with an adjusted P value < 0.05 were considered significant.

# Results

## Identification and validation of cancer-specific differential methylation of CpG sites

To explore NSCLC-specific DNA methylation markers, we first analyzed Infinium 450K methylation profiles obtained from The Cancer Genome Atlas (TCGA) (Fig. 1). We hypothesized that the most appropriate methylation differences (LUSC versus LUAD versus Normal) will lead to the best performance in both clustering and classification. Therefore, we started with 5 cutoff parameters of beta difference (greater than a cutoff) 0.10, 0.15, 0.20, 0.25, and 0.30 among LUAD versus LUSC versus Normal. The same cutoff parameters were utilized for LUAD and LUSC compared to normal controls. Four groups of features with each fixed parameter set were identified according to whether the difference in methylation met the cutoff: □ LUAD_LUSC_Normal specific; □ LUAD_specific; □ LUSC_specific; and □ Normal_specific. We assessed 4 combinations of specific probes (□; □+□+□; □+□+□+□; □+□) under each fixed parameter set, resulting in 5x5x4 = 100 mixed samples. Finally, we obtained the optimal parameters for LUAD_LUSC_diff ≥ 0.20; LUAD_Normal_diff ≥ 0.10; LUSC_Normal_diff > 0.10; $P < 0.05$. The clustering results are shown in Fig. 2a.

Under the optimal parameters, we detected 5426 differentially methylated CpG sites (DMCs), including 5426 LUAD-LUSC cancer-specific DMCs, 1409 LUAD-normal specific DMCs and 2919 LUSC-normal specific DMCs (Supplementary Table 2). Based on differential methylation of CpG sites, we were able to distinguish LUAD, LUSC and normal tissues with diagnostic accuracies of 97.5%, 95.7% and 100%, respectively (Fig. 2b and Supplementary Table 3).

To assess the diagnostic accuracy of the methylation marker panel, we then applied the methylation panel to 1/3 of TCGA validation cohort 1 and GEO validation cohort 2 (Fig. 2c and d). The diagnostic accuracy of the panel for LUAD, LUSC and normal tissues was 98.1%, 94.8% and 100%, respectively, in 1/3 of TCGA validation cohort 1 (Fig. 2c and Supplementary Table 4) and 86.2%, 87.5% and 98.6% in GEO validation cohort 2 (Fig. 2d and Supplementary Table 5), respectively. These results demonstrate the robust nature of the methylation panel in identifying the presence of malignancy and its NSCLC subtype classification.

## Cancer-specific methylation signature validation in tissues and CTCs

To validate the clinical value of the methylation panel, we next performed whole-genome bisulfite sequencing (WGBS) analysis of tumor tissue samples, matched normal lung tissue samples, CTCs and WBCs from 6 NSCLC patients, including 4 LUAD and 2 LUSC cases (Supplementary Table 1). To obtain CTCs, blood samples were drawn from the 6 patients with newly diagnosed lung cancer and processed with the immunostaining-FISH-based CTC technique [18–21], a CD45-based immunomagnetic system that combines leukocyte common antigen (CD45) immunostaining and fluorescence in situ hybridization

using unprocessed blood samples that is specifically adapted to achieve a capture rate of > 98% for single CTC and CTC clusters [18, 22]. Upon capture, fixed CTCs were stained with antibodies against CD45 to identify contaminating leukocytes (Supplementary Fig. 1). Upon staining verification, we identified CTCs in the six patients; the CTCs from each patient were individually captured and deposited in 10 μl lysis buffer for WGBS [18, 22]. The WGBS sequencing data comprised 30G raw data (10×) for the tissue samples and bulk WBCs and 90G raw data (30×) for the CTC samples. On average, we achieved 47.3% CpG coverage for CTCs, in line with a recent single-cell WGBS study [12]. For each individual methylation profile, only CpG sites ≥ 3× coverage were used for clustering and functional analyses (Supplementary Table 6).

To refine the tumor methylation signature in the matched CTCs, we identified similar methylation CpGs (SMCs) with methylation differences < 10% and P values > 0.05 between CTCs and the matched tumor tissue for each patient and included 450K CpG sites (Supplementary Table 7). Therefore, we obtained $CT_{450K}$ SMCs s present in both the CTC and tumor methylation profiles with similar methylation levels that were also included in the 450K CpG sites. Then, we merged the $CT_{450K}$ SMCs from the 6 patients and used these CpG sites as features to cluster matched normal and lung cancer tissues, WBCs and CTCs. After refinement of the $CT_{450K}$ methylation signatures, all 6 pairs of CTCs and matched tumor tissues clustered together, and the Pearson correlation coefficient of CTCs and tumor tissues for each patient was greater than 0.994 (Fig. 3b). These results suggest that CTCs inherit most of their methylation signatures from the primary tumors.

However, clustering of the $CT_{450K}$ methylation analysis did not group the CTCs and tumors together. The high Pearson correlation coefficient of CTCs and WBCs/Normal led us to hypothesize the existence of WBCs and normal tissue backgrounds in the CTC methylation pattern. To eliminate WBC backgrounds in the CTC methylation profile, we identified $CBT_{450K}$ DMCs for each patient (Supplementary Table 7). Next, we merged the $CBT_{450K}$ DMCs of the 6 patients together and used these merged DMCs to cluster the lung cancer tissues and matched normal tissues/WBCs/CTCs. The Pearson correlation coefficient of the CTCs and WBCs for each patient decreased from a minimum of 0.831 (Fig. 3b) to a maximum of -0.696 (Fig. 3c). The results showed that almost all 6 pairs of CTCs and matched tumors clustered together, except for the clustering of 2T and 2C due to ineffective bisulfite conversion of 2B (conversion ratio 91.46%). Figure 3c reveals that the WBC background is an important component of the CTC methylation pattern.

To further eliminate the normal tissue background from the CTC methylation profile, we identified $CBTN_{450K}$ DMCs in each patient (Supplementary Table 7) and then merged the $CBTN_{450K}$ DMCs of the 6 patients together and used these merged DMCs to cluster the matched normal and lung cancer tissues, WBCs and CTCs (Fig. 3d). The poor clustering performance suggested that a normal tissue background is not an effective component of the CTC methylation pattern.

Based on the above methylation profiles, our NSCLC tissue cohort showed a diagnostic accuracy of the methylation panel for all LUAD, LUSC and normal tissues of 100% (Supplementary Table 8); after

removing the matched WBC background, the diagnostic accuracy for LUAD and LUSC was also 100% based on the CTC cohort (Supplementary Table 9) (C&B diff > 0.1, P < 0.05).

