



## Transcriptomic profiling of single circulating tumor cells provides insight into human metastatic gastric cancer

Ryo Negishi<sup>1</sup>, Hitomi Yamakawa<sup>1</sup>, Takeru Kobayashi<sup>1</sup>, Mayuko Horikawa<sup>1</sup>, Tatsu Shimoyama<sup>2</sup>, Fumiaki Koizumi<sup>3</sup>, Takeshi Sawada<sup>2</sup>, Keisuke Oboki<sup>4</sup>, Yasushi Omuro<sup>2</sup>, Chikako Funasaka<sup>2</sup>, Akihiko Kageyama<sup>2</sup>, Yusuke Kanemasa<sup>2</sup>, Tsuyoshi Tanaka <sup>1</sup>, Tadashi Matsunaga<sup>1</sup> & Tomoko Yoshino <sup>1</sup>✉

Transcriptome analysis of circulating tumor cells (CTCs), which migrate into blood vessels from primary tumor tissues, at the single-cell level offers critical insights into the biology of metastasis and contributes to drug discovery. However, transcriptome analysis of single CTCs has only been reported for a limited number of cancer types, such as multiple myeloma, breast, hepatocellular, and prostate cancer. Herein, we report the transcriptome analysis of gastric cancer single-CTCs. We utilized an antigen-independent strategy for CTC isolation from metastatic gastric cancer patients involving a size-dependent recovery of CTCs and a single cell isolation technique. The transcriptomic profile of single-CTCs revealed that a majority of gastric CTCs had undergone epithelial-mesenchymal transition (EMT), and indicated the contribution of platelet adhesion toward EMT progression and acquisition of chemoresistance. Taken together, this study serves to employ CTC characterization to elucidate the mechanisms of chemoresistance and metastasis in gastric cancer.

<sup>1</sup>Division of Biotechnology and Life science, Institute of Engineering, Tokyo University of Agriculture and Technology, 2-24-16, Naka-cho, Koganei, Tokyo 184-8588, Japan. <sup>2</sup>Department of Medical Oncology, Tokyo Metropolitan Cancer and Infectious Diseases Center Komagome Hospital, Tokyo, Japan. <sup>3</sup>Department of Laboratory Medicine, Tokyo Metropolitan Cancer and Infectious Diseases Center Komagome Hospital, Tokyo, Japan. <sup>4</sup>Center for Medical Research Cooperation, Tokyo Metropolitan Institute of Medical Science, Setagaya-ku, Japan. ✉email: [y-tomoko@cc.tuat.ac.jp](mailto:y-tomoko@cc.tuat.ac.jp)

Circulating tumor cells (CTCs) are cells that migrate into blood vessels from the primary tumor site and are involved in the metastasis of cancer to distant tissues. CTCs are the main targets of liquid biopsy along with circulating tumor DNA (ctDNA) and exosomes<sup>1,2</sup>. Unlike ctDNA and exosomes, CTCs have an advantage in that they directly provide information on the metabolism and signaling of cancer cells. CTCs are used for genome-wide mutation<sup>3</sup> and transcriptome analysis<sup>4</sup>, as well as cell culture<sup>5,6</sup>, and could be a non-invasive source of a vast amount of cellular information on tumor composition, invasiveness, drug susceptibility, and resistance to therapy. CTCs are expected to be used not only for cancer diagnosis but also for the development of therapeutic drugs<sup>7</sup>.

Recently, transcriptome analysis of CTCs identified candidate genes associated with metastasis and signal pathways associated with resistance to chemotherapy<sup>8,9</sup>. Transcriptome analysis of CTCs at the single-cell level has shown that CTCs are a heterogeneous population between and within patient groups. This may be due to cells undergoing epithelial-mesenchymal transition (EMT); both epithelial and mesenchymal cell types have been identified in CTCs<sup>10,11</sup>. It has also been shown that EMT-transformed CTCs have concurrent cancer stem cell-like properties<sup>12</sup>. Transcriptome analysis of CTCs at a single-cell level is considered to be a powerful tool for obtaining critical insights, such as response to treatment, that cannot be acquired by analyzing cancer tissues or by in vitro experiments<sup>7</sup>. In contrast, transcriptome analysis of single CTCs has only been reported in small-scale clinical trials for a limited number of cancer types, such as breast<sup>10,13–15</sup>, hepatocellular<sup>16,17</sup>, and prostate cancer<sup>7,18,19</sup> and multiple myeloma<sup>20,21</sup>, and there is a need to expand this investigation to a variety of cancer types.

Gastric cancer is the third leading cause of death from cancer worldwide<sup>22</sup>, with a majority of patients concentrated in the Asia-Pacific region and developing countries. In particular, metastatic gastric cancer is known as a very diverse cancer type with a 5-year survival rate of <10%<sup>23</sup>. Gastric cancer has a high recurrence rate after resection<sup>24</sup>, and most cases of recurrence and metastasis are refractory to chemotherapy<sup>25</sup>. The CellSearch system, using EpCAM antibodies, is the gold standard for measuring CTCs in other cancer types; it has been reported to be effective in predicting prognosis and risk of recurrence for gastric cancer<sup>26,27</sup>. Notably, the number of EpCAM-positive CTCs is lower in gastric cancer with a lower cut-off value compared to other cancer types<sup>28</sup>. It has been noted that as a result of EMT, it is difficult to detect gastric cancer CTCs with mesenchymal phenotypes<sup>28,29</sup>. Recently, RNA-ISH analysis of gastric CTCs showed the presence of mesenchymal CTCs and revealed their potential for monitoring therapy response<sup>30</sup>. However, the study defined mesenchymal CTCs using only two genes (*vimentin* and *twist*) and underlying mechanism of EMT in gastric CTCs were still unknown. Till date, there have been no reports on the transcriptome analysis of gastric cancer CTCs as described above. Therefore, a single-CTC transcriptome analysis, including EMT-induced CTCs, is essential to elucidate the properties of gastric CTCs.

In this study, we aimed to characterize gastric cancer CTCs by single-cell isolation and transcriptome analysis of CTCs from cases of metastatic gastric cancer. Our research group has established the microcavity array (MCA)/gel-based cell manipulation (GCM) method, a size-dependent recovery of CTCs using microfilters and single cell isolation technique, which can be applied to single-cell genome analysis<sup>31,32</sup>. Recent commercialized single-cell isolation system (Droplet-based system, nanowell-based system and fluorescence activated cell sorting) is focused to isolate single-cells from cell population by random; therefore, it is difficult to isolate rare single cell such as CTCs. Some systems such as VyCap Puncher<sup>33</sup>, DEPArray<sup>34</sup>, ALS CellCollector<sup>35</sup>, and RareCytes CyteFinder<sup>36</sup>

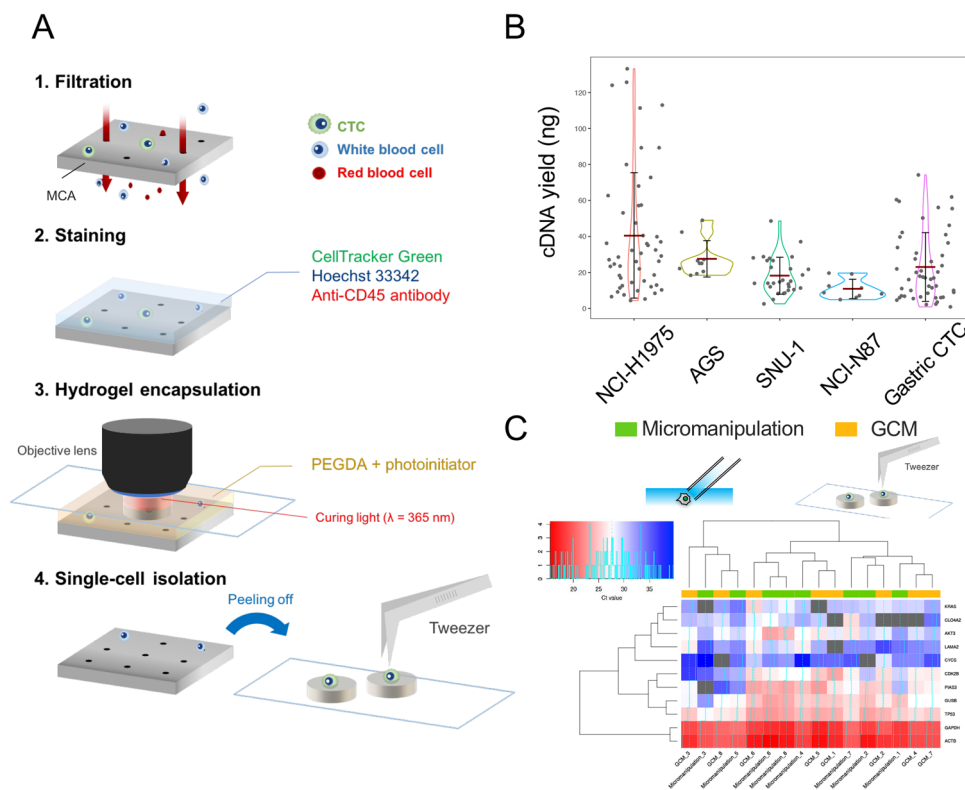
system have been developed for isolation of rare single CTCs. However, these systems are not integrated with the CTC enrichment system, necessitating CTC enrichment through other systems. This step may pose a risk of cell loss during sample handling. Although the RareCytes CytoFinder system was designed to perform the complete process of CTC isolation (CTC enrichment, staining, detection, and single-cell isolation), there is a risk of contamination of white blood cells because enriched cells were positioned at random. The MCA/GCM system was designed to perform the complete process of CTC isolation. Moreover, MCA can entrap single cells at even intervals; this allows isolation of single cells at high efficiency and without any contamination<sup>32</sup>. In this study, we separated gastric cancer CTCs by the surface antigen-independent MCA/GCM method and constructed a pipeline that enables RNA-seq analysis of single CTCs. Furthermore, our proposed method revealed the heterogeneity of the transcriptome of gastric cancer CTCs and characterized the gene expression pattern of EMT-induced CTCs that could reveal the underlying mechanism of EMT regulation in gastric cancer. The results of this study will also help to elucidate the mechanisms involved in the acquisition of drug resistance and metastasis of CTCs.

