README File:

Title: **Genetic and phenotypic profiling of single living circulating tumour cells from patients with microfluidics**

Authors: Zaizai Dong, Yusen Wang, Gaolian Xu, Bing Liu, Yang Wang, Julien Reboud, Pawel Jajesniak, Shi Yan, Pingchuan Ma, Feng Liu, Yuhao Zhou, Zhiyuan Jin, Kuan Yang, Zhaocun Huang, Minglei Zhuo, Bo Jia, Jian Fang, Panpan Zhang, Nan Wu, Mingzhu Yang, Jonathan M. Cooper, Lingqian Chang

This readme file aims to provide a guide to use the data associated with this publication, available at DOI 10.1073/pnas.2315168121.

Figure 2:

 “Figure 2.xlsx” contains the data associated with Figure 2b, d, f, h and i as separate sheets:

 In Figure 2b, column 1 is the target (1-3 and all together) included in the sample, whilst columns 2-4 contain the fluorescence intensity for different replicates, in FAM color. Columns 6-8 and 10-12 for Cy3 and Cy5 respectively.

 In Figure 2d, row 2 provides the efficiency (in %) for the condition “delivery” whilst row 3 is for “viability”. Columns 2-4 present the values for A549 cells, whilst columns 6-8 and 10-12 provide the values for HCC827 and 293ft respectively

In figure 2f, the first column provides the experimental conditions (rows 2-4), whilst the columns 2-4 provide the fluorescence intensity for GAPDH, and columns 6-8, that for PD-L1.

In figure 2h, the first column provides the cells used and studied in the experiment (rows 2-4). Columns 2-301 provide the ratio of fluorescence intensities for PD-L1/GAPDH for individual cells for each cells.

In figure 2i, the values of the difference between the Ct value of PD-L1 RNA and the Ct value of GAPDH RNA is provided in columns for each cell line (column 2 for A549 for example).

Figure 3:

“Figure 3.xlsx” contains the data associated with Figure 3d-j as separate sheets

 In figure 3d, the first column provides the time associated with the data for each condition. Columns 2-4 provide the data for the proliferation rate (PR, in %) for HCC827 cells for the 3 replicates (one per column) per time condition (in rows) for the sample “T cell + PD-1 inhibitor”. Columns 6-8, 10-12, 14-16 refer to the samples “T-cell”, “NEED”, and “control” respectively.

Figure 3e follows the same format to present the “inhibition rate” (in %) for the same samples and the same times, whilst

Figure 3f and 3g are the same as 3d and 3e respectively but for 293FT cells.

In Figure 3h, the first column provides the cells used and studied in the experiment (rows 2-4). Columns 2-301 provide the Amplitude of morphological elongation for individual cells for each cells.

In Figure 3i, the values for PR are provided in row 2, in triplicates for different cell lines (indicated in row 1). For example, the values for HCC827 cells are presented in columns 2-4.

Figure 3j is the same as 3i but providing data for IR.

Figure 4:

“Figure 4.xlsx” contains the data associated with Figure 4b-c, e-h as separate sheets.

 Figure 4b provides CTC numbers (column 2 onwards) for two conditions in two rows. Column 1 has headers.

 Figure 4c provides the ratios of the fluorescent intensity of PD-L1 (Cy3 channel) and GAPDH (FAM channel) from individual CTCs in columns, for all patients (column 2-81), 1 column per patient (top row is patient number).

 Figure 4e has the same format as Figure 4c and presents the elongation fold of individual CTCs.

 Figure 4f provides the values for different predictive indexes (in %), one index per row, for different patients, 1 patient per column.

 Figure 4g assembles the data for Receiver Operating Curves for different indexes (one per column, 3-7). The first column is sensitivity, whilst columns 3-7 have the specificity (in %).

 Figure 4h has the different indexes as rows in column 1 and the area under the curve associated with the ROC curve of Figure 4g in column 2 (starting at the second row).

Figure 5:

“Figure 5.xlsx” contains the data associated with Figure 5b-c, and h as separate sheets

 Figure 5b presents the number of CTCs at different timepoints. Column 1 rows 2-4 provide the timepoints headers, row 1 provides the patient number (one per column). The number of CTCs for each patient is provided in each column (starting at row 2).

 Figure 5c provides the NICHE index in the same format as Figure 5b.

Figure 5d presents the ratios of the fluorescent intensity of PD-L1 (Cy3 channel) and GAPDH (FAM channel) for each CTCs as columns. Column 1 is the time header (from row 2). The NICHE index for each time point is provided in row 7 (column 2-5).

Figure 5f is the same as 5d but for a different patient.

Supplementary data.