# Inheritance of CTCs

To explore the characteristics of CTCs, we specifically investigated the methylation profile distribution according to functional genomic features between CTCs and matched tumor tissues (Fig. 4). We observed that the number of CpG sites in each CT similar group was far larger than that in the matched CT difference group (Fig. 4a), which implies that CTCs have a large proportion of methylation signatures inherited from the primary tumor. In addition, 5-methylcytosine (5-mC) was most common in transcription factor binding sites (TFBSs) and intronic and intergenic regions, accounting for 42.8%, 47.3%, and 46.0% of all CpG sites in tumors and 47.4%, 47.6%, and 42.8% of all CpG sites in CTCs, respectively (Fig. 4a). 5-mC was also commonly found in enhancers, superenhancers, promoters, and CpG shores, accounting for 5.6%, 24.8%, 5.5%, and 6.6% of all CpG sites in tumors and 6.1%, 28.6%, 9.1%, and 8.0% of all CpG sites in CTCs, respectively (Fig. 4a).

For regulatory elements, loss of DNA methylation at TFBSs can designate active transcription factor networks or networks primed for activation at later stages, e.g., during processes such as the derivation of induced pluripotent stem cells from differentiated cells [23] or cancer progression [9]. We then analyzed SMCs at TFBSs using i-Cistarget [17] and used the clusterProfile R package to analyze Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways of global CTC hypomethylated TFBSs coexisting in CTCs and matched tumor tissues with methylation differences < 10% and P values > 0.05 (Fig. 5a). Our DNA methylation analysis revealed the MAPK signaling pathway and pathways regulating the pluripotency of stem cells, EGFR tyrosine kinase inhibitor resistance, and the cell cycle, among others. These pathways coexisted in both CTCs and matched tumor tissues, suggesting that CTC methylation originates from primary tumor tissues and is inherited as the cells move from the primary tumor tissues to peripheral blood (Fig. 5a). We also found that binding sites for stemness-associated transcription factors are specifically hypomethylated in SMCs in both CTCs and matched tumor tissues, including binding sites for NANOG and SOX2, which in previous reports were associated with CTC clusters compared to single CTCs [24].

To gain explore more subtle changes in DNA methylation occurring specifically within promoters, gene bodies, and superenhancer regions, we carried out hypergeometric-based gene set enrichment analysis of genomic features. Consistently, this analysis revealed hypomethylation and cell cycle progression (Supplementary Fig. 2), as previously observed for cancer specimens with stem-like and proliferative features.

# Evolution of CTCs

Nevertheless, CTCs showed many DMCs that differed from those of the primary tumor (Fig. 4b). To identify whether the characteristic-related transcription factor networks are also transcriptionally active in CTCs compared to matched tumors, we performed single-cell resolution methylation sequencing analysis of CTCs and matched tumors isolated from the 6 NSCLC patients (Fig. 5). We used the cluster Profile R

package to analyze KEGG pathways of global hypomethylated TFBSs in CTCs rather than matched tumor tissues with absolute methylation difference > 0.1 and P value 0.05 (Fig. 5b). KEGG analysis of TFBS DMCs in patient CTCs compared to matched tumors revealed enrichment of genes related to the T cell receptor signaling pathway, Th17 cell differentiation, TGF-beta signaling pathway, EGFR tyrosine kinase inhibitor resistance and canonical pathways (RAS/MAPK/PI3K/AKT), among others (Fig. 5b).

To gain insight into more subtle changes in DNA methylation occurring specifically within promoters, gene bodies, and superenhancer regions, KEGG analysis revealed enrichment of gene groups related to the TGF-beta signaling pathway, Th17 cell differentiation, T cell receptor signaling pathway, and EGFR tyrosine kinase inhibitor resistance, among others (Supplementary Fig. 3).

These results show that in the processes of dissemination from the primary tumor tissue to the peripheral blood, CTCs gradually develop their own unique methylation signatures with some unique characteristics different from those of the primary tumor.

# Discussion

Our study provides an alternative noninvasive approach by using CTC methylation signatures instead of biopsy for pathological classification of NSCLC patients. We found that these methylation signatures can identify LUAD and LUSC with extremely high accuracy. Our results raise the possibility that the detection of CTC methylation in peripheral blood may be expanded to aid in the diagnosis of a much larger number of tumor types. In addition, we uncovered two CTC methylation patterns of inheritance and evolution during the CTC migration process. These features of CTCs provide new insight into the mechanism of NSCLC metastasis.

In practice, the amount of genomic DNA in CTCs per patient is rather limited, typically in the range of tens of picograms. Moreover, amplification of damaged DNA beginning with a fixed CTC cell is even more difficult than amplification of integrated genomic DNA beginning with a live cell. In our study, we collected all CTCs from each patient into one tube. Then, we amplified and achieved an average of 47.3% CpG coverage for CTCs, in line with the close to 50% CpG site coverage in a single cell reported by a recent study [12].

Our study suggests that CTCs share several properties common in immune escape and in mesenchymal-shifted cells compared to matched tumors. For instance, T cell receptor signaling pathway and Th17 cell differentiation contributes to tumor immune escape [25–33]. TGF-beta pathway promotes the epithelial-mesenchymal transition (EMT) in tumor cells, which plays an important role in mediating tumor invasion and metastasis [34, 35]. EGFR tyrosine kinase inhibitor resistance suggests a form of acquired drug resistance, which is associated with the tumor cell EMT phase [36–38]. The canonical RAS/MAPK/PI3K/AKT signaling pathway is involved in EMT progression [39]. Previous reports have demonstrated that enhancement of mesenchymal-like features epigenetically reprograms epithelial cancer cells to adapt well to new microenvironments and thus may contribute to distant metastasis [40]. Several reports have focused on the relationship between EMT and immune escape [39, 41–44],

especially in NSCLC. However, previous studies only focused on a few genes or proteins in CTCs associated with immune escape, such as upregulation of CD47 expression as a potential escape mechanism in colorectal cancer based on qPCR [45] or downregulation of ULBP1 protein expression as a potential CTC evasion mechanism from NK cells [46]. Our study uncovered a CTC immune escape mechanism through CTC methylation signatures on a genome-wide scale, and we propose that the EMT status of CTCs and T cell receptor signaling ultimately leads to tumor immune escape and invasion.

Methylation profiling of ctDNA has been investigated in cancer diagnostics and is assessing therapeutic outcomes [47–52], but few methylation profiles of CTCs have been studied to date. We used CTC methylation profiles to identify LUAD and LUSC with extremely high accuracy in 6 NSCLC patients. In our study, CTCs, as one of the three liquid biopsy biomarker types, showed strong potential in terms of methylation origin and classification. Compared to ctDNA and exosomes, CTCs carry a complete genome, which is an incomparable advantage. Our study demonstrated that CTC traceability only requires deducing the matched WBCs; in contrast, ctDNA traceability needs a large training set and complex algorithm due to its rarity in the blood. Nonetheless, as an auxiliary diagnosis tool for benign and malignant lesions, CTC techniques should be strengthened. On the one hand, we may utilize other technologies, such as microfluidic technology, to count and capture CTCs with both high sensitivity and specificity and low damage to the cells. On the other hand, we should explore more analytical methods and models to improve coverage or change the analysis units from methylated cytosines to methylated regions [24, 53].