## Results

**Performance evaluation of the GCM-based method in single-cell whole transcriptome amplification (WTA).** Figure 1A showed schematic images of the GCM-based single-cell isolation method<sup>32</sup>. CTCs on the MCA were processed by RNA stabilization solution (CellCover) and stained with CellTracker Green and fluorescence-labelled antibody for CD45. Then, photopolymerizable hydrogel prepolymer was mounted and single cells were encapsulated into pillar-shaped hydrogels by light irradiation under the fluorescence microscope. Hydrogels were recovered by tweezers and subjected to subsequent single-cell WTA.

Prior to clinical trials, WTA of hydrogel-encapsulated single cells was performed using gastric cancer cell lines (SNU-1, NCI-N87, and AGS) and lung cancer cell line (NCI-H1975) to evaluate the performance of the MCA/GCM-based method. Although the cDNA yield varied depending on the cell lines, sufficient yield of cDNA from all cell lines were obtained for subsequent RNA-seq analysis (>10 ng/assay) (Fig. 1B). The average length of cDNA ranged from 823 to 1360 bp (Supplementary Fig. 1), indicating that sufficient cDNA length for gene expression analysis was obtained<sup>37</sup>. The amount of cDNA in lung cancer cell lines was relatively higher than that of gastric cancer cell lines (Fig. 1B). The WTA yields were proportional to the initial amount of mRNA in Quartz-Seq (Supplementary Fig. 2). A recent report suggests that mRNA amount in a single cell depends on the cell size<sup>38</sup>. The size of lung cancer cell line (NCI-H1975: 23.5  $\mu\text{m}$ ) was relatively larger than gastric cancer cells (AGS: 14.9  $\mu\text{m}$ , SNU-1: 11.7  $\mu\text{m}$ , NCI-N87: 12.2  $\mu\text{m}$ ), therefore, the initial amount of mRNA may be higher in lung cancer cells compared to gastric cancer cells, resulting in a higher yield of the resulting cDNA in WTA.

Next, we compared the GCM-based method and micromanipulation in terms of the yield and the average length of cDNA in WTA of lung cancer cell line NCI-H1975 cells. The cDNA yield in GCM was higher than that obtained by micromanipulation (Supplementary Fig. 3A), whereas no significant difference was observed in the average length (Supplementary Fig. 3B,  $p = 0.147$ ). Because the sample volume in single-cell Quartz-seq (WTA) is extremely low (0.4  $\mu\text{l}$ ), the reaction is largely unaffected by excess water. Excess water cannot be avoided during micromanipulation as the single-cell sample volume is transferred to a reaction tube, and lower cDNA yield in micromanipulation



**Fig. 1** Single-cell isolation and gene expression analysis based on microcavity array and gel-based cell manipulation. **A** Schematic overview of single-cell isolation based on MCA/GCM; **B** Violin plot shows the yield of WTA products from cancer cell lines and gastric CTCs. Each dot represents a single-cell. The red bar represents mean value. The black bar represents deviation. **C** Cluster analysis of single-cell gene expression patterns. Heat map shows Cq values of 11 genes by qPCR assays performed on single NCI-H1975 cells isolated by micromanipulation (green) and MCA/GCM (yellow). Red indicates high gene expression, white indicates median gene expression, blue indicates low gene expression, and gray represents undetectable expression.

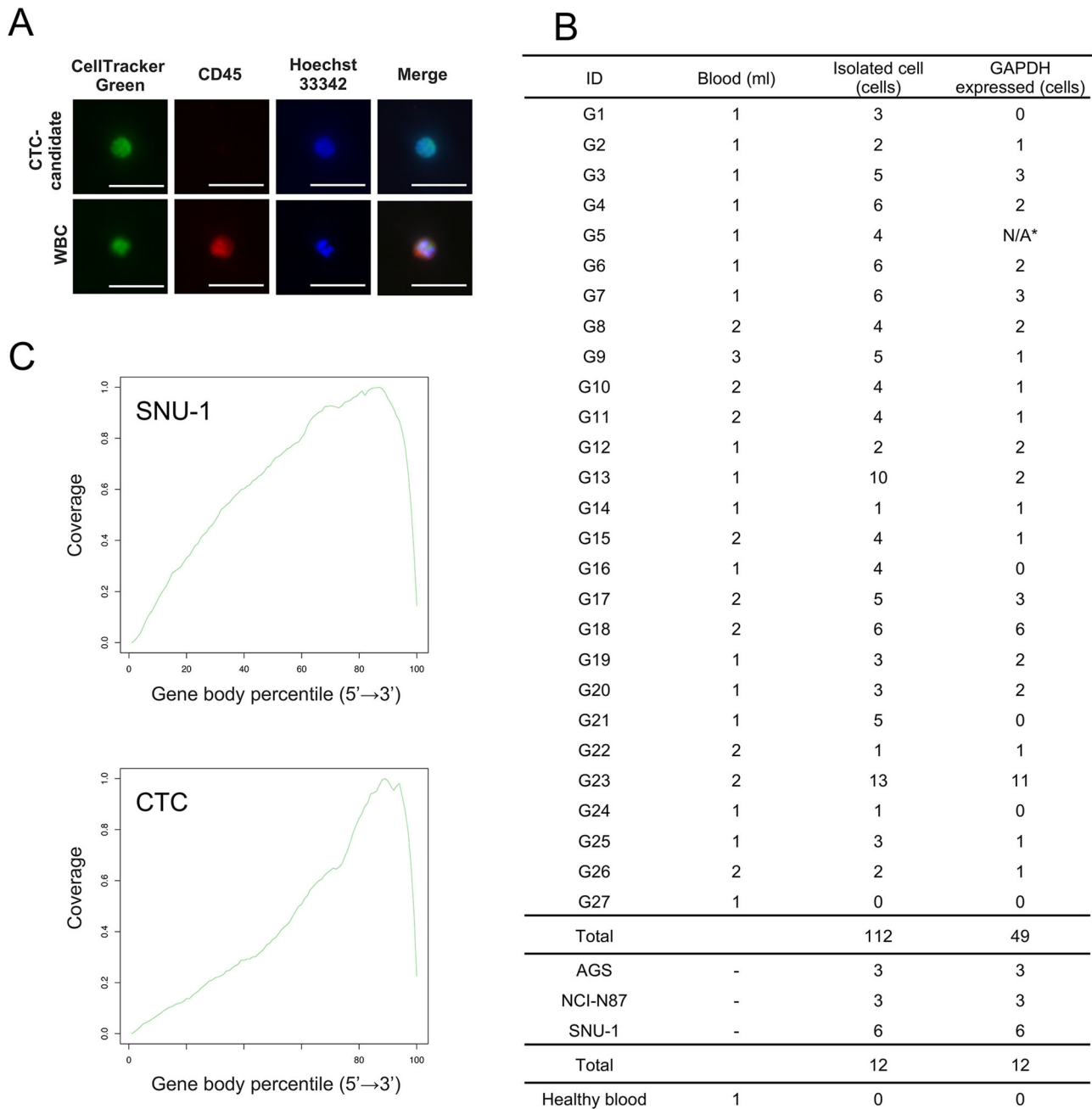
may be due to the water content in the WTA reaction. In other words, GCM-based single-cell manipulation is advantageous for extremely low volume reactions because hydrogel-encapsulated single-cells comprise of minimal water.

Furthermore, gene expression patterns of single-cells in the GCM-based method and micromanipulation were compared to assess WTA bias. Quantitative PCR (qPCR) analysis for 11 genes (including two housekeeping genes: *GAPDH* and *ACTB*) was carried out on the whole transcriptomes of NCI-H1975 cells (eight single-cells each). Figure 1C shows the results of cluster analysis on gene expression patterns of a total of 16 single-cells. Gene expression was represented as Cq values: red, high; white, median; blue, low; gray, undetectable gene expression. Cluster analysis revealed no difference in gene expression patterns between the GCM-based method and micromanipulation as no clustering appeared by either method. These results indicate that the GCM-based method exhibits a similar performance as micromanipulation in gene expression analysis. Besides, we evaluated the gene expression of *GAPDH* of GCM-isolated single gastric cancer cell lines ( $n = 12$ ) for the quality control (QC) for single-cell RNA-seq analysis. The *GAPDH* expression was successfully detected from all of gastric cancer cell lines (Fig. 2B).

**Isolation and characterization of CTC candidates from metastatic cancer patients.** Blood samples were collected from 27 gastric cancer patients between April 2017 and October 2019 (Supplementary Data 1). Out of these samples, 112 CTC candidates from 26 patients were detected by immunostaining (Fig. 2A & B) and were subjected to WTA. Besides, we also performed CTC isolation test for one healthy donor and did not detect any CTC candidates (Fig. 2B). The yield of amplified cDNA from

single-CTC candidates ranged from 0.1 to 99.3 ng (Fig. 1B). A large variation of CTC cDNA yield was observed compared to gastric cancer cell lines. Previously described, cDNA yield was affected by input amount of RNA. Besides, CTCs didn't originate from single cancer cell, unlike cancer cell line. Thus, these results indicated CTCs might be cell population with a wide range of mRNA abundance compared to cell lines. Then, the gene expression of *GAPDH* was evaluated as a QC for cDNA prior to RNA-seq (Fig. 2B). The 49 cells that passed the QC were defined as CTC candidates with sufficient quality RNA. The QC pass rate of CTCs was 43.8% (49/112 CTCs).

