Heliyon 9 (2023) e18305

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Gene expression networks involved in multiple cellular programs coexist in individual hepatocellular cancer cells

Jin Zhao [2], Ran Lu [2], Chen Jin, Siying Li, Yulin Chen, Qiaorong Huang, Xue Li, Wentong Meng, Hong Wu [*], Tianfu Wen [**], Xianming Mo [***,1]

Department of Liver Surgery, Laboratory of Stem Cell Biology, State Key Laboratory of Biotherapy, West China Hospital, Sichuan University, Chengdu, China

ARTICLE INFO

Keywords:
Hepatocellular carcinoma
Gene expression networks
Oncogene and anti-oncogene expression
Balanced combination of gene expressions
Individual cancer cell
Cellular program

ABSTRACT

The gene expression networks of a single cell can be used to reveal cell type- and condition-specific patterns that account for cell states, cell identity, and its responses to environmental changes. We applied single cell sequencing datasets to define mRNA patterns and visualized potential cellular capacities among hepatocellular cancer cells. The expressing numbers and levels of genes were highly heterogenous among the cancer cells. The cellular characteristics were dependent strongly on the expressing numbers and levels of genes, especially oncogenes and anti-oncogenes, in an individual cancer cell. The transcriptional activations of oncogenes and anti-oncogenes were strongly linked to inherent multiple cellular programs, some of which oppose and contend against other processes, in a cancer cell. The gene expression networks of multiple cellular programs proliferation, differentiation, apoptosis, autophagy, epithelial-mesenchymal transition, ATP production, and neurogenesis coexisted in an individual cancer cell. The findings give rise a hypothesis that a cancer cell expresses balanced combinations of genes and undergoes a given biological process by rapidly transmuting gene expressing networks.

1. Introduction

Cancer is a heterogeneous disease and displays differing histologic appearances defined as distinct subtypes of cancers from the same primary site of origin among patients [1]. Accumulated molecular and genetic data have modified the histology-based definitions of cancer subtypes and have revealed the tumor cell heterogeneity [2]. Human cancer has been documented comprehensively involving genomic and epigenetic alterations, including single-nucleotide variants (SNVs), small insertions or deletions, somatic mutations, gene copy number alterations, chromatin structural variants, chromatin regulator mutations, DNA methylation changes, and altered compositions of enhancers and repressors [3,4]. In addition, many cancer cells carry mutations in protein that modulate chromatin structures and activity of enhancers or repressors crossing genomes. For instance, in 38 types of human cancers, early oncogenesis is characterized by mutations in a constrained set of driver genes. A nearly fourfold diversification of driver genes and

* Corresponding author.
** Corresponding author.
*** Corresponding author.
E-mail addresses: wuhong@scu.edu.cn (H. Wu), wentianfu@scu.edu.cn (T. Wen), xmingmo@scu.edu.cn (X. Mo).
[1] leading contact.
[2] These authors contributed equally.

https://doi.org/10.1016/j.heliyon.2023.e18305
Received 28 October 2022; Received in revised form 11 July 2023; Accepted 13 July 2023
Available online 17 July 2023

increased genomic instability are features of later stages [5,6]. The tumor cell heterogeneity has been revealed across whole-genome sequences of 2,658 cancer samples spanning 38 cancer types. The sequencing data have revealed positive selection for subclonal driver gene mutations and copy number alterations as well as dynamic changes in gene mutational processes [7]. Furthermore, data derived from DNA methylation, bulk gene expression profiling, spatial gene expression profiling, and single-cell sequencing in many kinds of human tumors have also revealed multiple clones and subclones of cancer cells with distinct genomic and epigenetic mutations and alterations in samples of patients and animal models [8].

Genomic and epigenetic mutations and alterations result in gene transcriptional changes, especially in the transcription trans-mutations of oncogenes and anti-oncogenes [9]. Transcription of the eukaryotic genome is carried out by nuclear RNA polymerase I (Pol I), Pol II, and Pol III. Pol I transcribes the rRNA precursor, Pol III transcribes small non-coding RNAs, and Pol II transcribes protein-coding genes to produce mRNAs and some non-coding RNA. Pol II transcriptions involve tight control of multiple steps in the transcription cycle to produce cell type- and condition-specific patterns of mRNAs that account for cell differentiation, the maintenance of cell identity and the responses of cells to environmental changes [10,11]. Intensive bulk expression profiles reveal tumor cell heterogeneity based on Pol II transcriptional subtypes of cancer cells [12,13]. Gene expression-based cancer cell subtypes are nicely enriched for selected genetic events in cancer tissues in human patients. Within the same cancer specimen, the expressing subtypes of mRNA are also varied as multi-regions in a tumor. Multiple subtypes of cancer cells can be visualized in different regions of the same tumor. Temporal analyses have demonstrated that gene expression-based subtypes of cancer cells are altered over time and through therapies in patients [14,15]. Based on transcriptional patterns of protein-coding genes, single-cell RNA-sequencing (scRNA-seq) shows that distinct cells recount regimens of distinct subtypes and provides an approach to explore genetic and functional heterogeneity at a level of an individual cell in a tumor [16,17]. The gene expressing patterns of single cells typically reveal cell types with highly distinct clusters and cell states further heterogeneity within each cancer cell cluster in the same cancer tissue. Cell types and states are then distinguished based on the identity of preferentially expressed genes that can be used to resolve malignant from non-malignant cells, provide new insights into tumor composition, cancer stem cells and impact on crucial aspects of tumor progression, including drug resistance, metastasis, and self-renewal in human cancers [18,19]. Hepatocellular carcinoma (HCC) has high molecular heterogeneity documented in rich publications and has been identified to have different molecular subclasses that broadly categorize inter-patients, among tumor nodules within the same patient, and intra-tumor, between different regions of the same tumor [20–22]. Furthermore, the single-cell resolution has provided tremendous advances in characterizing the genetic groundwork of HCC heterogeneity [17]. Thus, HCC is a nice model for further portraying a detailed inter-patients and intra-tumor map of cell identities and states based on the singularity of preferentially gene expressed patterns in a cancer cell.