The file “Figure S1-S16.xlsx” contains the data associated with the supplementary information also available with the published manuscript. Each sheet contains the data for the whole figure.

Figure S1:

Rows 2-3 provide the data for Figure S1d as a frequency distribution of the number of CD45-MBs bound to 1000 Jurkat T cells (row 2) and the count in row 3.

 Rows 6-14 provide the data for Figure S1E. Column 1 is the flow rate (microlitre/min), whilst columns 2-4 are replicates of the WBC removal efficiency (in %)

 Rows 17-19 are for Figure S1f. The column 1 provides the cell types in rows 18 and 19, whilst the columns 2-4 are replicates of the WBC removal efficiency (in %).

Figure S2:

Rows 2-10 provide the data for Figure S2b. Column 1 is the flow rate (microlitre/min), whilst columns 2-4 are replicates of the WBC removal efficiency (in %), starting at row 3.

Rows 13-14 are for Figure S2f. Flow rate distribution across the traps : position in microns in row 13 (starting at column 2), flow rate in row 14.

Figures S2G, H and J follow the same format in rows 17-18, 21-22, and 25-26 respectively.

Rows 29-31 are for Figure S2K: Column 1 is the cell type (starting at row 30), whilst columns 2-4 are replicates of the cell capture efficiency (in %)

Rows 34-40 are for Figure S2m: column 1 (rows 35-40) provide the number of cells spiked and columns 2-4, 6-8, 10-12 provide the triplicate data of the number of captured cells for different buffers, PBS, PBMC and blood respectively.

Figure S3.

Rows 2-4 are for Figure S3c. Ratio of fluorescence intensity of Cy3 channel (row 3) and Cy5 (row 4) to that of FAM channel (normalised to 1 in row 2) for different cells (in columns starting at column 2)

Rows 7-9 provide the same data format for Figure S3d.

Figure S4

 Rows 2-27 are for Figure S4b. row 2 is column headers. Column 1 is time in min. the triplicate data is provided in 3 columns with one column empty between conditions.

 Rows 30-38 are for Figure S4c: row 30 is the wavelength (in nanometers), column 1 provides the conditions of each experiment. The fluorescence value is then provided in columns for each wavelength.

 Figure S4d and 3 follow the same format at S4c in rows 41-49 and 52-60 respectively

 Rows 63-70 are for Figure S4f. Row 63 is the time in min. the data starts in column 2 and rows 64-66 and rows 68-70 are triplicate data for each time and for the specific condition, respectively.

 Figures S4g and S4h follow the same format in rows 73-80 and 83-90 respectively.

 Rows 93-96 provide the data for Figure S4i. row 93 is header for concentrations in mM. Each column (starting at column 2) provide the triplicate data for fluorescence intensity.

 Figures S4j and k follow the same format as S4i, in rows 99-102 and 105-108 respectively.

Figure S5

 Rows 2-4 are for S5A. Row 3 provides the data (starting in column 2) for delivery efficiency using liposomes in triplicate for the cells A549. Data for other cells are in the following columns with one empty column in between each. Row 4 is for incubation condition.

 Rows 7 and 8 are for S5b. row 7 holds the size (in nm, starting column 2) and row 8 is the number of pores at that size.

 Rows 11-13 are for S5c. each row contains the data for 1 membrane (1-3) and columns (starting at 2) provide the number of pores.

 Rows 16-22 are for S5i. rows 17-22 are for each cell. Row 1 is column header (distance in microns). Data is electrical potential difference in V.

 Rows 35-31 are for S5j. Row 25 has headers (time in seconds, starting in column 2). Rows 26-28 and 29-31 are triplicates for NEED and incubation respectively. The data (starting column 2 is fluorescence intensity.

 Rows 34-35 are for S5k. Row 34 provides the fluorescence intensity in triplicates (column 2-4) for NEED, whilst row 35 is for incubation.

Figure S6

 Rows 2-7 are for S6a. column 1 is row headers (here voltage). Row 2 is column headers. Data is in triplicate (column 2-4) for different cell types, separated by an empty column.

 Rows 10-15, 18-23, 26-31, 24-39 follow the same format for Figures S6b-e, respectively.

 Rows 42-44 are for S6f. each row is an independent experiment and fluorescence intensity data is in columns starting from 2 for 100 cells (1 column per cell)..

 Rows 47-50 are for S6g. column 1 has row headers (cell line). Row 47 has column header and the data is presented in triplicate (e.g. column 2-4) for each condition, separated by an empty column.

 Rows 53-56 and 59-62 follow the same format for figures S6h and I respectively.

Figure S7

 Rows 3-4 are for S7B. column 1 has row headers (gene), whilst columns present the average fluorescence intensity without probe for 100 cells and 3 cell types (columns 2-101 are for A549 for example).