**Evaluation of quality and contents of RNA-seq data.** We performed RNA-seq analysis for 49 single CTC-candidates, three AGS cells (gastric cancer cell line), three NCI-N87 cells (gastric cancer cell line), and six SNU-1 cells (gastric cancer cell line, three cells were isolated from PBS, others were isolated from cancer cell-spiked bloods). Approximately 5 M raw reads were obtained from 49 CTC-candidates and 12 single cancer cells. The alignment rate of raw reads from single CTC-candidates to the human genome were  $68.9 \pm 17.4\%$ , and was comparable to the raw reads from single cancer cells ( $73.8 \pm 22.9\%$ ). Figure 3C shows the distribution of mapped reads of single-cell RNA-seq data. Although SNU-1 showed 3' bias due to poly(A) tailing-based WTA, data from CTC-candidates showed a stronger 3' bias. These data suggested that RNA in CTCs tended to be degraded, therefore, we chose RPM as a gene expression unit. To evaluate transcript abundance, we designated transcripts with RPM  $< 1.0$  as detectable transcripts. We detected 4376–12,084 transcripts from single cancer cells ( $n = 12$ ), whereas,  $4064 \pm 1967$  transcripts were detected from single CTC-candidates (with a range of 1183 to



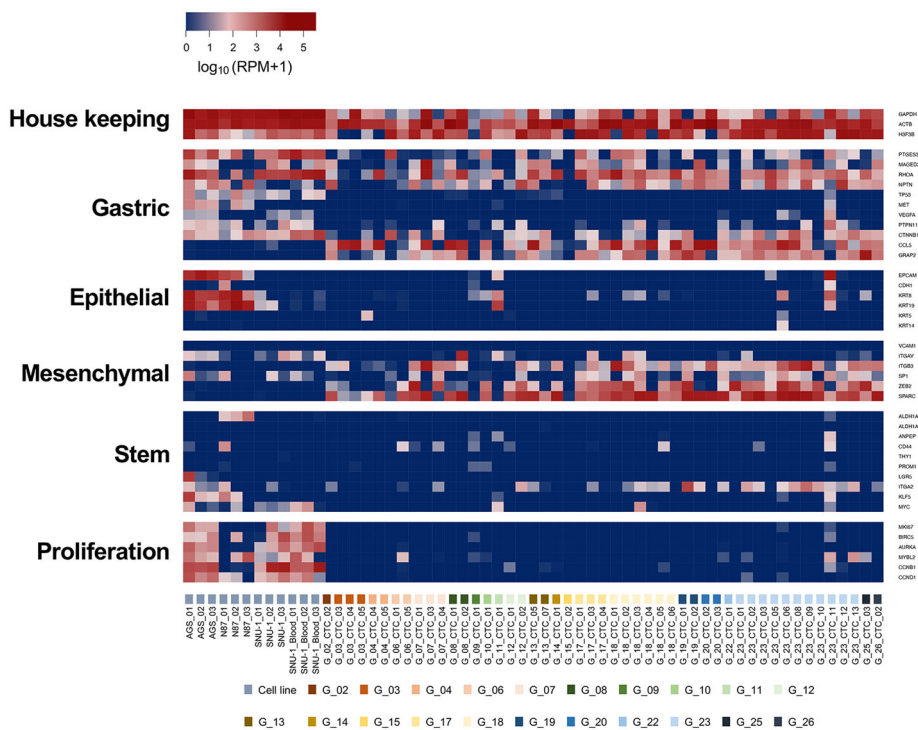
**Fig. 2 Isolation of gastric CTCs and RNA-seq analysis. A** Fluorescence image of a representative CTC-candidate and WBC. Scale bar: 50  $\mu$ m. **B** Summary of CTC isolation and quality control. Samples in G5 were lost by technical error (N/A\*). **C** Gene body coverage from a single SNU-1 cell and gastric CTC.

11,230 transcripts,  $n = 49$ ). Besides, there was no correlation between the number of detected transcripts and the yield of amplified cDNA. We evaluated the gene expression value of the WBC marker CD45 (*PTPRC*) to detect contaminated WBCs. Almost all CTC candidates ( $n = 48$ ) were negative for the WBC marker CD45 (*PTPRC*). One CTC candidate (G\_18\_CTC\_4) showed weak expression of the *PTPRC* gene. However, we did not exclude these cells from gene expression analysis because their expression levels were very low. Therefore, we confirmed that CTC candidates were not contaminated with WBCs. In addition, we measured the ratio of mitochondrial gene expression to evaluate the quality of the single-cell library (Supplementary Fig. 4). The ratio of mitochondrial genes of CTC-candidates was range from 0.02 to 68.17% (average: 5.07%; median: 1.33%). We excluded two single cells (G\_04\_CTC\_05 and G\_07\_CTC\_03)

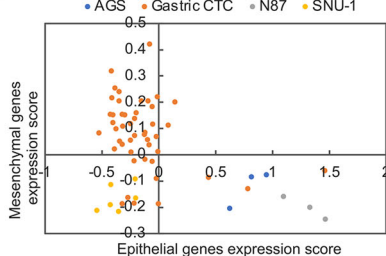
with >25% mitochondrial gene ratio as low-quality cells from further analysis.

**Expression patterns of gene markers in single-CTCs.** To characterize gastric cancer CTCs, we first evaluated the expression patterns of known markers. Figure 3A illustrates the expression of housekeeping genes, gastric cancer gene markers, epithelial marker genes, mesenchymal marker genes, and stem cell/cell proliferation-related genes<sup>10,12,39–41</sup>. Blue color indicates low expression level of genes, whereas red represents high expression level. *RHOA* or *CTNNB1* was expressed in all CTCs (47/47). Analysis of CTCs in prostate cancer and melanoma also shows that CTCs retain the gene expression pattern characteristic of the primary tumor<sup>18,42</sup>. Thus, gastric cancer CTCs have been suggested to possess characteristics of gastric cancer cells. As expected, almost all CTCs

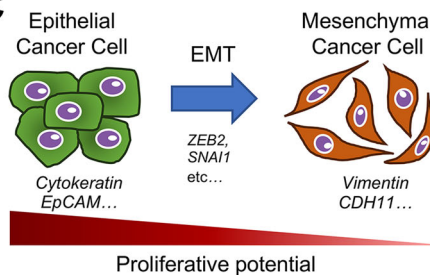
A



B



C



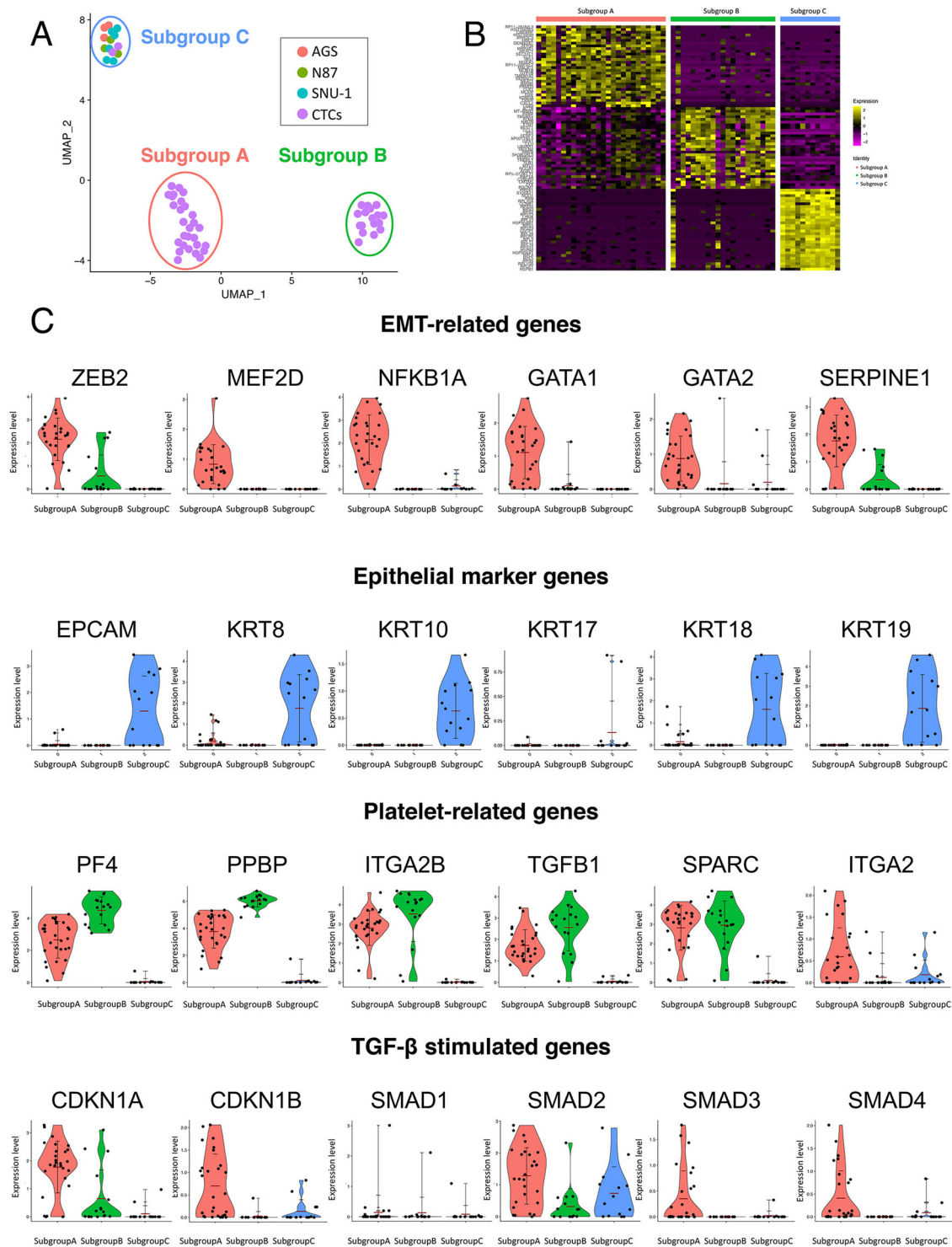
**Fig. 3 Targeted analysis of single-cell RNA-seq data.** **A** The heat map shows expression of housekeeping genes, gastric, epithelial, epithelial, stem cell, and proliferation markers. Scale in  $\log_{10}(\text{RPM} + 1)$ . Expression level increases from blue to red. **B** Scatter plots based on the expression levels of epithelial and mesenchymal genes. **C** Schematic of EMT induction and proliferative potential of cancer cells.

displayed low expression of epithelial marker genes (*KRT* family and *EpCAM*) and high expression of mesenchymal genes such as *ZEB2*, which is one of the widely known EMT-inducing factors, and *SERPINE1*. Furthermore, 23 CTCs expressed epithelial genes such as the *KRT* family of genes (Fig. 3A). To evaluate the EMT signature of CTCs, we measured the gene expression scores of epithelial and mesenchymal genes (Fig. 3B). Most CTCs showed stronger expression of mesenchymal genes than the cell lines. However, three CTCs showed expression of epithelial genes that was comparable with the expression observed in epithelial gastric cancer cell lines (AGS and N87). Furthermore, two CTCs showed weak expression of epithelial genes, despite expressing mesenchymal genes. These results indicate that most gastric cancer CTCs had undergone EMT, consistent with the results of previous reports<sup>43,44</sup>. However, the expression of proliferation markers in CTCs was much lower than the cell lines. In particular, the expression of *Ki67* (*MKI67*), which has an integral role in cell cycle progression, was not detected, suggesting that CTCs are a cell population in which growth has been arrested. In general, mesenchymal cancer cells have lower proliferative potential than epithelial cancer cells (Fig. 3C)<sup>45</sup>. This low proliferative potential of

CTCs may be due to EMT. Biphenotypic properties and arrested cell proliferation have been confirmed in CTCs of pancreatic, prostate, and breast cancer<sup>7,12,13</sup>. These characteristics may also provide CTCs with important phenotypes for survival in blood vessels. In addition, expression of stem cell markers such as *CD44* was observed in 16% (8/47) CTCs and *ITGA2* in 45% (22/47) CTCs. This result suggested that a portion of gastric cancer CTCs have a cancer stem cell-like property. Besides, *ITGA2* is not only a mesenchymal stem cell marker, but also a protein that contributes to binding to extracellular matrix and platelets. Recent studies on cancer tissues suggest that *ITGA2* is related to the invasiveness of gastric and pancreatic cancer cells<sup>46,47</sup>. This result suggests that some external factors may influence CTCs and that gastric cancer CTCs have heterogeneous properties.