# Conclusion

In summary, our study provides insight into the potential of CTCs to replace invasive biopsy for pathological classification of NSCLC patients., and CTC primary tumor inheritance and CTC evolution affect metastasis and immune escape.

# Abbreviations

WBC, white blood cell; TCGA, The Cancer Genome Atlas; GEO, Gene Expression Omnibus; WGBS, whole-genome bisulfite sequencing; DMCs, differentially methylated CpG sites; SMCs, similarly methylated CpG sites LUAD, lung adenocarcinoma; LUSC, lung squamous cell carcinoma; Normal, normal tissues; CTC, circulating tumor cell; KEGG, Kyoto Encyclopedia of Genes and Genomes.

# Declarations

### Ethics approval and consent to participate

Ethical approval was obtained from the Zhongshan Hospital Research Ethics Committee, and written informed consent was obtained from each patient.

### Consent for publication

Not applicable.

### Availability of data and materials

All data generated or analyzed during this study are included either in this article or in the supplementary information files.

### *Competing interests*

The authors declare no conflicts of interest.

### *Funding*

This work was supported by the National Natural Science Foundation of China (81972168) and National Key R&D Program of China (2016YFA0502202, H.W.).

### Authors' contributions

JHJ, JYD and YFY, conceived and designed the experiments; JHJ, CYC and JG performed the experiments; JL, YL and WF analyzed the data; JHJ and CYC

wrote the paper. All authors read and approved the final manuscript.

### Acknowledgements

Not applicable.

# References

1. Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jemal A. Global cancer statistics, 2012. CA Cancer J Clin. 2015;65:87–108.
2. Reck M, Rabe KF. Precision diagnosis and treatment for advanced non-small-cell lung cancer. N Engl J Med. 2017;377:849–61.
3. Zhang Z, Xiao Y, Zhao J, Chen M, Xu Y, Zhong W, et al. Relationship between circulating tumour cell count and prognosis following chemotherapy in patients with advanced non-small-cell lung cancer. Respirology. 2016;21:519–25.
4. Matthew EM, Zhou L, Yang Z, Dicker DT, Holder SL, Lim B, et al. A multiplexed marker-based algorithm for diagnosis of carcinoma of unknown primary using circulating tumor cells. Oncotarget. 2016;7:3662–76.
5. Vidal E, Sayols S, Moran S, Guillaumet-Adkins A, Schroeder MP, Royo R, et al. A DNA methylation map of human cancer at single base-pair resolution. Oncogene. 2017;36:5648–57.
6. Ehrlich M. DNA methylation in cancer: too much, but also too little. Oncogene. 2002;21:5400–13.
7. Ehrlich M. DNA hypomethylation in cancer cells. Epigenomics. 2009;1:239–59.

8. Farlik M, Halbritter F, Muller F, Choudry FA, Ebert P, Klughammer J, et al. DNA methylation dynamics of human hematopoietic stem cell differentiation. Cell Stem Cell. 2016;19:808–22.
9. Feinberg AP, Vogelstein B. Hypomethylation distinguishes genes of some human cancers from their normal counterparts. Nature. 1983;301:89–92.
10. Fernandez AF, Assenov Y, Martin-Subero JI, Balint B, Siebert R, Taniguchi H, et al. A DNA methylation fingerprint of 1628 human samples. Genome Res. 2012;22:407–19.
11. Moran S, Martinez-Cardus A, Sayols S, Musulen E, Balana C, Estival-Gonzalez A, et al. Epigenetic profiling to classify cancer of unknown primary: a multicentre, retrospective analysis. Lancet Oncol. 2016;17:1386–95.
12. Clark SJ, Smallwood SA, Lee HJ, Krueger F, Reik W, Kelsey G. 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. Krueger F, Andrews SR. Bismark: a flexible aligner and methylation caller for bisulfite-seq applications. Bioinformatics. 2011;27:1571–2.
14. Langmead B, Salzberg SL. Fast gapped-read alignment with bowtie 2. Nat Methods. 2012;9:357–9.
15. Akalin A, Kormaksson M, Li S, Garrett-Bakelman FE, Figueroa ME, Melnick A, et al. methylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles. Genome Biol. 2012;13:R87.
16. Wang HQ, Tuominen LK, Tsai CJ. SLIM: a sliding linear model for estimating the proportion of true null hypotheses in datasets with dependence structures. Bioinformatics. 2011;27:225–31.
17. Herrmann C, Van de Sande B, Potier D, Aerts S. i-cisTarget: an integrative genomics method for the prediction of regulatory features and cis-regulatory modules. Nucleic Acids Res. 2012;40:e114.
18. Wan JF, Li XQ, Zhang J, Yang LF, Zhu J, Li GC, et al. Aneuploidy of chromosome 8 and mutation of circulating tumor cells predict pathologic complete response in the treatment of locally advanced rectal cancer. Oncol Lett. 2018;16:1863–8.
19. Lin PP, Gires O, Wang DD, Li L, Wang H. Comprehensive *in situ* co-detection of aneuploid circulating endothelial and tumor cells. Sci Rep. 2017;7:9789.
20. Lin PP. Erratum to: integrated EpCAM-independent subtraction enrichment and iFISH strategies to detect and classify disseminated and circulating tumors cells. Clin Transl Med. 2016;5:6.
21. Lin PP. Integrated EpCAM-independent subtraction enrichment and iFISH strategies to detect and classify disseminated and circulating tumors cells. Clin Transl Med. 2015;4:38.
22. Chen C, Li J, Wan J, Lu Y, Zhang Z, Xu Z. A low cost and input tailing method of quality control on multiple annealing, and looping-based amplification cycles-based whole-genome amplification products. J Clin Lab Anal. 2019;33:e22697.
23. Lee DS, Shin JY, Tonge PD, Puri MC, Lee S, Park H, et al. An epigenomic roadmap to induced pluripotency reveals DNA methylation as a reprogramming modulator. Nat Commun. 2014;5:5619.