 Rows 7-35 are for S7C. the data is arrange in sets of 4 rows separated by an empty row for the different markers and cell types. Each column contains triplicate data for fluorescence intensity for 100 cells (1 per column).

 Rows 38-40 are for S7D. column 1 is row header. Data is in triplicate (e.g. columns 2-4) for each cell type, separated by an empty column.

Figure S8

 Rows 2-4 are for S8A, the first column provides the genes of interest. Columns 2-4 provide the Ct values in triplicates for A549 cells. Other cell types are provided in later columns with a blank column between them.

 Rows 7-9 are for S8B, with conditions in column 1 and triplicate Delta Ct in columns 2-4

 Rows 12-36 are for S8d, rows 13 to 23 provide the data for GAPDH, whilst 25-36 are for PDL1. Rows 13-15 are independent experiments and columns 2-101 provide data for 100 individual cells. Other conditions are provided after an empty row (i.e. next condition is row 17-19).

 Rows 39-49 are for Figure S8E and follow the same format as S8d but provide the ratios.

Figure S9

Rows 2-10 are for Figure S9B, each row is a different target (header in column 1). Results for 100 cells (fluorescence intensity) in columns 2-101. Rows 8-10 are for a different condition.

 Rows 13-16 are for Figure S9C, rows 14-16 provide the target header in column 1, The Ct data in each row is in triplicate for HCC827cells columns 2-4, whilst condition “HCC827+CK SiRNA” is in columns 6-8.

Figure S10

 Rows 2-5 are for Figure S10B, each row provides PR (in %) data for different times (header in column 1), in triplicate over columns (e.g. column 2-4) separated by a blank column between each condition.

 Rows 8-11 are for S10C and follow the same format but for the IR in %.

Figure S11

 Rows 2-5 are for S11B. each row provides data for different times (header in column 1), in triplicate over columns (e.g. column 2-4) separated by a blank column between each condition.

 Other figures (S11C-G) follow the same format with 1 empty row and a title row between each subfigure (e.g. S11C is in rows 8-11)

Figure S12

Rows 2-4 are for S12B and provide the relative quantification of RNA expression in different conditions (in rows, header in column 1), in triplicate over columns 2-4.

Rows 7-9 are for S12C. rows 8 and 9 provide each a target (header in column 1) and results in triplicate over columns (e.g. first condition 2-4) with an empty column between conditions (e.g. second are at 6-8).

Rows 12-14 are for S12E and provide the data for fold of morphological elongation in rows (headers in column 1) for 300 cells (1 per column)

Rows 17-20 are for S12f, each row provides PR (in %) data for different times (header in column 1), in triplicate over columns (e.g. column 2-4) separated by a blank column between each condition.

 Rows 23-26 are for S12g and follow the same format but for the IR in %.

Figure S13

 Rows 2-6 are for figure S13a, the cells under study are provided in rows (header in column 1) whilst the columns provide the intensity of protein levels (protein header in row 2).

 Rows 9-11 are for S13b and provide data for the same proteins but in comparison (condition header in rows 10-11 column 1), as a ratio.

 Rows 14-18 are for S13c and follow the same format as for S13a but for down regulated proteins.

 Rows 21-23 are for S13d and has the data for the same proteins as S13c for the ratio explained for S13b.

Figure S15

 Rows 2-3 are for S15b, row headers are in column 1 and data is provided for mean of PD-L1 expression (as a ratio to GAPDH) in all CTCs from each patient divided into 2 groups (rows).

 Rows 6-8 are for S15c. Patient number is in row 6 (starting column 2). Column 1 has row headers for patient classification. The data is % of CTCs.

 Rows 11-19 are for S15d with again patients in columns (starting from column 1 on row 11). Each column contains the elongation fold of individual CTCs for patients with stage I-II cancer. Note that the number of CTCs for each patient is different so some cells are empty.

 Rows 22-41 are for S15e and follow the same format as d, but presenting the data for patients with stage II-IV cancer.

 Rows 44-47 are for S15f a ROC curve.

Figure S16

 Rows 2-17 are for S16a. row 2 provides the patient number as a header. Row 3 provides the time as header. The data in each column is the ratio of the fluorescent intensity of PD-L1 (Cy3 channel) and GAPDH (FAM channel) from each CTC for that patient at that time. Again patients have different number of CTCs so some cells are empty.

 Rows 20-24 provide the data for S16b. Patient number as column header is in row 20. Column 1 provides the time header. The data for each patient is the NICHE index at different times. Note that for T3, some data was not available so the cells are empty, as patients would come out of the study before T3.