2. Results

2.1. The heterogeneous gene expression patterns in cancer cells of hepatocellular carcinoma

To depict the intrinsic characteristics of cancer cells, we collected 6 samples of human hepatocellular carcinoma from 5 patients and dissociated each sample into single cells. The acquired single cell suspensions were then sorted by fluorescence-activated cell sorting (FACS) to enrich $CD45^{-}$ cell populations to generate single-cell labeled cDNA libraries with Singleron GEXSCOPE for sequencing. The procedures eliminated the immune cells, which consisted of considerably large cell populations in cancer tissues, to increase the ratio of the cancer cell population. We obtained single-cell transcriptomes of 38,077 cells after quality controls defined by gene expressing patterns and mitochondrial gene expressing patterns detected in each cell. Cells with more than 20% mitochondrial gene expression were excluded from the dataset. Principle-component analysis (PCA) and t-distributed stochastic neighbor embedding (tSNE) analysis were conducted to construct the panoramic view of hepatocellular carcinoma tissues. We identified multiple cell types including hepatocellular carcinoma cells, hepatic stellate cells (HSCs), endothelial cells (ECs), plasma cells, natural killer cells (NKs), and macrophages (MPs) (Figs. S1A and B) in the cancer tissues. After excluding other cell types, 30,429 cancer cells of the 6 human hepatocellular carcinoma samples were further analyzed. The heterogeneity of cancer cells derived from each sample was quite recognizable via differentially expressed genes (DEGs) among cancer cells in the cancer tissues (Fig. S1C). Comprehensive analysis of copy number variations (CNV) showed that the gene expressing regions of chromatins in cancer cells of an individual sample manifest a unique pattern. A distinctive CNV pattern was visualized in each cancer cell in a sample of cancer tissues (Fig. S1D). Detection of highly variable genes (HVG) showed exclusive patterns of gene expression in cancer cells with distinctive signaling pathways in each cancer tissue (Figs. S1E, S1F, S2A and data not shown). The cancer cells of each sample displayed high heterogeneity and were grouped into a distinctive population (Fig. S2A). All the assays show that human hepatocellular carcinoma (HCC) has high molecular heterogeneity of inter-samples and intra-sample, consistent with previous observations [23,24].

2.2. Random distribution of oncogenes and anti-oncogenes in cancer cells of HCC

We analyzed in detail the expressions of a panel of genes, which have previously been shown to be the most common molecular anomalies to drive cancer development and progresses in human hepatocellular carcinoma. The expression of such a gene showed random distribution among cancer cells in a sample (Figs. S2A and B). The ratio of cells expressing a given gene was highly varied inter-samples and intra-sample of human hepatocellular carcinoma. For instance, *KRAS* had relatively high levels of expression in three samples, but a small fraction of cells were still negative. Notably, there were no clear patterns in the distribution of positive and negative cells in cell populations and in the sections of cancer tissues from a given sample. *TP53*, *KEAP1*, and *ARID1A* were expressed only in a small fraction of cancer cells in all samples detected. *CCND1*, *CTNNB1*, *PTEN*, and *CDKN2A* were also expressed in all samples

but carried different ratios of positive cells in cancer tissues. We then extracted cancer cells from one of the patients for verification. The scatter plots of gene expression showed the anti-oncogene *KLF6*, *ELF3*, and the oncogene *JUN* exhibit scattered random distributions. Furthermore, immunofluorescent staining showed that the cancer cells with protein expression of these genes also exhibited scattered random distributions in cancer tissue sections from a given sample (Fig. S2C, data not shown). We compared more samples by