24. Gkountela S, Castro-Giner F, Szczerba BM, Vetter M, Landin J, Scherrer R, et al. Circulating tumor cell clustering shapes DNA methylation to enable metastasis seeding. Cell. 2019;176:98–112.e14.
25. Han Y, Ye A, Bi L, Wu J, Yu K, Zhang S. Th17 cells and interleukin-17 increase with poor prognosis in patients with acute myeloid leukemia. Cancer Sci. 2014;105:933–42.
26. Kim JM, Chen DS. Immune escape to PD-L1/PD-1 blockade: seven steps to success (or failure). Ann Oncol. 2016;27:1492–504.
27. Grywalska E, Pasiarski M, Gozdz S, Rolinski J. Immune-checkpoint inhibitors for combating T-cell dysfunction in cancer. Onco Targets Ther. 2018;11:6505–24.
28. He X, Xu C. Immune checkpoint signaling and cancer immunotherapy. Cell Res. 2020;30:660–9.
29. Price DA, West SM, Betts MR, Ruff LE, Brenchley JM, Ambrozak DR, et al. T cell receptor recognition motifs govern immune escape patterns in acute SIV infection. Immunity. 2004;21:793–803.
30. Qin A, Coffey DG, Warren EH, Ramnath N. Mechanisms of immune evasion and current status of checkpoint inhibitors in non-small cell lung cancer. Cancer Med. 2016;5:2567–78.
31. Reuben A, Zhang J, Chiou SH, Gittelman RM, Li J, Lee WC, et al. Comprehensive T cell repertoire characterization of non-small cell lung cancer. Nat Commun. 2020;11:603.
32. Spranger S. Mechanisms of tumor escape in the context of the T-cell-inflamed and the non-T-cell-inflamed tumor microenvironment. Int Immunol. 2016;28:383–91.
33. Tian Y, Zhai X, Han A, Zhu H, Yu J. Potential immune escape mechanisms underlying the distinct clinical outcome of immune checkpoint blockades in small cell lung cancer. J Hematol Oncol. 2019;12:67.
34. Colak S, Ten Dijke P. Targeting TGF-beta signaling in cancer. Trends Cancer. 2017;3:56–71.
35. Drabsch Y, Ten Dijke P. TGF-beta signalling and its role in cancer progression and metastasis. Cancer Metastasis Rev. 2012;31:553–68.
36. Gainor JF, Dardaei L, Yoda S, Friboulet L, Leshchiner I, Katayama R, et al. Molecular mechanisms of resistance to first- and second-generation ALK inhibitors in ALK-rearranged lung cancer. Cancer Discov. 2016;6:1118–33.
37. Liu X, Li J, Cadilha BL, Markota A, Voigt C, Huang Z, et al. Epithelial-type systemic breast carcinoma cells with a restricted mesenchymal transition are a major source of metastasis. Sci Adv. 2019;5:eaav4275.
38. Fischer KR, Durrans A, Lee S, Sheng J, Li F, Wong ST, et al. Epithelial-to-mesenchymal transition is not required for lung metastasis but contributes to chemoresistance. Nature. 2015;527:472–6.
39. Jiang Y, Zhan H. Communication between EMT and PD-L1 signaling: new insights into tumor immune evasion. Cancer Lett. 2020;468:72–81.
40. Mitra A, Mishra L, Li S. EMT. CTCs and CSCs in tumor relapse and drug-resistance. Oncotarget. 2015;6:10697–711.
41. Mak MP, Tong P, Diao L, Cardnell RJ, Gibbons DL, William WN, et al. A patient-derived, pan-cancer EMT signature identifies global molecular alterations and immune target enrichment following

epithelial-to-mesenchymal transition. Clin Cancer Res. 2016;22:609–20.
42. Lou Y, Diao L, Cuentas ER, Denning WL, Chen L, Fan YH, et al. Epithelial-mesenchymal transition is associated with a distinct tumor microenvironment including elevation of inflammatory signals and multiple immune checkpoints in lung adenocarcinoma. Clin Cancer Res. 2016;22:3630–42.
43. Kim S, Koh J, Kim MY, Kwon D, Go H, Kim YA, et al. PD-L1 expression is associated with epithelial-to-mesenchymal transition in adenocarcinoma of the lung. Hum Pathol. 2016;58:7–14.
44. Tiwari N, Gheldof A, Tatari M, Christofori G. EMT as the ultimate survival mechanism of cancer cells. Semin Cancer Biol. 2012;22:194–207.
45. Steinert G, Scholch S, Niemietz T, Iwata N, Garcia SA, Behrens B, et al. Immune escape and survival mechanisms in circulating tumor cells of colorectal cancer. Cancer Res. 2014;74:1694–704.
46. Hu B, Tian X, Li Y, Liu Y, Yang T, Han Z, et al. Epithelial-mesenchymal transition may be involved in the immune evasion of circulating gastric tumor cells via downregulation of ULBP1. Cancer Med. 2020;9:2686–97.
47. Taylor WC. Comment on 'sensitive and specific multi-cancer detection and localization using methylation signatures in cell-free DNA' by M. C. Liu et al. Ann Oncol. 2020;31:1266–7.
48. Liu L, Toung JM, Jassowicz AF, Vijayaraghavan R, Kang H, Zhang R, et al. Targeted methylation sequencing of plasma cell-free DNA for cancer detection and classification. Ann Oncol. 2018;29:1445–53.
49. Jiang P, Sun K, Tong YK, Cheng SH, Cheng THT, Heung MMS, et al. Preferred end coordinates and somatic variants as signatures of circulating tumor DNA associated with hepatocellular carcinoma. Proc Natl Acad Sci U S A. 2018;115:E10925-33.
50. Xu RH, Wei W, Krawczyk M, Wang W, Luo H, Flagg K, et al. Circulating tumour DNA methylation markers for diagnosis and prognosis of hepatocellular carcinoma. Nat Mater. 2017;16:1155–61.
51. Guo S, Diep D, Plongthongkum N, Fung HL, Zhang K, Zhang K. 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.
52. Lee EJ, Luo J, Wilson JM, Shi H. Analyzing the cancer methylome through targeted bisulfite sequencing. Cancer Lett. 2013;340:171–8.
53. Hao X, Luo H, Krawczyk M, Wei W, Wang W, Wang J, et al. DNA methylation markers for diagnosis and prognosis of common cancers. Proc Natl Acad Sci U S A. 2017;114:7414–9.

# Supplementary Tables

**Supplementary Table 1**. Information for the patients enrolled in this study

| Patient ID | Gender | Pathological Diagnosis | Captured CTCs | Tumor Size (cm) | EGFR Mutation | TNM Stage (AJCC 8th) |
|---|---|---|---|---|---|---|
| 1 | Female | lung adenocarcinoma | 13 | 1.0 | 19del | IA |
| 2 | Male | lung squamous cell carcinoma | 7 | 2.5 | none | IA |
| 3 | Male | lung adenocarcinoma | 40 | 2.0 | none | IA |
| 4 | Female | lung adenocarcinoma | 12 | 1.5 | none | IA |
| 5 | Male | lung squamous cell carcinoma | 10 | 3.0 | none | IA |
| 6 | Female | lung adenocarcinoma | 5 | 2.5 | 19del | IA |

**Supplementary Table 3**. Confusion table of training cohort

| | LUAD | LUSC | Normal lung | Total |
|---|---|---|---|---|
| **LUAD** | 310 | 7 | 0 | 317 |
| **LUSC** | 5 | 225 | 0 | 230 |
| **Normal lung** | 3 | 3 | 60 | 66 |
| **Total** | 318 | 235 | 60 | 613 |
| **Correct** | 310 | 225 | 60 | 595 |
| **Correct(%)** | 97.48428 | 95.74468 | 100 | 97.06362 |