**Cluster analysis of single-CTCs and clinical relevance.** Next, unsupervised analysis of the transcriptome of single-CTCs was performed. In recent years, unsupervised clustering by transcriptome analysis has enabled the discovery of new cell sub-populations based on expression patterns<sup>48</sup>. We classified CTCs

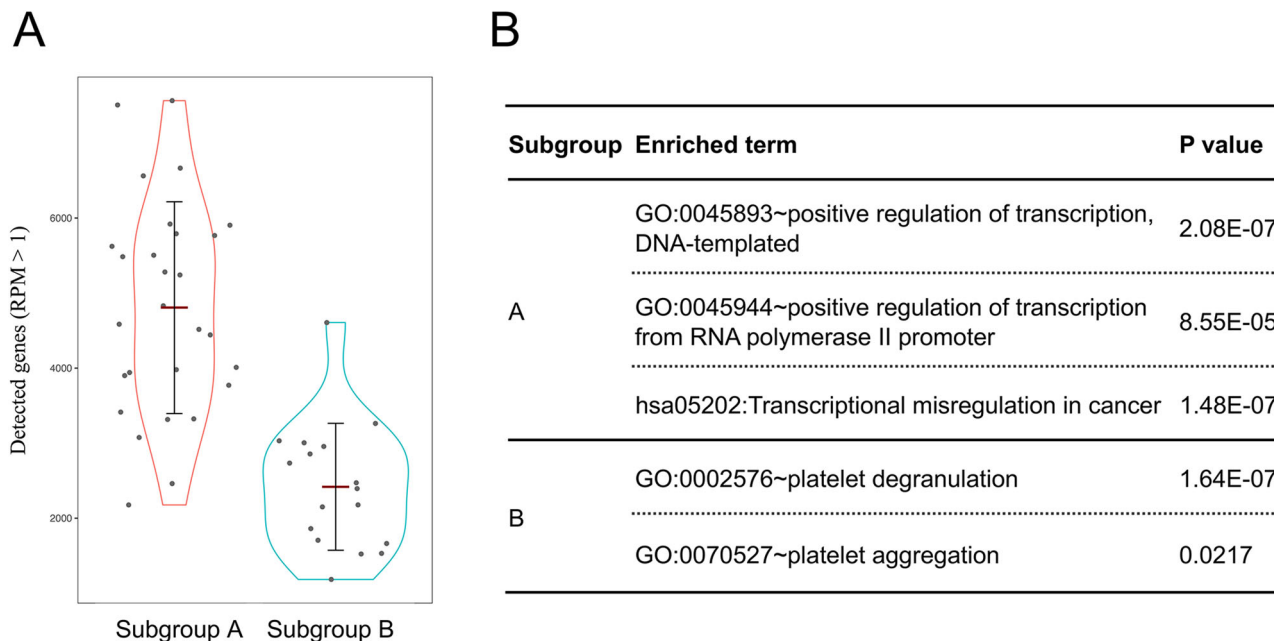




**Fig. 4** Gene expression and clustering of gastric CTCs. **A** UMAP plot of single gastric cancer cells and gastric CTCs. (12 cancer cells and 47 CTCs from 21 patients. Each dot represents a single cell. Salmon pink: AGS, green: NCI-N87, blue: SNU-1, purple: gastric CTCs) **B** Heatmap of overexpressed genes of each subgroup. Top 30 overexpressed genes from each subgroup are shown. Heatmap showed log scaled RPM that increases from purple to yellow. **C** Violin plot for expression of EMT-related genes, epithelial genes, platelet related-genes, TGF- $\beta$  stimulated genes. Each dot represents a single-cell. The red bar represents mean value. The black bar represents deviation.

into multiple subtypes, and performed clustering by Seurat, a package for single-cell RNA-seq analysis, and extracted genes that characterize each cluster<sup>49</sup>. Figure 4A is a Uniform Manifold Approximation and Projection (UMAP) plot that illustrates the clustering of CTCs and cancer cell lines. Each dot indicates one cell, G (green) indicates gastric cancer CTCs, and

AGS (salmon pink), N87 (blue), and SNU1 (purple) indicates gastric cancer cell lines. Single cells were classified into three subgroups; most of the CTCs belonged to subgroups A and B. The cell lines and only two CTCs belonged to subgroup C. We measured nCount and nFeature of CTCs to confirm that clustering analysis revealed biological differences. We confirmed



**Fig. 5 Comparison of subgroup A and B.** **A** Comparison of the number of detected genes (RPM > 1) between subgroup A and B. Each dot represents a single-cell. The red bar represents mean value. The black bar represents deviation. **B** Gene ontology analysis of enriched genes.

that there were no remarkable differences in nCount between each subgroup, but the nFeature was larger for subgroups A and C than for subgroup B (Supplementary Fig. 5). CTCs in subgroup B were detected in patients with CTCs in subgroups A and C (Supplementary Table 1). Thus, we conclude that our clustering results did not show technical differences but biological differences in CTCs. Figure 4B is a heat map displaying the characteristic overexpressed genes in each subgroup (names of overexpressed genes was shown in Supplementary Table 2). Purple color indicates low expression level of genes, whereas yellow represents high expression level. Subgroup A was found to express EMT-related transcription factors such as *ZEB2*, *MEF2D*, *GATA1*, and *GATA2*, and transcription-coupling factors such as *KMT2A* (Fig. 5B, C). Interestingly, the expression of *NFKBIA*, an inhibitory protein of NF- $\kappa$ B, was observed. *NFKBIA* expression is induced by activation of NF- $\kappa$ B<sup>50</sup>, suggesting that transcriptional induction by the NF- $\kappa$ B pathway occurs in this subgroup of cells. Subgroup B showed strong expression of genes related to platelet activation and aggregation, such as *ITGA2B*, *PF4* (Platelet Factor 4), and *PPBP* (pro-platelet basic protein, isoform of *CXCL7*) (Fig. 4C). Furthermore, platelet-related genes such as *TGFBI* and *SPARC* were also expressed in subgroup A, suggesting that subgroups A and B are cells to which platelets adhere. In recent years, it is becoming evident that contact between platelets and cancer cells leads to progression of EMT. According to Labelle et al., TGF- $\beta$  released from platelets activates the SMAD signaling pathway in cancer cells, and NF- $\kappa$ B activates EMT in the cancer cells-platelet interaction<sup>51</sup>. Besides, recent studies have also revealed the importance of the interaction between cancer cells and platelets in EMT<sup>52</sup>. Thus, we inferred that EMT-related gene expression and *NFKIB* expression observed in subgroup A were due to contact with platelets. This observation suggests that gastric CTCs undergo EMT through platelet adhesion. In contrast, expression of EMT-related genes was not detected in subgroup B, and the number of expressed genes (nFeature) tended to be lower than that of subgroup A (subgroup A:  $4805 \pm 1412$  genes; subgroup B:  $2418 \pm 845$  genes) (Fig. 5A). Besides, the expression level of platelet-related genes in subgroup B

was higher than that of subgroup A (Fig. 4C). Gene Ontology analysis also showed that subgroup A strongly expressed transcription-related genes (GO:0045893, positive regulation of transcription, DNA-templated; GO:0045944, positive regulation of transcription from RNA polymerase II promoter) (Fig. 5B, Supplementary Fig. 6A and Supplementary Data 2) and subgroup B strongly expressed platelet-related genes (GO: 0002576, platelet degranulation; GO:0070527, platelet aggregation) (Fig. 5B, Supplementary Fig. 6B and Supplementary Data 3). These results suggested that cells in subgroup B had more adhered platelets or had a lower CTC transcriptional activity. Interestingly, the expression of cell cycle regulators such as *CDKN1B* and *CDKN1A*, which are known to be induced by TGF- $\beta$ , was observed in subgroup A (Fig. 4C). These regulators cause cell cycle arrest in the G1 phase; *p21* is known to contribute to resistance to chemotherapy<sup>53</sup>. Furthermore, subgroup A CTCs were detected in patients who failed to respond to first-line treatment and had a longer duration of therapy before CTC collection. (Supplementary Table 3 and Supplementary Fig. 7). From this result, it was inferred that CTCs in subgroup A were resistant to chemotherapy.

In subgroup C, which includes cancer cell lines, many genes related to metabolic pathways and epithelial markers were detected as overexpressed genes. We concluded that subgroup C comprises of actively metabolizing cells. Platelet-related genes were not expressed in CTCs belonging to subgroup C, and high expression of epithelial markers such as *EpCAM*, *KRT8*, and *KRT18*, absent in subgroups A and B, was detected (Fig. 4C). This suggests that CTCs in subgroup C have completely different characteristics from other CTCs. We note that all patients ( $n = 2$ ) in subgroup C died within 1 month of blood collection (Supplementary Table 1). Therefore, classification of patients into subgroups according to the gene expression pattern of cancer cells may be effective in predicting the prognosis of gastric cancer patients.