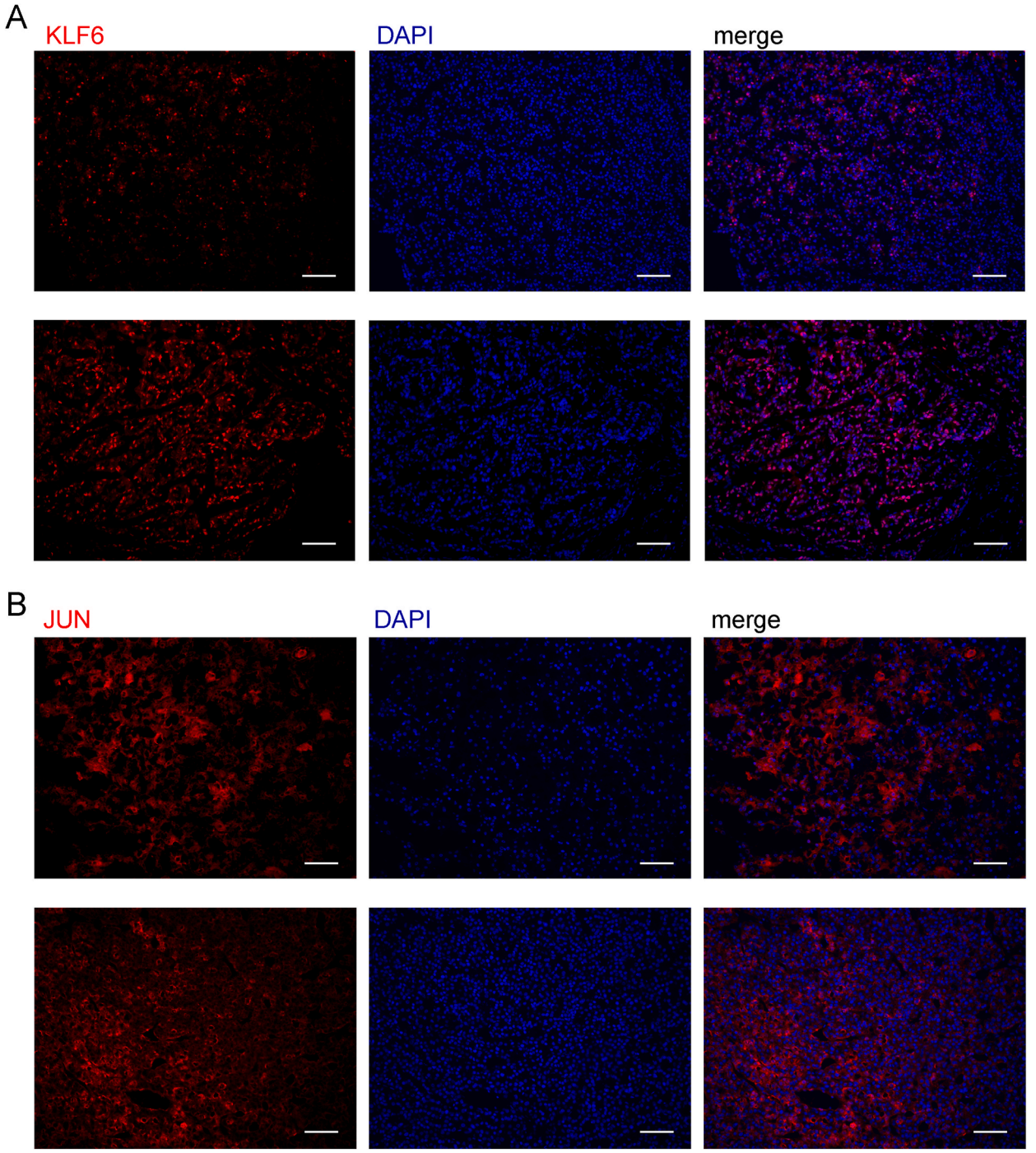
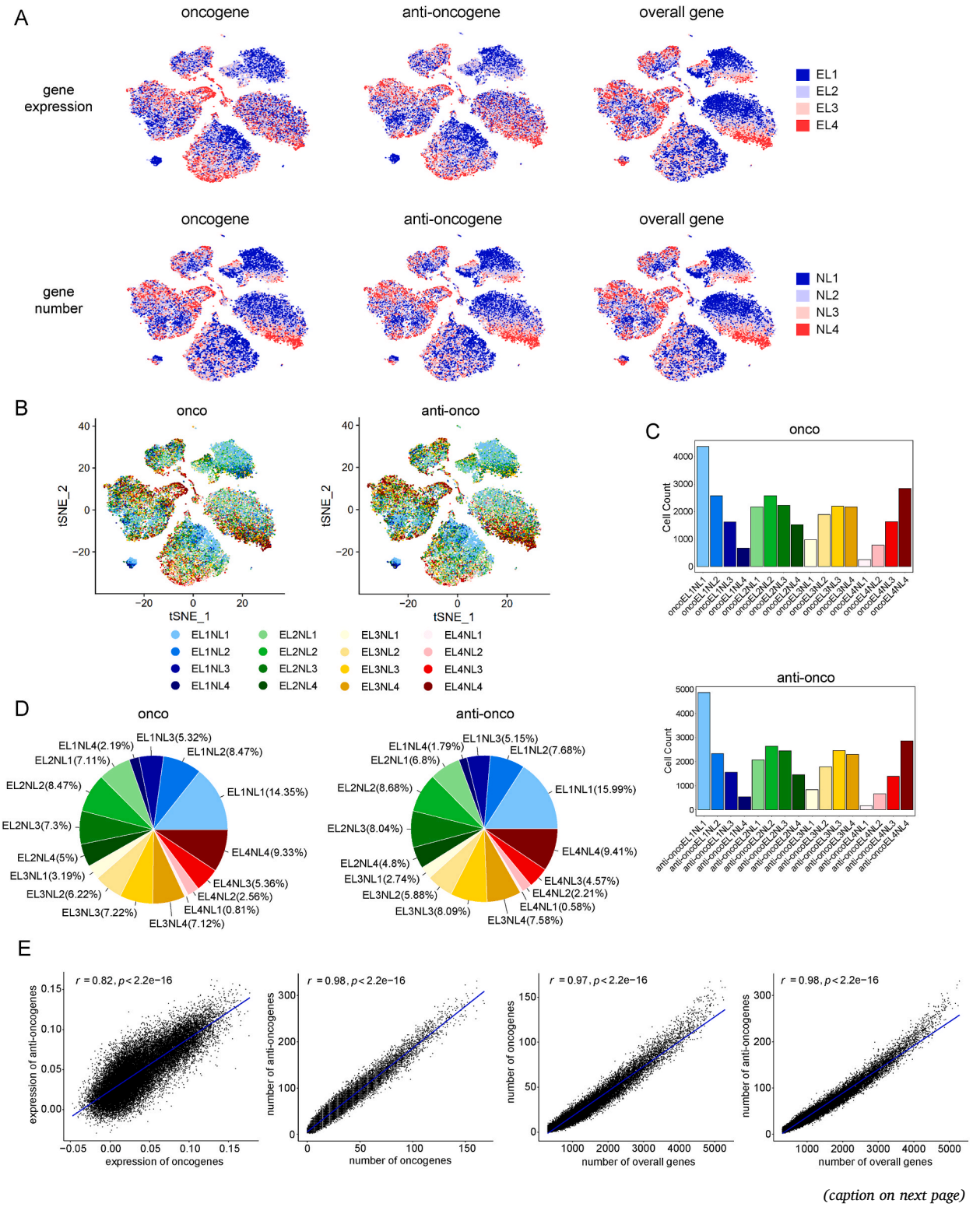