**Supplementary Table 4**. Confusion table of validation cohort 1

| | LUAD | LUSC | Normal lung | Total |
|---|---|---|---|---|
| **LUAD** | 151 | 5 | 0 | 156 |
| **LUSC** | 2 | 127 | 0 | 129 |
| **Normal lung** | 1 | 2 | 14 | 17 |
| **Total** | 154 | 134 | 14 | 302 |
| **Correct** | 151 | 127 | 14 | 292 |
| **Correct(%)** | 98.05195 | 94.77612 | 100 | 96.68874 |

**Supplementary Table 5**. Confusion table of validation cohort 2

| | LUAD | LUSC | Normal lung | Total |
|---|---|---|---|---|
| **LUAD** | 25 | 1 | 1 | 27 |
| **LUSC** | 2 | 7 | 0 | 9 |
| **Normal lung** | 2 | 0 | 73 | 75 |
| **Total** | 29 | 8 | 74 | 111 |
| **Correct** | 25 | 7 | 73 | 105 |
| **Correct(%)** | 86.2069 | 87.5 | 98.64865 | 94.59459 |

**Supplementary Table 6**. Summary of the basic sequencing parameters, including the sequencing depth, for all six patients

| Case no. | Raw_bases | Conversion rate (%) | Mappability (%) | Duplication rate (%) | Sequence depth | 1xCpG coverage |
|---|---|---|---|---|---|---|
| 01B | 34023049106 | 99.51% | 88.62% | 16.50% | 7.001361 | 95.608 |
| 02B | 31388659920 | 91.46% | 29.14% | 52.24% | 1.156164 | 44.667 |
| 03B | 33872628040 | 99.77% | 87.78% | 18.54% | 6.717153 | 94.709 |
| 04B | 31534218786 | 99.73% | 88.77% | 16.83% | 6.30206 | 93.076 |
| 05B | 34806104906 | 99.75% | 87.05% | 16.96% | 6.863009 | 94.18 |
| 06B | 33407950002 | 99.72% | 88.71% | 22.55% | 6.349538 | 93.927 |
| 01N | 40379812376 | 99.68% | 87.23% | 21.45% | 7.645562 | 93.645 |
| 02N | 1436336462 | 98.76% | 81.85% | 32.54% | 0.202227 | 14.748 |
| 03N | 34455483812 | 99.69% | 85.31% | 17.22% | 6.717548 | 92.865 |
| 04N | 37161722120 | 99.67% | 87.90% | 18.91% | 7.343653 | 93.019 |
| 05N | 33460825068 | 99.68% | 87.93% | 21.77% | 6.38553 | 91.777 |
| 06N | 36764933682 | 99.68% | 87.54% | 19.69% | 7.145472 | 93.673 |
| 01T | 34563878558 | 99.65% | 87.21% | 19.57% | 6.687883 | 92.707 |
| 02T | 31370116818 | 99.51% | 76.32% | 15.80% | 5.738217 | 92.61 |
| 03T | 34458690146 | 99.72% | 80.61% | 18.77% | 6.189226 | 90.995 |
| 04T | 36849359292 | 99.61% | 88.02% | 19.81% | 7.190256 | 92.878 |
| 05T | 34271216700 | 99.71% | 86.99% | 22.66% | 6.35076 | 91.395 |
| 06T | 38044759550 | 99.68% | 86.86% | 19.38% | 7.373278 | 93.425 |
| 01C | 90723626720 | 98.99% | 62.06% | 66.06% | 3.856598 | 77.455 |
| 02C | 91474230506 | 98.00% | 59.38% | 86.49% | 1.408053 | 45.289 |
| 03C | 97534784928 | 98.98% | 71.55% | 88.42% | 1.777489 | 45.97 |
| 04C | 1.00386E+11 | 98.88% | 71.49% | 91.78% | 1.254073 | 26.617 |
| 05C | 1.0637E+11 | 98.94% | 71.95% | 94.55% | 0.863636 | 16.98 |
| 06C | 93647437304 | 99.20% | 67.79% | 79.47% | 2.988654 | 71.377 |

**Supplementary Table 7**. Parameter conditions for the clustering of cancer tissues, normal tissues, WBCs and CTCs

| Optimized Condition | CT-similar | CB-diff | CN-diff |
|---|---|---|---|
| $Raw_{450K}$ | - | - | - |
| $CT_{450K}$ | P>0.05;diff<10% | - | - |
| $CBT_{450K}$ | P>0.05;diff<10% | P<0.05;diff>0.15 | - |
| $CBNT_{450K}$ | P>0.05;diff<10% | P<0.05;diff>0.15 | P<0.05;diff>0.15 |

Note: diff, difference.

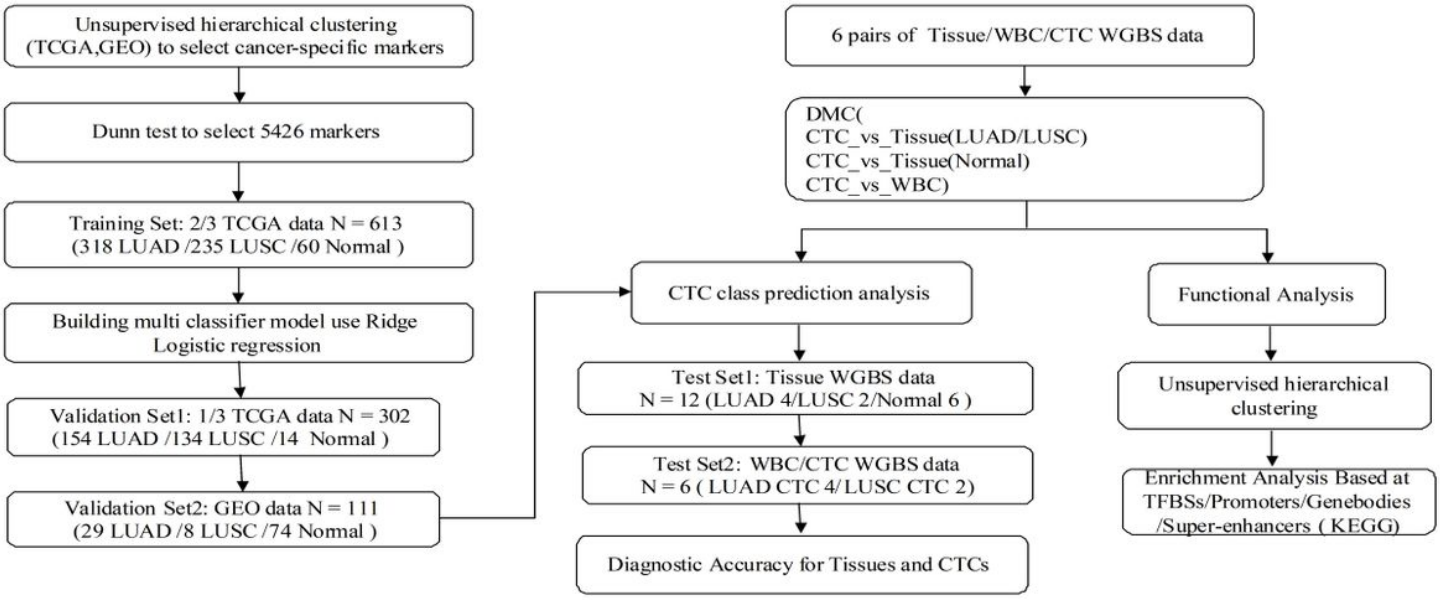
**Supplementary Table 8**. Confusion table of our NSCLC tissue cohort