## Discussion

In this study, we characterized the CTCs of patients with metastatic gastric cancer by single-cell RNA-seq analysis. There are several reports of the recovery of gastric cancer CTC by the

CellSearch System<sup>27</sup>. However, the low detection rate of CTCs and low cut-off value suggest that there are numerous CTCs that cannot be detected due to EMT<sup>28</sup>. Recently, clinical studies revealed the advantage of a size-selective CTC enrichment method for the detection of gastric CTCs<sup>54,55</sup>. Thus, the size-selective method has the potential to achieve higher sensitivity in CTC detection from metastatic gastric cancer patients than the CellSearch system. In particular, Cheng and colleagues showed CTCs with mesenchymal gene (vimentin and twist) expression by RNA-in situ hybridization (RNA-ISH) and correlated the number of mesenchymal-CTCs with disease progression<sup>30</sup>. Although RNA-ISH is powerful tool for identifying the EMT-status of CTCs, it is difficult to obtain information about the molecular mechanism of EMT progression because of the small number of marker genes. The size-selective CTC enrichment method has great potential for understanding of the mechanism of EMT in gastric CTCs. Therefore, we adopted the size-selective CTC enrichment method and the surface antigen-independent CTC detection method (CellTracker Green-positive and CD45-negative). As a result, CTC candidates were detected in 96% (26/27) of the cases. The detection rate of CTC in gastric cancer was remarkably high compared to previous reports (11–85%), and the size-selective CTC recovery method was found to be effective for the detection of gastric cancer CTCs.

About half of the obtained CTCs were low-quality cells in which no amplification product was obtained or no housekeeping gene expression was observed. The QC pass rate was much lower than that of cancer cell lines. These results suggested that the RNA quality of single-CTCs were found to be lower than that of cancer cell lines. It is reported that cytoplasmic mRNA are degraded as an early part of the apoptotic response<sup>56</sup>. Thus, CTCs with low-quality RNA may be in the early phase of apoptosis owing to several factors such as shear stress in blood vessels<sup>57,58</sup>. Previous studies with single CTCs from breast cancer and prostate cancer have reported analysis issues with 30–70% of CTCs due to poor quality<sup>7,18,59</sup>. The QC pass rate obtained in this study was comparable to previous findings, suggesting that it is a common property among all cancer types.

We expected that the detection of CD45-negative cells as CTCs will be influenced by contamination of hemocyte cells, however, this effect can be negated by identifying the expression of hemocyte-related genes. All CTC-candidates in the present study were of a different species from PBMC as shown by unsupervised clustering. Till date, most single-cell RNA-seq analyses of CTCs utilize antibody staining of epithelial markers such as EpCAM and E-cadherin to detect CTC, and many isolated CTCs exhibit epithelial gene expression<sup>7,12,18</sup>. Conversely, most of the CTCs detected in this study were cells that lost epithelial gene expression due to EMT. Thus, isolation of CD45-negative cells as CTC candidates allows the inclusion of EMT-induced CTCs.

The majority of CTCs expressed genes associated with platelet binding. As platelet-specific genes such as *PF4* and *PPBP* were expressed, we assumed that the platelets were attached to the separated CTCs. In recent years, it has become clear that the adherence between cancer cells and platelets not only protects cancer cells from immune cells and shear stress, but also promotes EMT. Previous studies have reported cases in which platelet adhesion was suggested in CTCs isolated from mice model of pancreatic cancer<sup>12</sup>. It has been reported that CTCs or circulating tumor microemboli (CTM) with attached platelets were detected in human samples, but gene expression analysis of these cells was not performed<sup>60</sup>. In the present study, CTCs with predicted platelet adhesion did not express epithelial marker genes. This result shows that it is possible to detect CTCs independent of epithelial markers. Besides, a strong expression of *TGFBI* was observed in CTCs with advanced EMT. TGF- $\beta$  is mainly secreted

by platelets and induces EMT by activating the SMAD pathway in cancer cells<sup>51</sup>. In recent years, it has been revealed that binding of cancer cells to platelets promotes the production of TGF- $\beta$  by cancer cells and synergizes EMT with platelet-derived TGF- $\beta$ <sup>61</sup>. Taken together, we propose that platelet adhesion plays a major role in the induction of EMT in CTCs. Recent study detected CTM from 24% of patients with stage IV gastric cancer before chemotherapy<sup>62</sup>. Abdallah et al. reported decreasing of CTM detection rate in local gastric cancer patient after the neoadjuvant chemotherapy and surgery<sup>63</sup>. These studies implicate that the chemotherapy might decrease CTMs. We did not detect CTM from our cohort including patient without chemotherapy. However, our previous study demonstrated the detection of CTM in metastatic non-small cell lung cancer patients by MCA system<sup>64</sup>. Thus further study will reveal the relationship between EMT and platelet-CTM interaction in gastric cancer patients.

Recently, Sun et al. have shown that hepatocellular carcinoma CTCs possess epithelial properties immediately after they are released from tissues, but EMT may induced within pulmonary capillaries. Also, Aceto et al. have shown that the TGF- $\beta$  pathway activity of CTC is high in cases of metastasis to other organs in breast cancer. Based on these findings, it was suggested that CTC activate the TGF- $\beta$ /SMAD pathway and induces EMT by adhesion to platelets in blood vessels. Generally, platelet-cancer cell interaction is analyzed by in vitro experiments or mouse models. Our results showed that single-cell analysis of CTC may reveal more information about the function of platelets in metastasis and may contribute to the development of new therapeutic targets. Recent studies have highlighted the importance of tumor-educated platelets in cancer progression<sup>52</sup>. Our results suggest that this single CTC isolation system could contribute to revealing the direct communication between CTCs and platelets. Moreover, the expression of *CDKN1A* (p21) and *CDKN1B* (p27) was observed in most EMT-induced CTCs, suggesting that the cell cycle is arrested in the G1 phase. These genes are known to be involved in chemoresistance of cancer cells and dormancy. Therefore, gene expression analysis of CTC may be effective in analyzing the mechanism of acquisition of resistance to chemotherapy in gastric cancer, and may aid decisions involving treatment selection for patients.

The clinical cases in which CTCs exhibited epithelial gene expression (G\_12, G\_25) had a considerably worse prognosis than other cases. The epithelial CTCs showed a similar gene expression pattern to the cell lines, with a strong expression of genes associated with various metabolic pathways. This suggests that these CTCs are more metabolically active than other CTCs. Among cases of metastatic breast cancer with extremely high number of epithelial CTCs, death has been reported within 1 month of blood collection<sup>1,65,66</sup>. It was also confirmed that the number of epithelial CTCs was significantly high in patients who did not show response to therapeutics in cases of metastatic gastric cancer<sup>67</sup>. Furthermore, single-cell gene expression analysis of CTCs in pancreatic cancer cases also report a poor prognosis in cases wherein CTCs show epithelial characteristics<sup>59</sup>. Taken together, these findings suggest that CTCs with epithelial gene expression may be an indicator of poor prognosis. The CTC detection method using the CellSearch system or immunostaining detects CTCs based on the expression of an epithelial protein, however, it is difficult to quantitatively evaluate the expression level. As single-cell gene expression analysis facilitates a more quantitative characterization of CTC, it may be used for the pursuit of new prognostic markers. Moreover, no expression of platelet adhesion-related genes was observed in epithelial CTCs. Therefore, we propose that these CTCs indeed avoid the immune system and fluid shear caused by blood flow. However, a conclusion cannot be drawn on the exact mechanism of this function



of CTCs and further analysis is required in this direction. In addition, our analysis involved a limited cohort size and a small number of baseline patients; a larger scale analysis is warranted in future. We plan to perform a large-scale clinical study for validation of marker genes extracted in this study. Furthermore, we noted that the size of gastric CTCs tended to be smaller than that of cancer cell lines. This observation suggests that there might be small CTCs that are difficult to entrap using the size-selective method. Further optimization for size-selective CTC enrichment process is still required.

In summary, we constructed a pipeline for size-selective and antigen-independent CTC separation and single-cell RNA-seq analysis based on the MCA/GCM method. A majority of the gastric cancer CTCs were EMT-induced cells, and we suggest that the adhesion of platelets initiates a signal transduction, which leads to the induction of EMT, a phenomenon mainly occurring in blood vessels. Furthermore, our results suggest that CTCs exhibit cell cycle arrest and are associated with resistance to chemotherapy. Furthermore, some CTCs were epithelial and were found to be useful markers for prognosis prediction.

## Methods

**Cell lines and culture.** Human lung cancer cell lines, NCI-H1975 (ATCC<sup>®</sup> CRL5908<sup>™</sup>) and PC9 (ECACC 90071810), and human gastric cancer cell lines, AGS (ATCC<sup>®</sup> CRL-1739<sup>™</sup>), SNU-1 (ATCC<sup>®</sup> CRL-5971<sup>™</sup>), and NCI-N87(ATCC<sup>®</sup> CRL-5822<sup>™</sup>) were used in this study. The cell lines were cultured in RPMI-1640 media containing 10% (v/v) fetal bovine serum (Invitrogen Corp., Carlsbad, CA) and 1% (v/v) penicillin/streptomycin (Invitrogen Corp.) for 3–4 days at 37 °C with 5% CO<sub>2</sub> supplementation. Prior to experiments, all cells, except for SNU-1 (floating cells), grown up to 80% confluency were trypsinized and re-suspended in PBS. SNU-1 cells were directly recovered from the media and re-suspended in PBS. Cells were then stained with 5 μM CellTracker Green 5-chloromethylfluorescein diacetate (CMFDA; Thermo Fisher Scientific, Waltham, MA) and 1 μg/ml Hoechst 33342 (Thermo Fisher Scientific) for 20 min and centrifuged at 400 g for 3 min to obtain cell pellets. After washing the cells twice with PBS, they were re-suspended in PBS containing 2 mM EDTA and 0.5% bovine serum albumin (BSA; PBS-EB). For spike-in experiments, SNU-1 (100 cells) was spiked into human blood (1 mL). Human blood samples were collected from healthy donors at the Tokyo University of Agriculture and Technology. Experimental protocols were approved by the Institutional Review Board of the Tokyo University of Agriculture and Technology (Approval code: No. 30-10).