Fig. 1. The protein expression level of anti-oncogene KLF6 and oncogene JUN among different cancer tissues of hepatocellular carcinoma. **A**. The top and bottom rows show the immunofluorescence staining results of anti-oncogene KLF6 (red) among different cancer tissues of hepatocellular carcinoma. Nuclei are blue (DAPI). Scale bar indicates 100 μm. **B**. The top and bottom rows show the immunofluorescence staining results of oncogene JUN (red) among different cancer tissues of hepatocellular carcinoma. Nuclei are blue (DAPI). Scale bar indicates 100 μm. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)



(caption on next page)

Fig. 2. The potentialities of oncogene and anti-oncogene expression in cancer cells of HCC. A. The tSNE projections show all cancer cells derived from 6 HCC samples, and we mapped the expression levels (ELs) and number levels (NLs) of oncogene, anti-oncogene or overall gene in it. **B.** The tSNE projections show 16 categories of oncogene or anti-oncogene obtained by combining 4 gene expression levels and 4 number levels. **C.** The bar plots show the count of cells of 16 classifications of oncogene or anti-oncogene. **D**. The pie charts show the percentage of cells of 16 classifications of oncogene or anti-oncogene. **E**. The first two scatter plots show the correlations between the expression level or expression number of oncogenes or anti-oncogenes, respectively. The last two scatter plots show the correlations between the expression number of overall genes and the expression number of oncogenes or anti-oncogenes, respectively. All plots with Pearson correlation coefficient (*r*-value) and *p*-value labeled at the top.

detecting the cancer cells with the expression of anti-oncogene KLF6 (a nuclear-localized protein) (Fig. 1A) and the oncogene JUN (a cytoplasmic-localized protein) (Fig. 1B). The results clearly revealed the random distribution features of positive cells in cancer tissues that were examined. No clear distribution patterns of positive and negative cells were detected in cancer tissues of HCC. Taken together, the examination of the expression of recognized driver genes as well as additional oncogenes and anti-oncogenes in human hepatocellular carcinoma revealed that the cancer cells with the expression of a given gene were randomized distributed in cancer tissues. The assays indicate that cancer cells do not constantly carry the expression of a specific driven gene detected anomalies and suggest that anomalous genes might substitutively express to drive or maintain the malignant identity in a cancer cell in human hepatocellular carcinoma.

2.3. The potentialities of oncogene and anti-oncogene expression in cancer cells of HCC

In order to thoroughly explore all oncogenes and anti-oncogenes expressed in individual cancer cells, we started a new definition and classification based on the gene expressing number and expressing level in a cancer cell. First, the expressing levels and expressing numbers of oncogenes, anti-oncogenes, and overall genes expressed in each single cancer cell were measured and counted separately within each sample, and the expression level (EL) and number level (NL) were classified into four classes according to their respective quartiles (Fig. 2A). EL or NL1 to 4 indicated increasing expressing level or number of genes, respectively. We identified similar tendencies in the distribution of oncogenes, anti-oncogenes, and overall genes in terms of gene number rather than gene expression. The expressing number of overall genes showed a strong correlation with the expressing numbers of oncogenes and anti-oncogenes. The results indicate that the number of oncogenes and anti-oncogenes expressed in cells was affected to an intense extent by the total number of expressed genes. Combining the EL and NL labels of oncogenes or anti-oncogenes respectively in an individual cell, the cancer cells of human HCC were clustered into 16 groups (Fig. 2B). After counting the number of cancer cells in the 16 clusters, cancer cells with the lowest expressing level and the lowest expressing number of oncogenes (defined as oncoEL1NL1) consisted of the largest population (14.35%). Cancer cells with the highest expressing level and the highest expressing number of oncogenes (defined as oncoEL4NL4) accounted for the second largest population (9.33%). It was also true for the cancer cells with anti-oncogenes expression. Cancer cells with the lowest expressing level and the lowest expressing number of anti-oncogenes (defined as anti-oncoEL1NL1) accounted for the biggest population (15.99%). The cancer cells with the highest expressing level and the highest expressing number of anti-oncogenes (defined as anti-oncoEL4NL4) accounted for 9.41%, the second biggest population of cancer cells (Fig. 2C and D). The results suggest that cancer cells tend to group into two cellular states: either high activations in gene expression or high silences in gene expression in the cancer cells of human HCC. Cancer cells in other cellular states are more likely transiting states in cancer tissues.

Based on the number of genes expressed in an individual cell, the expressing number of anti-oncogenes was strongly correlated to the expressing number of oncogenes in a cancer cell (Fig. 2E). At the expressing level of genes, the expressing levels of anti-oncogenes also showed a nice correlation with the expressing levels of oncogenes (Fig. 2E). The results indicate that cancer cells that are easier to express oncogenes had a higher possibility to express anti-oncogenes. The measurements suggest that the expressing numbers and levels of oncogenes and anti-oncogenes are the consequence of gene transcription activation in a cancer cell.

2.4. The potentialities of proliferation and hepatocytic differentiation in cancer cells of HCC

Next, we detected the proliferation status of cancer cells via gene expression patterns. The results showed that the proliferating potentialities of cancer cells were heterogeneous among samples (Fig. 3A). We also measured the hepatocytic signatures of a cancer cell to ascertain the differentiation properties of cancer cells in HCC. Based on the averaged expression of genes, the hepatocytic signatures were scored in each cancer cell (Fig. 3A). In contrast to the proliferative potential, the distribution of the hepatocytic signatures, i.e., differentiation, showed a clear tendency in tSNE projection. The results suggest that a significant number of hepatocytic signature-related genes are involved in the formation of the dimension reduction results. Additionally, correlation analysis showed a clear negative correlation between differentiation and proliferation of cancer cells (Fig. 3B). The cancer cells with differentially hepatocytic signatures were able to divide into four subtypes according to the quartiles of scores. A subtype of cancer cells with very low hepatocytic signatures (hepaEL1), a subtype of cancer cells with low hepatocytic signatures (hepaEL2), a subtype of cancer cells with high hepatocytic signatures (hepaEL3), and a subtype of cancer cells with very high hepatocytic signatures (hepaEL4). In the meantime, we divided the cells with different proliferative capacities into four levels using the quartiles of gene expression networks of proliferation in each cell. Cells at levels from level1 to level4 displayed increasing capabilities to grow. The results confirmed that hepaEL1 did exhibit higher proliferative potential while the cancer cells with the highest hepatocytic signatures exhibited the lowest abilities to grow. (Fig. 3C). To be more convincing, we scored the potentialities of fatty acid metabolism and cellular response to the