| | LUAD | LUSC | Normal lung | Total |
|---|---|---|---|---|
| **LUAD** | 4 | 0 | 0 | |
| **LUSC** | 0 | 2 | 0 | |
| **Normal lung** | 0 | 0 | 6 | |
| **Total** | 4 | 2 | 6 | 12 |
| **Correct** | 4 | 2 | 6 | 12 |
| **Correct(%)** | 100 | 100 | 100 | 100 |

**Supplementary Table 9**. Confusion table of the CTC validation cohort, including both C&B diff (C&B diff > 0.1, P < 0.05) and 5426 methylation markers

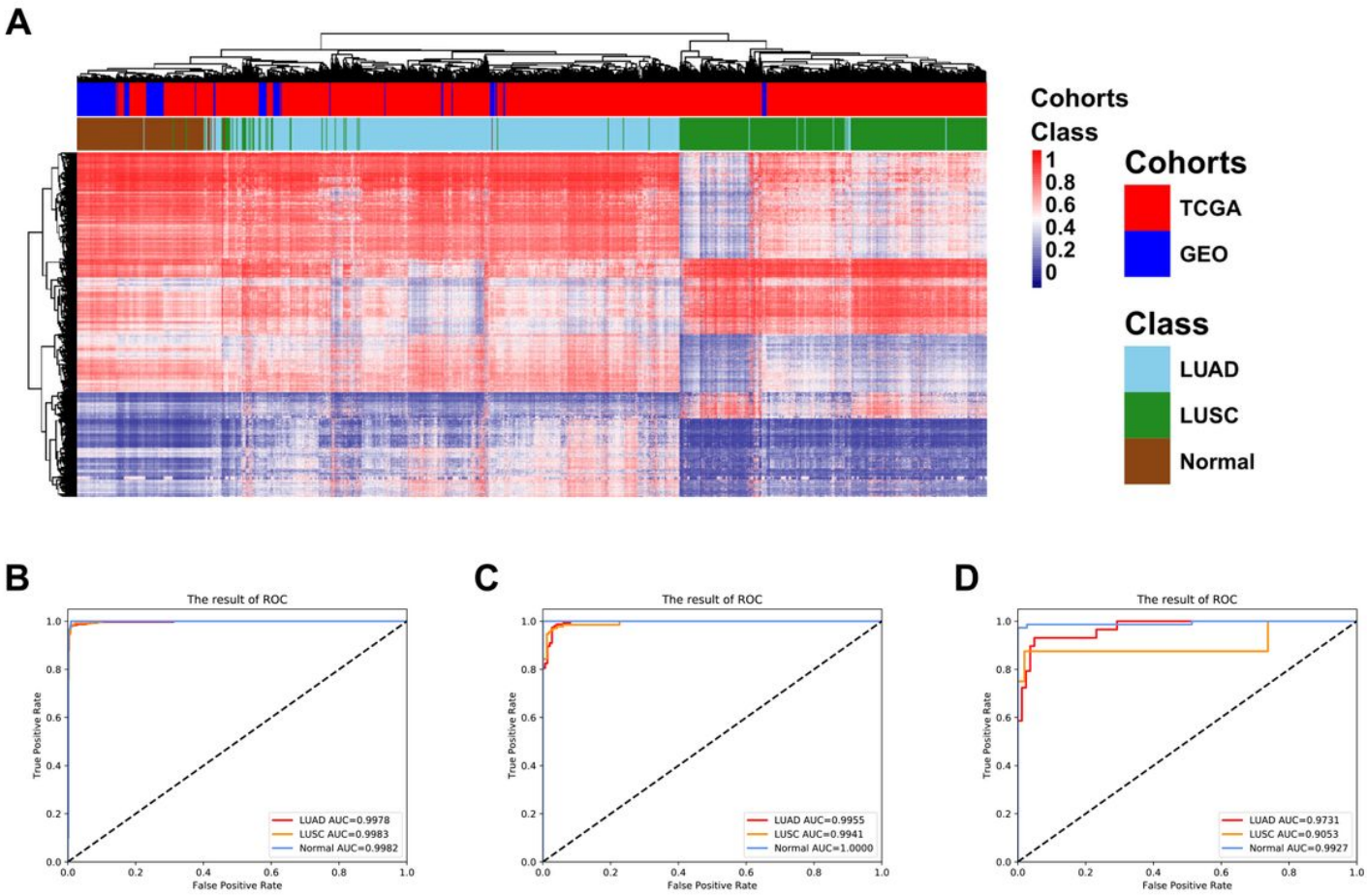
| B | LUAD | LUSC | Total |
|---|---|---|---|
| **LUAD** | 4 | 0 | |
| **LUSC** | 0 | 2 | |
| **Normal lung** | 0 | 0 | |
| **Total** | 4 | 2 | 12 |
| **Correct** | 4 | 2 | 12 |
| **Correct(%)** | 100 | 100 | 100 |