**Patient samples.** The clinical study was conducted in Tokyo Metropolitan Cancer and Infectious Diseases Center, Komagome Hospital. The study design involving human subjects was approved by the Institutional Review Board at Tokyo Metropolitan Cancer and Infectious Diseases Center Komagome Hospital (Approval Number: 1441). Blood samples were collected from patients with metastatic gastric cancer treated with chemotherapy ( $n = 24$ ) from April 2017 to October 2019. We also obtained blood samples from three metastatic gastric cancer patients without any chemotherapy. Detailed disease status and therapy are provided in Table 1. Written informed consent was obtained from all patients prior to blood sampling. For each patient, blood sample was collected by venipuncture into Venoject II 5 ml EDTA-2Na tubes (VP-NA050KN, Terumo, Japan). Peripheral blood samples for CTC analysis were collected after withdrawal of the first several milliliters of blood for clinical use to avoid potential skin cell contamination from the venipuncture. All the blood samples were processed within 2 h of collection.

**Single-cell isolation by micromanipulation.** Prior to micromanipulation-based single-cell isolation, a spacer seal (a slide seal for in situ PCR, inner space: 9 × 9 mm<sup>2</sup>, thickness: 300 μm, Thermo Fisher Scientific, Waltham, MA, USA) was affixed at the glass slide (Matsunami Glass, Osaka, Japan). Cellular RNA was stabilized using Cell-Cover (AL Anacyte Laboratories UG, Hamburg, Germany), if required, according to the manufacturer's protocol. After determination of cell concentration by a hemocytometer, ~100 cancer cells were suspended in 100 μl of PBS-EB. The cell suspension was mounted only on the inside of the spacer seal and observed under a fluorescence microscope (IX71; Olympus Co., Tokyo, Japan). Single cancer cells exhibiting strong fluorescence were manually picked up by a micromanipulator equipped with a 30 μm diameter glass capillary (PicoPipet; Nepa Gene Co., Ltd., Chiba, Japan). Then, single cells were transferred into 0.4 μl of a lysis buffer (0.5% NP40) in 200 μl PCR tubes (N8010840, Thermo Fisher Scientific). In this study, we use 0.5% NP40 as lysis buffer for subsequent Quartz-seq based whole transcriptome amplification. Isolated single-cells were immediately frozen with liquid nitrogen and stored at –80 °C for further whole transcriptome amplification.

**Fabrication of CTC recovery device.** An MCA made of nickel by electroforming (Optonics Precision Co. Ltd.) was used for CTC recovery. The circular pore with a diameter of 8 μm was fabricated for entrapment of the CTCs. The pore diameter was previously optimized for size- and deformability-based CTC enrichment<sup>31</sup>. The distance between pores was 125 μm, with a total of 3969 cavities arranged in each 63 × 63 array for subsequent GCM steps. The CTC recovery device was fabricated by integration of MCA, PMMA-made mold and PDMS-made flow channel using spacer tape<sup>31</sup>.

**CTC enrichment and staining.** One ml of cancer cell-spiked PBS-EB (or blood) or patient blood was introduced into the PBS-EB pre-filled CTC recovery device. Subsequently, negative pressure was applied to the samples with a peristaltic pump (MINIPULS 3; Gilson, Wisconsin, USA) which was connected to a vacuum line. The sample was passed through the MCA at a flow rate of 150 μl/min. To remove any residual cells, 2 mL of PBS-EB was passed through the MCA at a flow rate of 200 μl/min for 10 min. For the patient blood samples, 1 mL PBS-EB containing 5 μM CellTracker Green CMFDA and 1 μg/ml Hoechst 33342 was introduced into the MCA and incubated for 20 min and the excess dye was washed with 1 mL PBS-EB. Then, RNA of recovered cells was stabilized with CellCover throughout the MCA for 15 min. After washing with 1 mL PBS-EB, cancer cell-spiked PBS-EB samples were moved to the GCM step. For patient blood samples, cells were stained with anti-CD45 antibody (Hitachi Chemical, Tokyo, Japan) and Texas Red-labeled secondary antibody (Hitachi Chemical). Finally, the array was washed with 1 mL of PBS-EB and PBS to remove excess antibodies.

**Single-cell isolation based on GCM.** Prior to GCM, polyethylene glycol diacrylate (PEGDA, Mn = 700; Sigma-Aldrich) with 0.5% Irgacure 2959 (photoinitiator) was introduced onto the MCA. Then, the CTC recovery device was covered with a coverslip (24 × 40 mm; Matsunami Glass). Coverslips were modified with 3-(trimethoxysilyl)propyl methacrylate to provide a stiff connection between the coverslip and hydrogel. The MCA was observed under a fluorescence microscope (BX53; Olympus Co.), and cancer cells and CTC-candidates were determined as CellTracker Green- and Hoechst-33342 positive and CD45-negative cells. The subsequent steps were performed as previous study as follow<sup>32</sup>. Briefly, single-cells were encapsulated into the hydrogel by light irradiation ( $\lambda = 365$  nm, 12.7 mW/cm<sup>2</sup> for 30 s). After peeling off the coverslip, hydrogel-encapsulated single-cells were resuspended into 0.4 μl of lysis buffer (0.5% NP40) in 200 μl PCR tubes (N8010840, Thermo Fisher Scientific) using tweezers. Isolated single-cells were immediately frozen with liquid nitrogen and stored at –80 °C for further whole transcriptome amplification.

**Single-cell whole transcriptome amplification.** A recent study revealed the superior performance of poly A tailing-based WTA in detection of genes<sup>68</sup>. In addition, there are several reports of single-cell transcriptome analysis of CTCs using poly A tailing-based WTA<sup>7,12,14,15</sup>. Thus, we selected Quartz-seq, based on the poly A tailing reaction, for single-cell WTA in the present study<sup>37</sup>. In brief, single cells in lysis buffer were placed in an aluminum PCR rack on ice and subsequently mixed with 0.8 μl priming buffer (1.5× PCR buffer with MgCl<sub>2</sub> [TaKaRa Bio], 41.67 pmol/L of RT primer, 4 U/μl of RNase inhibitor [RNasin Plus; Promega Corp., Madison, WI, USA], and 50 μmol/L dNTPs). After mixing and centrifuging (3000 × g, 10 s), the mRNA from single cells was denatured and captured by RT primer, oligo dT linked to an adaptor sequence, using a thermal cycler (70 °C for 90 s and 35 °C for 15 s). Reaction mixtures were placed in an aluminum PCR rack on ice, and treated with 0.8 μl of RT buffer (1× PCR buffer, 25 U/μl reverse transcriptase [SuperScript III; Life Technologies], and 12.5 mmol/L DTT). First-strand cDNA was synthesized using a reverse-transcription reaction (35 °C for 5 min and 45 °C for 20 min). After heat inactivation at 70 °C for 10 min, the reaction tubes were placed in an aluminum PCR rack on ice and treated with 1 μl of the exonuclease solution (1× Exonuclease buffer [TaKaRa Bio] and 1.5 U/μl exonuclease I [TaKaRa Bio]). The primer digestion reaction was performed to remove excess primers (37 °C). Reaction tubes were placed in an aluminum PCR rack on ice after heat inactivation at 80 °C for 10 min, and subsequently treated with 2.5 μl of poly-A-tailing buffer (1 × PCR buffer, 3 mmol/L dATP, 33.6 U/μl terminal transferase [Roche Applied Science], and 0.048 U/μl RNase H [Invitrogen]). The poly A tailing reaction was performed at 37 °C for 50 s and heat inactivated at 65 °C for 10 min. The reaction samples were placed in an aluminum PCR rack on ice and mixed with 23 μl of the second strand buffer (1.09× MightyAmp Buffer v2 [TaKaRa] and 0.054 U/μl MightyAmp DNA polymerase [TaKaRa]). The second-strand cDNA reaction was performed (98 °C for 130 s, 40 °C for 1 min, and 68 °C for 5 min), and reaction tubes were placed in an aluminum PCR rack on ice. Then, 25 μl of PCR buffer (1× MightyAmp Buffer v2 and 1.9 mol suppression PCR primer) was immediately added, and PCR was performed under the following conditions: 98 °C for 10 s, 65 °C for 15 s, and 68 °C for 5 min (21 cycles). Amplified cDNA was purified by the MinElute PCR Purification Kit (QIAGEN, Hilden, Germany) and eluted in 24 μl of buffer EB. Then, yield and average length of amplified cDNA were measured using the High Sensitivity DNA Kit (Agilent, California, USA). For quality checks, amplification of *GAPDH* by qPCR using 50 pg of amplified cDNA as template was performed. Only cells with detectable levels of *GAPDH* were subjected to further qPCR and selected for library construction for RNA-seq analysis.

**RT-qPCR of WTA products.** WTA products (50 pg) were subjected to qPCR analysis using MightyAmp<sup>™</sup> for Real Time (TB Green Plus) kit (TaKaRa Bio, Shiga, Japan).