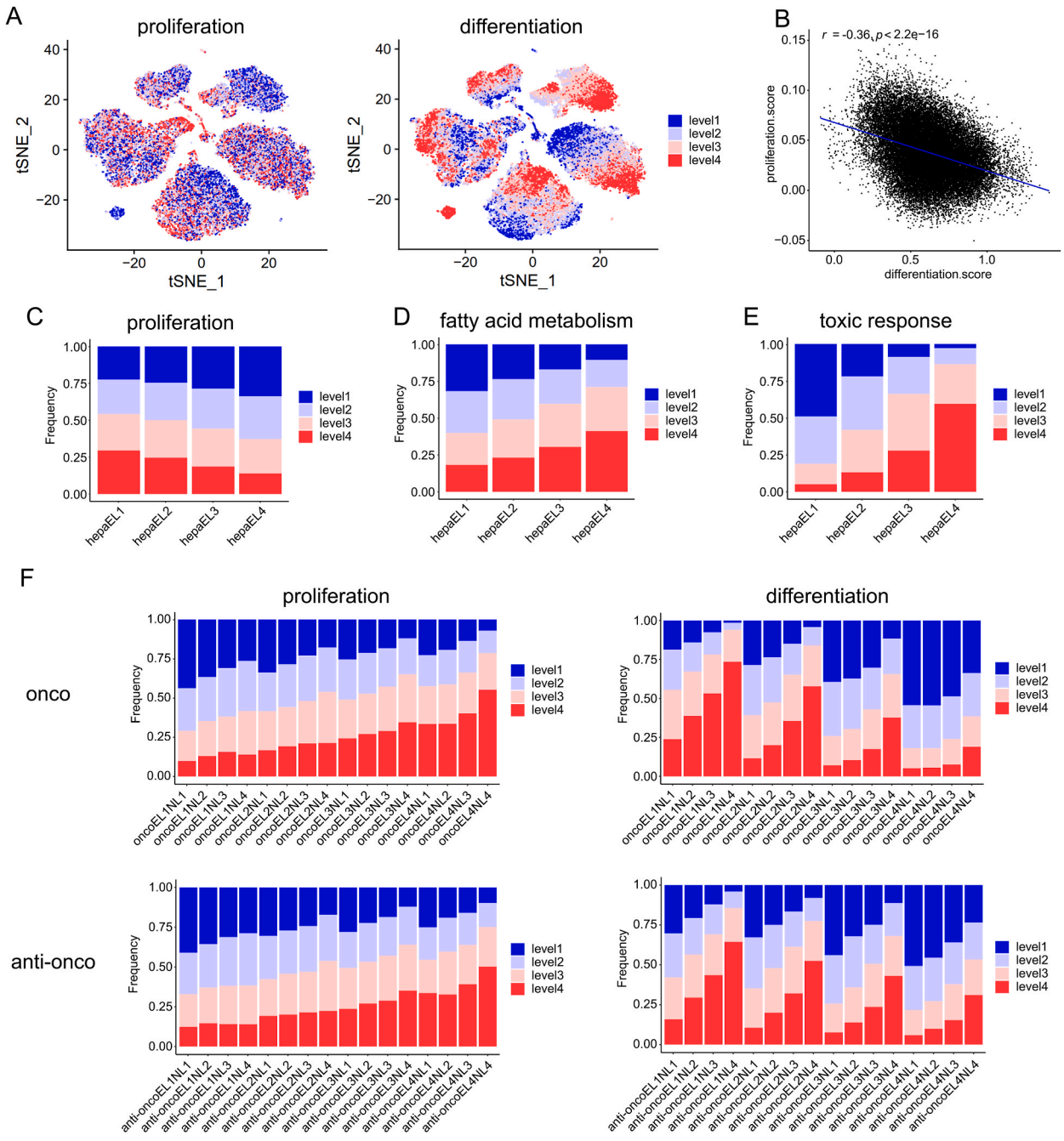
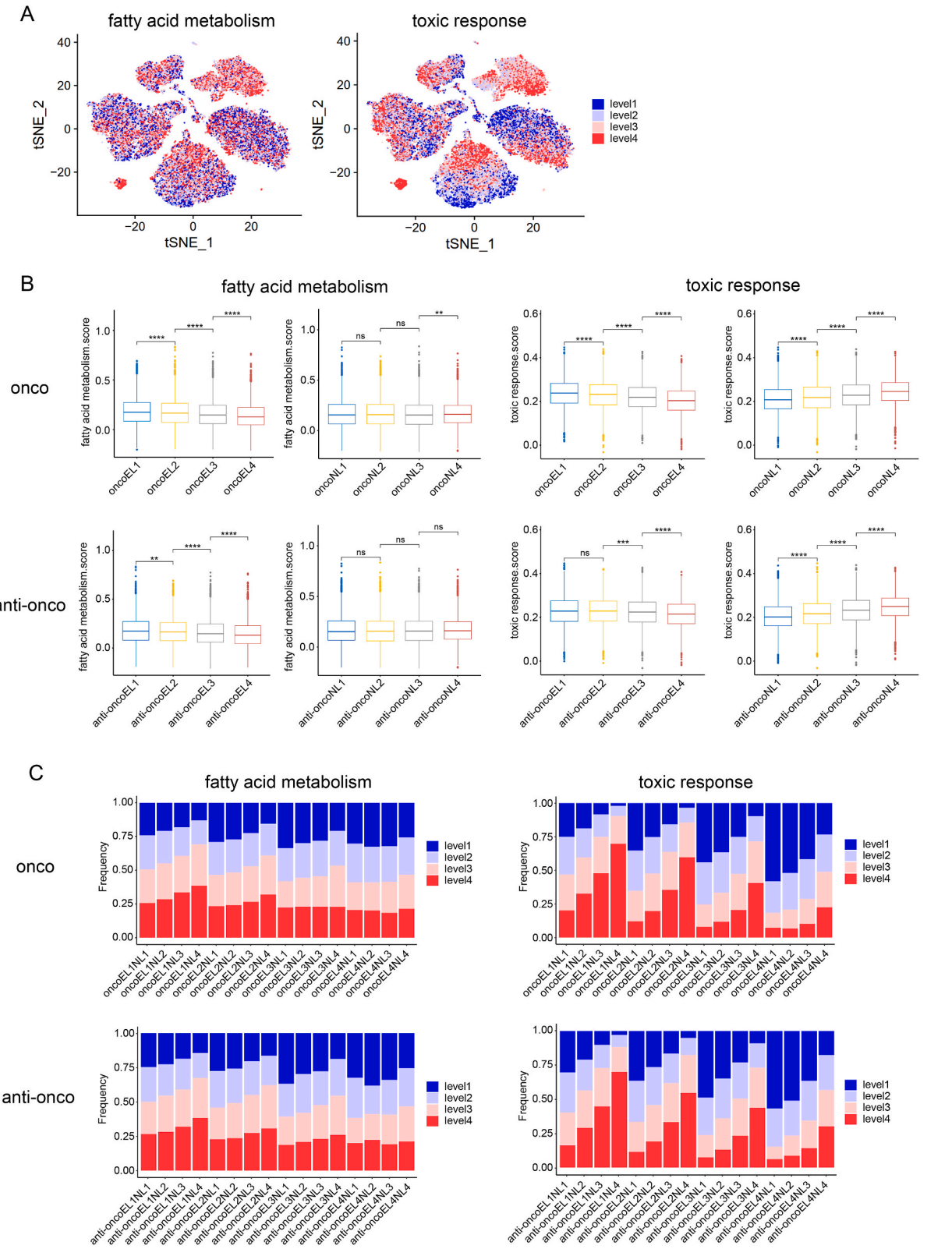