# Figures



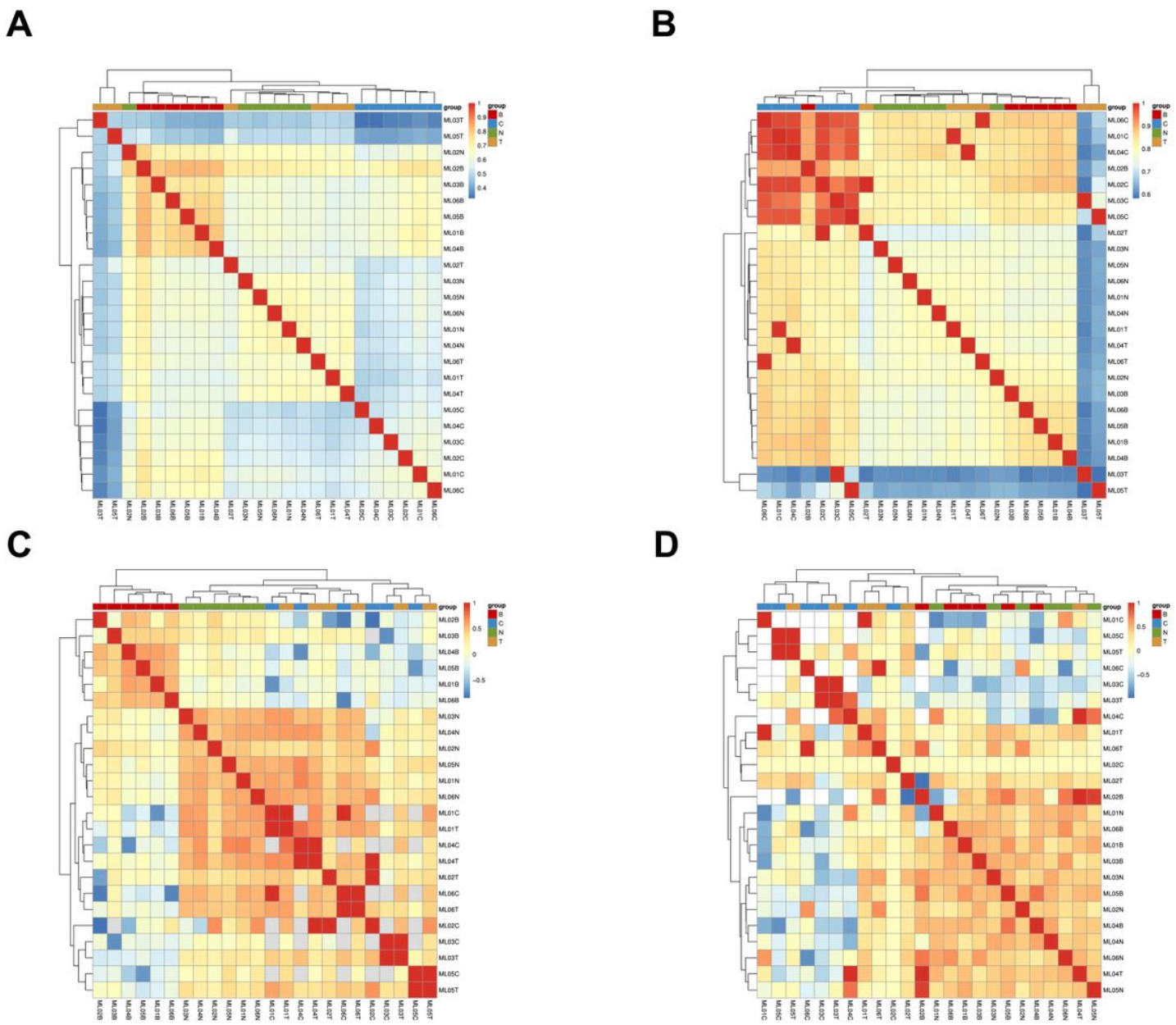

## Figure 1

Workflow chart of data generation and analysis. TCGA methylation data was used to cluster and identify 5426 features. Building a multiclassification model based on the above selected feature. A total of 2/3 of the TCGA data were used for training; 1/3 of the TCGA and the GEO data were used for validation. Patient-WGBS data were used for prediction model testing and functional analysis. WBC, white blood cell; TCGA, The Cancer Genome Atlas; GEO, Gene Expression Omnibus; WGBS, whole-genome bisulfite sequencing; DMCs, differentially methylated CpG sites; LUAD, lung adenocarcinoma; LUSC, lung squamous cell carcinoma; Normal, normal tissues; CTC, circulating tumor cell; KEGG, Kyoto Encyclopedia of Genes and Genomes.