**Table 1 Summary of patients.**

Patient ID	Sex	Age	Metastasis	Treatment at time of CTC collection	Response	Prior treatment
G1	Man	66	Liver	SOX	PD	Trastumab + FP, SOX
G2	Woman	68	Liver	No	N/A	TS-1
G3	Man	75	Liver, Lymph node	No	N/A	TS-1
G4	Woman	81	Peritoneal dissemination	No	N/A	None
G5	Woman	63	Lung, Pancreas	PTX	PD	SOX
G6	Woman	60	Esophagus	No	N/A	None
G7	Woman	60	Esophagus	No	N/A	None
G8	Man	73	Peritoneal dissemination	No	N/A	S-1
G9	Man	65	Peritoneal dissemination, Lymph node	SOX	Response	SOX
G10	Man	58	Peritoneal dissemination, Lymph node, Liver	SOX	Response	SOX
G11	Woman	52	Peritoneal dissemination	RAM + PTX	PD	SOX, RAM + PTX
G12	Man	71	Peritoneal dissemination	No	N/A	FOLFOX, nab-PTX
G13	Man	69	Peritoneal dissemination, Lymph node, Liver	FOLFOX + BV	PD	FOLFOX + BV
G14	Man	76	Paraortic lymph node, Liver	SOX	Response	SOX
G15	Man	71	Peritoneal dissemination, Paraortic lymph node, Pancreas	SOX	Response	SOX
G16	Woman	68	Liver	CPT-11	Response	TS-1, PTX + RAM, Nivolumab, CPT-11
G17	Man	75	Liver, Lymph node	No	N/A	TS-1, PTX + RAM, Nivolumab, CPT-11
G18	Woman	68	Liver	No	N/A	TS-1, PTX + RAM, Nivolumab, CPT-11
G19	Man	72	Liver	5-FU + Tmab	Response	5-FU + Trastumab
G20	Woman	60	Peritoneal dissemination	FOLFOX	Response	FOLFOX
G21	Woman	68	Liver	No	N/A	TS-1, PTX + RAM, Nivolumab, CPT-11
G22	Woman	81	Peritoneal dissemination	No	N/A	SOX, PTX + RAM
G23	Man	70	Peritoneal dissemination	No	N/A	FOLFOX, RAN + nab-PTX, Nivolumab
G24	Man	75	Peritoneal dissemination	RAM + nabPTX	Response	S-1, RAM + nab-PTX
G25	Man	81	Lymph node	SOX	Response	SOX
G26	Man	66	Peritoneal dissemination	SOX	Response	FOLFOX, SOX
G27	Man	78	Peritoneal dissemination, Colon	No	N/A	SOX

Briefly, 5 µl of diluted WTA products were mixed with 0.8 µl of primers (10 µM of each primer), 10 µl of MightyAmp for Real Time (TB Green Plus) master mix (2X), 0.4 µl of ROX reference dye II (50X), and 3 µl of ultrapure water. The thermal cycler was set to carry out PCR under the following conditions: 98 °C for 1 min; 40 cycles of 98 °C for 10 s and 60 °C for 20 s min; 60 °C for 30 s and Cq values were measured. Eleven genes were selected from PrimerArray® Small Cell Lung Cancer & Non-small Cell Lung Cancer (Human) (TaKaRa) kit for evaluation of amplification bias. The expression pattern of 96 genes in total RNA enriched from NCI-H1975 cells were evaluated according to the manufacturer's protocol. Then, we selected *GAPDH* and *ACTB* as high-expression genes, *TP53* and *GUSB*, *PLA3*, and *CDK2B* as intermediate-expression genes and *KRAS*, *CLO4A2*, *AKT3*, *LAMA2*, and *CYCS* as low-expression genes. All the primers were re-designed using PrimerBLAST to amplify regions <150 bp within the 250 bp stretch from the 3' end of mRNA (Supplementary Table 4) for polyA-based WTA products. The specificity of amplification of each RT-qPCR was evaluated by melting-curve analysis.

**RNA-seq.** Ten nano grams of amplified cDNA was used for library construction with NEBNext Ultra DNA Library Prep Kit for Illumina (New England BioLabs, Massachusetts, USA). The cDNA was sheared by the Covaris system<sup>69</sup>, and then the sheared fragments were end-repaired, A-tailed, and ligated to sequencing adaptors according to manufacturer's protocol. A size selection of about 200 bp was performed before the PCR enrichment (12 cycles). Concentration of the cDNA library was quantified using a Qubit 2.0 fluorometer (Thermo Fisher), and diluted to 2 ng/µl before checking the insert size on an Agilent 2100 and quantifying by qPCR for greater accuracy (library activity > 2 nM). Libraries were subjected to Illumina Hiseq 2500 and 50-base single-ended reads to a depth of 5 M reads per samples were obtained. These experimental procedures were performed by Novogene Corporation.

**Analysis of RNA-seq data.** All raw RNA-seq reads were subjected to trimmomatic-0.36.5 to remove the sequencing adaptors and WTA primers. Trimmed reads were aligned to the human reference genome (hg19) using HISAT2 2.1.0 with default parameters. Aligned reads per genes were counted using HTseq.

Transcriptomic reads were aligned using a GTF file with transcript annotations obtained from GENCODE (release 19, [ftp://ftp.ebi.ac.uk/pub/databases/genecode/Gencode\\_human/release\\_19/genecode.v19.annotation.gtf.gz](ftp://ftp.ebi.ac.uk/pub/databases/genecode/Gencode_human/release_19/genecode.v19.annotation.gtf.gz)). The read counts were divided by the total number of reads and multiplied by one million to form the reads-per-million (rpm). We used rpm rather than rpkm because 3' bias in the alignments was observed as described in a previous study<sup>12</sup>. Gene body coverage was calculated using RSeQC-2.6.5 for RefSeq registered 63,169 transcripts. In addition, we excluded single cells with a high mitochondrial gene ratio (>25%) from further gene expression analysis according to a previous study<sup>68</sup>. Visualization of gene expression data was performed with R software. Heatmap of known target genes was drawn by heatmap.2 package. Unsupervised hierarchical clustering based on the shared nearest neighbor method was performed by Seurat package<sup>49</sup>.

**Statistics and reproducibility.** All statistical analysis except for RNA-seq data analysis were performed by Welch's *t*-test (two tailed) with an alpha level of 0.05. The statistical analysis were performed using stats v3.6.2 package on R software. In RNA-seq analysis, the gene expression was log-normalized for principal component analysis (PCA) and UMAP methods. For gene ontology enrichment analysis, the significantly differential genes ( $p < 0.05$ ) enriched by Seurat were applied to DAVID<sup>70,71</sup>. All *N* values defined in the legends and manuscript refer to biological replicates unless otherwise indicated.

**Reporting summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

### Data availability

All data supporting the findings of this study are available within this article, Supplementary Information and Supplementary Data. We provided an original source data for Fig. 1B, C as Supplementary Data 4 and 5. Source data for Fig. 3A, B as Supplementary Data 6 and 7. Source data for Fig. 4B, C and Supplementary Fig. 6 as Supplementary Data 8. Source data for 5 A as Supplementary Data 9. Source data for

Supplementary Figs. 1–5 as Supplementary Data 10–14. Single-cell RNA-seq data were deposited at [DRA011720](https://doi.org/10.1038/s42003-021-02937-x).

Received: 15 December 2020; Accepted: 1 December 2021;