Fig. 3. The potentialities of proliferation and hepatocytic differentiation in cancer cells of HCC. A. The tSNE projections show 4 levels of cell proliferation and hepatocytic differentiation, respectively. **B**. The scatter plot shows the correlation of differentiation. score and proliferation. score with Pearson correlation coefficient (*r*-value) and *p*-value labeled at the top. **C.D.E.** The stacked bar charts show the percentage of 4 levels of proliferation, fatty acid metabolism, and toxic response among cells with different expression level of hepatic signatures (hepaEL1 to hepaEL4 represent increased cell maturity). **F**. The stacked box plots show the differences in proliferation. score or differentiation. score among cells defined with different NLs or ELs of oncogenes or anti-oncogenes.

toxic substances, which represent the physiological functions of hepatocytes, in cancer cells and classified all cells into 4 categories based on the quartiles of the scores. The results showed the active metabolism of fatty acid and toxic response were highly correlated with hepatocytic signatures in a cancer cell (Fig. 3D and E).

The effects of the expression level and gene number of oncogenes and anti-oncogenes on the proliferation and differentiation of cancer cells were assessed. The result charts indicated that the expression level and gene number of both oncogenes and anti-oncogenes



(caption on next page)

Fig. 4. The potentialities of fatty acid metabolism and toxic response in cancer cells of HCC. A. The tSNE projections show 4 levels of fatty acid metabolism and toxic response potentialities, respectively. **B**. The box plots show the differences in fatty acid metabolism. score or toxic response. score among cells defined with different NLs or ELs of oncogenes or anti-oncogenes. Statistical differences are indicated by asterisks. *p*-values were calculated by Wilcoxon test. ns: not significant. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$, $^{****}p < 0.0001$. **C**. The stacked bar charts show the percentage of 4 levels of fatty acid metabolism or toxic response among cells with 16 classifications of oncogenes or anti-oncogenes, respectively.

influenced the proliferation ability of a cancer cell. An increase in the expression level and gene number of both oncogenes and anti-oncogenes was associated with increased proliferation abilities of cancer cells. The expressing number of oncogenes or anti-oncogenes also significantly and profoundly influenced the degree of cell differentiation. An increase in the expressing gene number of both oncogenes and anti-oncogenes was associated with increased differentiation abilities of cancer cells. When ignoring the expressing number differences in oncogenes or anti-oncogenes, the higher expression levels of oncogenes or anti-oncogenes were always associated with the lower differentiation index in cancer cells (Fig. 3F, Fig. S3A).

In addition, we monitored cell cycle phases of all cancer cells (Fig. S3B). When measuring the expressing levels of oncogenes, a cancer cell with higher expressing level of oncogenes was more probable to undergo mitotic phases. When determining the expressing numbers of oncogenes, a cancer cell with more expressing oncogenes had less probability to undergo into G2/M phase. The cells had almost equal probability to undergo into S phase, regardless of the expressing numbers of oncogenes. The expressing levels of oncogenes increased the probability of a cancer cell going into the G2/M phase. In contrast, when measuring the expressing level of anti-oncogenes and expressing number of anti-oncogenes, a cancer cell with higher expressing anti-oncogenes had a larger probability to go into the G2/M phase, and a cell with less expressing anti-oncogenes was more likely to undergo mitotic phases. Taken together, the assays indicate that a cancer cell carrying fewer expressing oncogenes/anti-oncogenes with higher expressing levels had a larger probability to undergo mitotic cell cycle in cancer tissues of HCC.