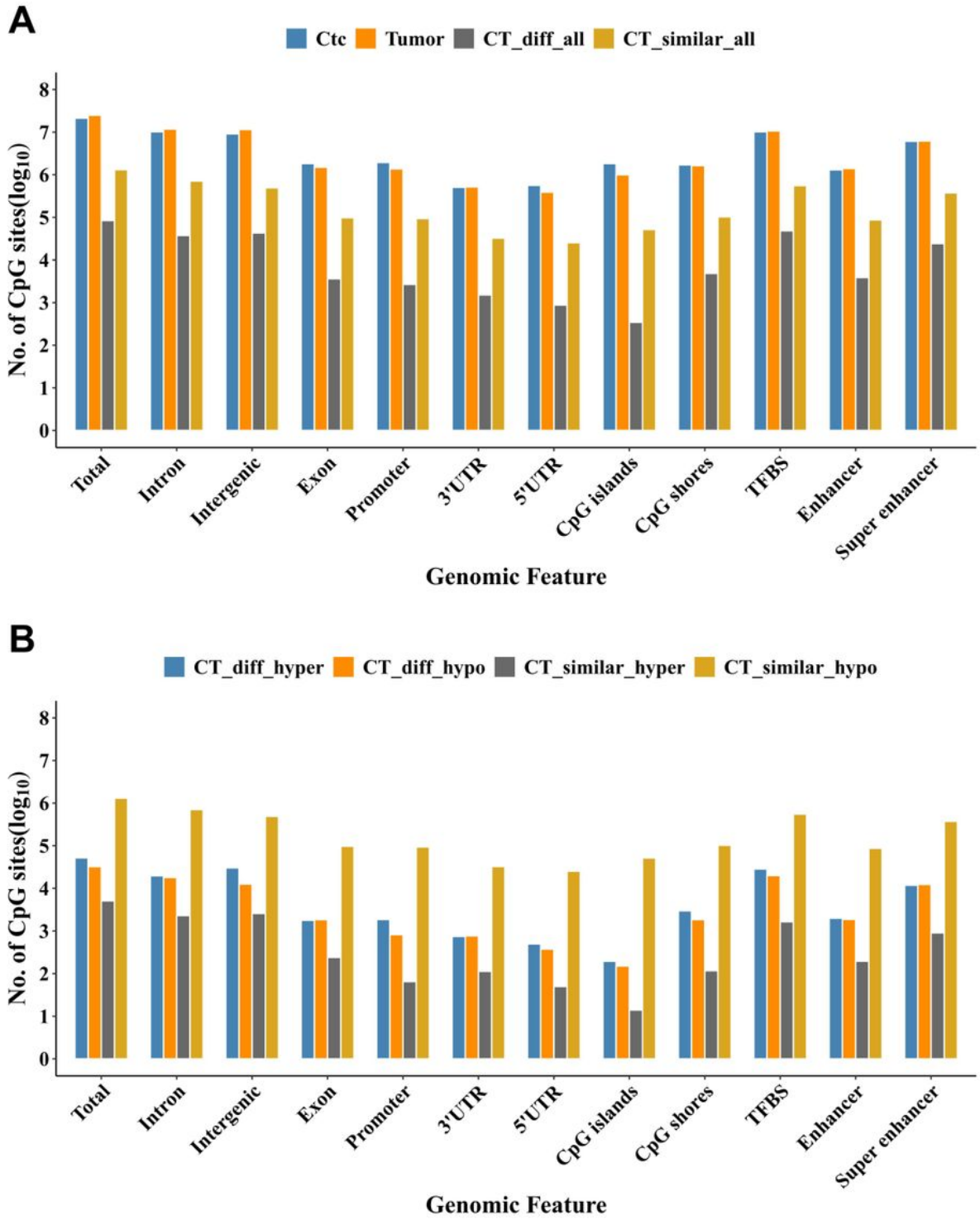

## Figure 2

Clustering and ROC analyses of discovery and validation sets using 5426 CpG markers identified in TCGA cohort. (a) DNA methylation signatures can identify LUAD, LUSC and NORMAL in TCGA and GEO cohorts. Shown are unsupervised hierarchical clustering and heat maps associated with the methylation profile of 501 LUAD samples (sky-blue) and 377 LUSC samples (green) and 148 normal samples (brown) in TCGA (red) and GEO (blue) cohorts with a panel of 5426 CpG markers. Each column represents an individual patient, and each row represents an individual CpG marker. The color scale shows DNA methylation level. (b) ROC curve of the diagnostic prediction model with methylation markers in 2/3 of the TCGA training cohort; (c) 1/3 of the TCGA validation cohort 1; (d) Gene Expression Omnibus (GEO) validation cohort 2.




## Figure 3

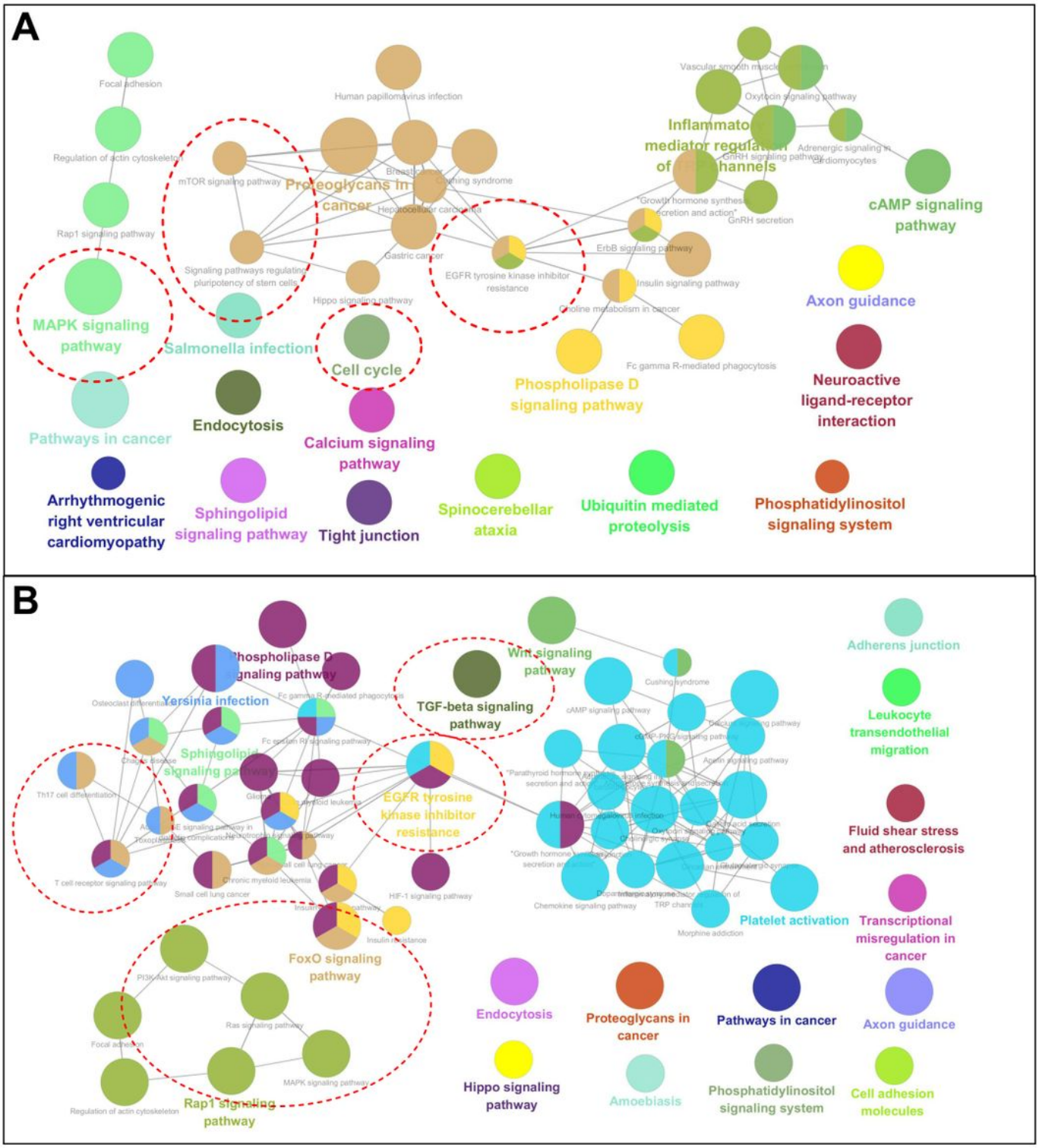
Unsupervised hierarchical clustering associated with the methylation profile (a) included in the 450K methylation array (according to the color scale shown) in cancer tissue (T), normal tissue (N), WBC bulk (b) and CTC (c) data for 6 patients. (b) Included in CT similar methylation markers selected for use in the 24 samples from the 6 patients. (c) Included in CBT methylation markers selected for use in the 24 samples from the 6 patients. (d) Included in CBNT methylation markers selected for use in the 24 samples from the 6 patients. The color scale shows Pearson correlation coefficients.




## Figure 4

CpG sites with known genomic features in 6 patients. CpG sites with no less than 3× coverage were counted, as shown in this figure. Similar CpG sites (CT similar) corresponding to genomic features with a normal P value >0.05 and a methylation difference <10% were counted between CTCs and matched tumor tissues. Differential CpG sites (CT diff) corresponding to genomic features with a normal P value

<0.05 and an absolute methylation difference >0.1 were counted in CTCs compared to matched tumor tissues. DMCs that appeared in the 2 of the 6 patients are illustrated in this figure.




## Figure 5

Pathway enrichment analysis of TFBSs on a genome-wide scale identified using i-cisTarget in hypomethylated regions of CT-similar (a) and CT-differential (b). CT-similar displayed a <10% methylation difference (P value > 0.05) in CTCs compared to matched tumor tissues among the 6 patients. CT-

differential displayed a >0.1 methylation difference (P value < 0.05) in CTCs compared to matched tumor tissues among the 6 patients. Gene sets with an adjusted P value< 0.01 were considered significant.

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryFigure1.jpg
- SupplementaryFigure2.jpg
- SupplementaryFigure3.jpg
- SupplementaryTable2.xlsx