Published online: 11 January 2022

## References

- Cristofanilli, M. et al. Circulating tumor cells, disease progression, and survival in metastatic breast cancer. *N. Engl. J. Med.* **351**, 781–791 (2004).
- Nagrath, S. et al. Isolation of rare circulating tumour cells in cancer patients by microchip technology. *Nature* **450**, 1235–1239 (2007).
- Carter, L. et al. Molecular analysis of circulating tumor cells identifies distinct copy-number profiles in patients with chemosensitive and chemorefractory small-cell lung cancer. *Nat. Med.* **23**, 114–119 (2017).
- Kalinich, M. et al. An RNA-based signature enables high specificity detection of circulating tumor cells in hepatocellular carcinoma. *Proc. Natl Acad. Sci. USA* **114**, 1123–1128 (2017).
- Yu, M. et al. Ex vivo culture of circulating breast tumor cells for individualized testing of drug susceptibility. *Science* **345**, 216–220 (2014).
- Khoo, B. L. et al. Expansion of patient-derived circulating tumor cells from liquid biopsies using a CTC microfluidic culture device. *Nat. Protoc.* **13**, 34–58 (2018).
- Miyamoto, D. T. et al. RNA-Seq of single prostate CTCs implicates noncanonical Wnt signaling in antiandrogen resistance. *Science* **349**, 1351–1356 (2015).
- Franses, J. W. et al. Pancreatic circulating tumor cell profiling identifies LIN28B as a metastasis driver and drug target. *Nat. Commun.* **11**, 3303 (2020).
- Szczerba, B. M. et al. Neutrophils escort circulating tumour cells to enable cell cycle progression. *Nature* **566**, 553–557 (2019).
- Cheng, Y. H. et al. Hydro-Seq enables contamination-free high-throughput single-cell RNA-sequencing for circulating tumor cells. *Nat. Commun.* **10**, 2163 (2019).
- Gorges, T. M. et al. Accession of Tumor Heterogeneity by Multiplex Transcriptome Profiling of Single Circulating Tumor Cells. *Clin. Chem.* **62**, 1504–1515 (2016).
- Ting, D. T. et al. Single-cell RNA sequencing identifies extracellular matrix gene expression by pancreatic circulating tumor cells. *Cell Rep.* **8**, 1905–1918 (2014).
- Powell, A. A. et al. Single cell profiling of circulating tumor cells: transcriptional heterogeneity and diversity from breast cancer cell lines. *PLoS ONE* **7**, e33788 (2012).
- Aceto, N. et al. Circulating tumor cell clusters are oligoclonal precursors of breast cancer metastasis. *Cell* **158**, 1110–1122 (2014).
- Aceto, N. et al. AR Expression in Breast Cancer CTCs Associates with Bone Metastases. *Mol. Cancer Res.* **16**, 720–727 (2018).
- D'Avola, D. et al. High-density single cell mRNA sequencing to characterize circulating tumor cells in hepatocellular carcinoma. *Sci. Rep.* **8**, 11570 (2018).
- Qi, L. N. et al. Circulating Tumor Cells Undergoing EMT Provide a Metric for Diagnosis and Prognosis of Patients with Hepatocellular Carcinoma. *Cancer Res.* **78**, 4731–4744 (2018).
- Cann, G. M. et al. mRNA-Seq of single prostate cancer circulating tumor cells reveals recapitulation of gene expression and pathways found in prostate cancer. *PLoS ONE* **7**, e49144 (2012).
- Chen, C. L. et al. Single-cell analysis of circulating tumor cells identifies cumulative expression patterns of EMT-related genes in metastatic prostate cancer. *Prostate* **73**, 813–826 (2013).
- Lohr, J. G. et al. Genetic interrogation of circulating multiple myeloma cells at single-cell resolution. *Sci. Transl. Med.* **8**, 363ra147 (2016).
- Garces, J. J. et al. Transcriptional profiling of circulating tumor cells in multiple myeloma: a new model to understand disease dissemination. *Leukemia*, <https://doi.org/10.1038/s41375-019-0588-4> (2019).
- Bray, F. et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* **68**, 394–424 (2018).
- Jim, M. A. et al. Stomach cancer survival in the United States by race and stage (2001–2009): findings from the CONCORD-2 study. *Cancer* **123**, 4994–5013 (2017).
- Liu, D. et al. The patterns and timing of recurrence after curative resection for gastric cancer in China. *World J. Surgical Oncol.* **14**, 305 (2016).
- Saito, M., Kiyozaki, H., Takata, O., Suzuki, K. & Rikiyama, T. Treatment of stage IV gastric cancer with induction chemotherapy using S-1 and cisplatin followed by curative resection in selected patients. *World J. Surg. Oncol.* **12**, doi:Artn 406 <https://doi.org/10.1186/1477-7819-12-406> (2014).
- Okabe, H. et al. Circulating Tumor Cells as an Independent Predictor of Survival in Advanced Gastric Cancer. *Ann. Surg. Oncol.* **22**, 3954–3961 (2015).
- Uenosono, Y. et al. Clinical significance of circulating tumor cells in peripheral blood from patients with gastric. *Cancer Cancer* **119**, 3984–3991 (2013).
- Lee, M. W., Kim, G. H., Jeon, H. K. & Park, S. J. Clinical Application of Circulating Tumor Cells in Gastric Cancer. *Gut Liver* **13**, 394–401 (2019).
- Thanh Huong, P. et al. Emerging Role of Circulating Tumor Cells in Gastric Cancer. *Cancers* **12**, <https://doi.org/10.3390/cancers12030695> (2020).
- Cheng, B. et al. Enumeration And Characterization Of Circulating Tumor Cells And Its Application In Advanced Gastric Cancer. *OncoTargets Ther.* **12**, 7887–7896 (2019).
- Hosokawa, M. et al. Size-selective microcavity array for rapid and efficient detection of circulating tumor cells. *Anal. Chem.* **82**, 6629–6635 (2010).
- Yoshino, T. et al. Manipulation of a Single Circulating Tumor Cell Using Visualization of Hydrogel Encapsulation toward Single-Cell Whole-Genome Amplification. *Anal. Chem.* **88**, 7230–7237 (2016).
- Stevens, M. et al. VyCAP's puncher technology for single cell identification, isolation, and analysis. *Cytom. Part A: J. Int. Soc. Anal. Cytol.* **93**, 1255–1259 (2018).
- Polzer, B. et al. Molecular profiling of single circulating tumor cells with diagnostic intention. *EMBO Mol. Med.* **6**, 1371–1386 (2014).
- Donato, C. et al. Hypoxia Triggers the Intravasation of Clustered Circulating Tumor Cells. *Cell Rep.* **32**, 108105 (2020).
- van der Toom, E. E. et al. Analogous detection of circulating tumor cells using the AccuCyte((R)) -CyteFinder((R)) system and ISET system in patients with locally advanced and metastatic prostate cancer. *Prostate* **78**, 300–307 (2018).
- Sasagawa, Y. et al. Quartz-Seq: a highly reproducible and sensitive single-cell RNA sequencing method, reveals non-genetic gene-expression heterogeneity. *Genome Biol.* **14**, R31 (2013).
- Padovan-Merhar, O. et al. Single mammalian cells compensate for differences in cellular volume and DNA copy number through independent global transcriptional mechanisms. *Mol. Cell* **58**, 339–352 (2015).
- Kanda, M. & Kodera, Y. Recent advances in the molecular diagnostics of gastric cancer. *World J. Gastroenterol.* **21**, 9838–9852 (2015).
- Matsuoka, T. & Yashiro, M. Biomarkers of gastric cancer: Current topics and future perspective. *World J. Gastroenterol.* **24**, 2818–2832 (2018).
- Wang, X. et al. Prognostic value of a 25-gene assay in patients with gastric cancer after curative resection. *Sci. Rep.* **7**, 7515 (2017).
- Ramskold, D. et al. Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells. *Nat. Biotechnol.* **30**, 777–782 (2012).
- Li, Y. L. et al. Clinical significance of phenotyping and karyotyping of circulating tumor cells in patients with advanced gastric cancer. *Oncotarget* **5**, 6594–6602 (2014).
- Zhou, J., Ma, X., Bi, F. & Liu, M. Clinical significance of circulating tumor cells in gastric cancer patients. *Oncotarget* **8**, 25713–25720 (2017).
- Chanrion, M. et al. Concomitant Notch activation and p53 deletion trigger epithelial-to-mesenchymal transition and metastasis in mouse gut. *Nat. Commun.* **5**, 5005 (2014).
- Chuang, Y. C. et al. Blockade of ITGA2 Induces Apoptosis and Inhibits Cell Migration in Gastric Cancer. *Biol. Proced. Online* **20**, 10 (2018).
- Ren, D. et al. Overexpressed ITGA2 promotes malignant tumor aggression by up-regulating PD-L1 expression through the activation of the STAT3 signaling pathway. *J. Exp. Clin. Cancer Res.* **38**, <https://doi.org/10.1186/s13046-019-1496-1> (2019).
- Zheng, G. X. et al. Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.* **8**, 14049 (2017).
- Stuart, T. et al. Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888–1902 e1821 (2019).
- Moss, B. L. et al. Interrogation of inhibitor of nuclear factor kappaB alpha/nuclear factor kappaB (IkappaBalpha/NF-kappaB) negative feedback loop dynamics: from single cells to live animals in vivo. *J. Biol. Chem.* **287**, 31359–31370 (2012).
- Labelle, M., Begum, S. & Hynes, R. O. Direct signaling between platelets and cancer cells induces an epithelial-mesenchymal-like transition and promotes metastasis. *Cancer Cell* **20**, 576–590 (2011).
- Sabrkhany, S., Kuijpers, M. J. E., Oude Egbrink, M. G. A. & Griffioen, A. W. Platelets as messengers of early-stage cancer. *Cancer Metastasis Rev.* **40**, 563–573 (2021).
- Huang, P. et al. Chemotherapy-driven increases in the CDKN1A/PTN/PTPRZ1 axis promote chemoresistance by activating the NF-kappaB pathway in breast cancer cells. *Cell Commun. Signal* **16**, 92 (2018).
- Zheng, L. et al. Inflammation-based indexes and clinicopathologic features are strong predictive values of preoperative circulating tumor cell detection in gastric cancer patients. *Clin. Transl. Oncol.: Off. Publ. Federation Span. Oncol. Societies Natl Cancer Inst. Mex.* **19**, 1125–1132 (2017).
- Lu, R. et al. Detection of circulating stage III–IV gastric cancer tumor cells based on isolation by size of epithelial tumor: using the circulating tumor cell biopsy technology. *Transl. Cancer Res.* **8**, 1342–1350 (2019).
- Bushell, M., Stoneley, M., Sarnow, P. & Willis, A. E. Translation inhibition during the induction of apoptosis: RNA or protein degradation? *Biochemical Soc. Trans.* **32**, 606–610 (2004).



57. Cognart, H. A., Viovy, J.-L. & Villard, C. Fluid shear stress coupled with narrow constrictions induce cell type-dependent morphological and molecular changes in SK-BR-3 and MDA-MB-231 cells. *Sci. Rep.* **10**, <https://doi.org/10.1038/s41598-020-63316-w> (2020).
58. Wirtz, D., Konstantopoulos, K. & Searson, P. C. The physics of cancer: the role of physical interactions and mechanical forces in metastasis. *Nat. Rev. Cancer* **11**, 512–522 (2011).
59. Lapin, M. et al. Single-cell mRNA profiling reveals transcriptional heterogeneity among pancreatic circulating tumour cells. *BMC Cancer* **17**, 390 (2017).
60. Jiang, X. et al. Microfluidic isolation of platelet-covered circulating tumor cells. *Lab a Chip* **17**, 3498–3503 (2017).
61. Zuo, X. X. et al. Platelets promote breast cancer cell MCF-7 metastasis by direct interaction: surface integrin alpha2beta1-contacting-mediated activation of Wnt-beta-catenin pathway. *Cell Commun. Signal* **17**, 142 (2019).
62. Zheng, X. et al. Detection of Circulating Tumor Cells and Circulating Tumor Microemboli in Gastric Cancer. *Transl. Oncol.* **10**, 431–441 (2017).
63. Abdallah, E. A. et al. The Potential Clinical Implications of Circulating Tumor Cells and Circulating Tumor Microemboli in Gastric Cancer. *Oncologist* **24**, e854–e863 (2019).
64. Hosokawa, M. et al. Size-based isolation of circulating tumor cells in lung cancer patients using a microcavity array system. *PLoS ONE* **8**, e67466 (2013).
65. Cristofanilli, M. et al. Circulating tumor cells: a novel prognostic factor for newly diagnosed metastatic breast cancer. *J. Clin. Oncol.: Off. J. Am. Soc. Clin. Oncol.* **23**, 1420–1430 (2005).
66. Sawada, T. et al. Prognostic Impact of Circulating Tumor Cell Detected Using a Novel Fluidic Cell Microarray Chip System in Patients with Breast Cancer. *EBioMedicine* **11**, 173–182 (2016).
67. Hiraiwa, K. et al. Clinical significance of circulating tumor cells in blood from patients with gastrointestinal cancers. *Ann. Surg. Oncol.* **15**, 3092–3100 (2008).
68. Mereu, E. et al. Benchmarking single-cell RNA-sequencing protocols for cell atlas projects. *Nat. Biotechnol.* <https://doi.org/10.1038/s41587-020-0469-4> (2020).
69. Fisher, S. et al. A scalable, fully automated process for construction of sequence-ready human exome targeted capture libraries. *Genome Biol.* **12**, R1 (2011).
70. Huang da, W., Sherman, B. T. & Lempicki, R. A. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* **37**, 1–13 (2009).
71. Huang da, W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **4**, 44–57 (2009).

## Acknowledgements

This work was supported by JST, CREST (JPMJCR14G5), and JST-Mirai Program Grant Number JPMJMI18G9, Japan.

## Author contributions

R.H., F.K., and T.Y. designed the research. R.H., H.Y., T.K., and M.H. performed laboratory experiments, and R.N., T.K., and H.Y. performed sequence alignment and bioinformatic analysis. T.S., F.K., and T.S. designed the clinical study and K.O., Y.O., C.F., A.K., and Y.K. performed patient recruitment. T.T., T.M., and T.Y. supervised the study. R.N., H.Y., T.T., and T.Y. wrote the paper. All authors discussed and contributed to the paper.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s42003-021-02937-x>.

**Correspondence** and requests for materials should be addressed to Tomoko Yoshino.

**Peer review information** *Communications Biology* thanks the anonymous reviewers for their contribution to the peer review of this work. Primary Handling Editor: George Inglis.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022