2.5. The potentialities of fatty acid metabolism and toxic response in cancer cells of HCC

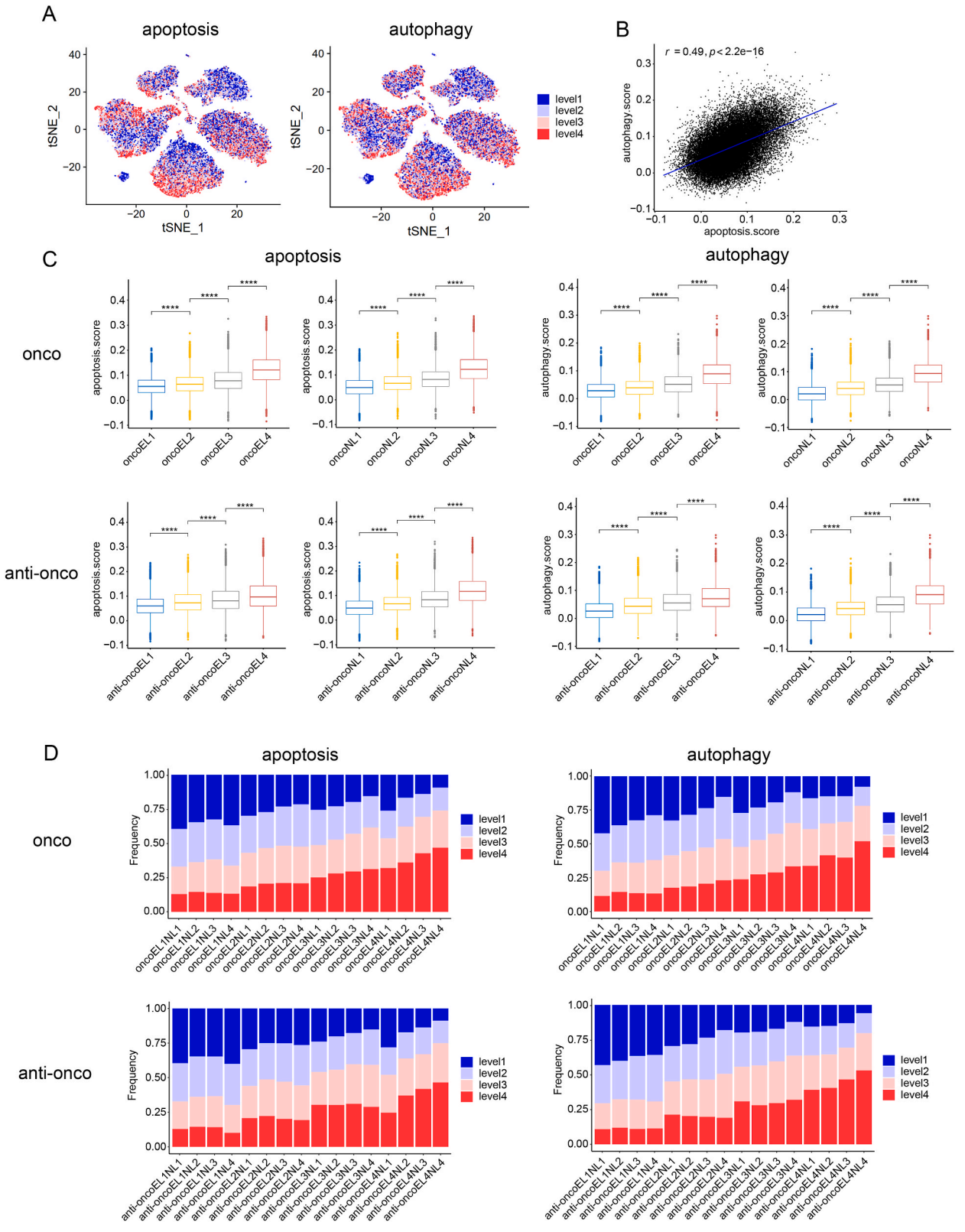
As mentioned above, we scored the potentialities of fatty acid metabolism and toxic response using corresponding gene sets. Subsequently, we classified the cancer cells into 4 levels based on their respective quartiles of the scores (Fig. 4A). From level1 to level4 indicates an increasing capacity of fatty acid metabolism or cellular response to toxic. The analysis showed that a cancer cell with lower expression level of oncogenes/anti-oncogenes manifested stronger fatty acid metabolism and toxic response, but the expressing gene number of oncogenes/anti-oncogenes did not show an influence on the cellular fatty acid metabolism in cancer cells (Fig. 4B and C). In contrast, the potentiality of cellular response to toxic could be facilitated with increased expressing numbers of oncogene/anti-oncogene (Fig. 4B and C).

2.6. The potentialities of apoptosis and autophagy in cancer cells of HCC

Previous observations show that programmed cell death by apoptosis serves as a natural barrier to cancer development [25]. The cancer cells need to overcome the barrier of cell death to expand cancer tissues. Thus, we evaluated gene expressing patterns of both upstream regulators and downstream effector components in apoptotic machinery and weighed the probability of apoptosis in a cancer cell (Fig. 5A). We also measured the gene expressing patterns of regulatory and effector components in autophagy machinery and showed that autophagic phenotypes manifested similar appearances of apoptosis in cancer cells of hepatocellular carcinoma (Fig. 5A). Autophagy was highly correlated to apoptosis in a cancer cell (Fig. 5B). The expressing levels of oncogenes and anti-oncogenes were correlated with apoptotic signatures and autophagic signatures in a cancer cell (Fig. 5C). A cancer cell with higher expressing levels of oncogenes or anti-oncogenes had an increased probability to be subjected to apoptosis or autophagy. The increasing number of expressing oncogenes or anti-oncogenes in a cancer cell was also related to apoptosis or autophagy (Fig. 5C). When examining the tendency of apoptosis or autophagy in cells labeled by 16 classifications of oncogenes or anti-oncogenes, we identified that both the expressing level and the expressing number of oncogenes or anti-oncogenes played influential roles on the apoptosis or autophagy in cells (Fig. 5D). Enhancing oncogene/anti-oncogene expressing levels along with increasing expressing numbers of oncogene/anti-oncogene in a cancer cell were highly related to cellular apoptosis and autophagy. A cancer cell with the highest expressing level of oncogene/anti-oncogene plus the most expressing number of oncogene/anti-oncogene had the most probability to undergo apoptosis or autophagy.

2.7. The potentialities of energy metabolism in cancer cells of HCC

Cancer cells have been shown to alter energy metabolism to support their growth and division in cancer tissues [26]. In this way, we evaluated the gene expressing networks involved in the processes of glycolysis, tricarboxylic acid (TCA) cycle, ATP production, and oxidative phosphorylation, key programs of energy metabolism in a cell (Fig. 6A). When comparing changes in energy metabolism in cells with 16 classifications of the expression of oncogenes or anti-oncogenes, we found strikingly uniform tendencies in four types of energy metabolisms. When the oncogenes or anti-oncogenes were expressed at a given level, the more oncogenes or anti-oncogenes expressed in a cancer cell, the cancer cell carried a higher energy metabolism level. While the lower expression level of oncogenes or anti-oncogenes represented higher potentialities of energy metabolisms (Fig. 6B, Fig. S4). Thus, the cells labeled with EL1NL4 manifested the highest activations of energy metabolisms.



(caption on next page)

Fig. 5. The potentialities of apoptosis and autophagy in cancer cells of HCC. A. The tSNE projections show 4 levels of apoptotic and autophagic potentialities, respectively. **B**. The scatter plot shows the correlation of apoptosis. score and autophagy. score with Pearson correlation coefficient (*r*-value) and *p*-value labeled at the top. **C**. The box plots show the differences in apoptosis. score or autophagy. score among cells defined with different NLs or ELs of oncogenes or anti-oncogenes. Statistical differences are indicated by asterisks. *p*-values were calculated by Wilcoxon test. ns: not significant. $*p < 0.05$, $**p < 0.01$, $***p < 0.001$, $****p < 0.0001$. **D**. The stacked bar charts show the percentage of 4 levels of apoptosis or autophagy among cells with 16 classifications of oncogenes or anti-oncogenes, respectively.

2.8. The potentialities of epithelial-mesenchymal transition and neurogenesis in cancer cells of HCC

Epithelial-mesenchymal transition (EMT) is crucial for malignant progression [27]. During EMT, a new transcriptional program is activated to promote the mesenchymal fate in epithelial cells. In the cancer context, EMT confers on cancer cells increased tumor-initiating and metastatic potential and greater resistance to elimination by several therapeutic approaches. Recent evidence has shown that cancer cells express neurotrophic markers such as nerve growth factors, brain-derived neurotrophic factor, glial cell-derived neurotrophic factor, and axon guidance molecules to promote neurogenesis in cancer tissues [28,29]. Thus, we examined the gene expressing networks involved in EMT and neurogenesis in a hepatocellular cancer cell (Fig. 7A). First, the correlation analysis indicated a strong correlation between neurogenesis and EMT in a cancer cell (Fig. 7B). Next, we examined whether the expressing levels and the expressing numbers of oncogenes or anti-oncogenes played roles in the potentialities of EMT and neurogenesis in cancer cells. The box plots showed that the expressing levels of oncogenes/anti-oncogenes mainly influenced the potentiality of EMT in a cancer cell (Fig. 7C). Increasing expressing levels and expressing numbers of oncogenes/anti-oncogenes significantly enhanced the neurogenesis in a cancer cell (Fig. 7C). From the perspective of 16 classifications, higher expressing level and less expressing number of oncogenes/anti-oncogenes gave a cell carrying higher potential of EMT and neurogenesis (Fig. 7D).

2.9. Similarities of cellular potential patterns between two samples derived from one patient

We also explored additional cellular programs including ROS pathway, redox homeostasis, peroxisome, stemness, senescence, inflammatory response, and angiogenesis to verify cell states and cell identities in cancer cells of HCC (data not shown). Because of the heterogeneity of these cellular programs or responses among samples, the combined results of all data of samples did not show clear tendencies. Thus, we analyzed the data of each sample and identified the same potentials of cellular characters including ROS pathway, redox homeostasis, peroxisome, stemness, senescence, inflammatory response, and angiogenesis in a cancer cell (data not shown, Fig. 8, S5). For instance, the results of the primary and recurrent samples of patient P5 (primary: P5-1; recurrent: P5-2) showed high similarities in cellular potential patterns (Fig. 8, S5). The correlation heatmaps showed that, in good agreement with the previous combined results, the total gene expressing numbers of cancer cells in both samples were highly correlated with the expressing numbers of oncogenes and anti-oncogenes. Besides, in both samples, the hepatocytic signatures, such as differentiation, fatty acid metabolism, and toxic response, showed a significant negative correlation with the expression levels of oncogenes and anti-oncogenes. While cellular programs such as apoptosis, autophagy, stemness, EMT, senescence, inflammatory response, and neurogenesis showed a positive correlation with the expressing levels of oncogenes/anti-oncogenes. In contrast, angiogenesis was not significantly correlated with the expression levels of oncogenes/anti-oncogenes but had a clear positive correlation with EMT. We also identified that the recurrent sample showed higher correlations (higher *r*-values) between the cellular potentials and gene expression in almost all aspects. For example, the *r*-values of the correlation between three hepatocytic signatures (differentiation, fatty acid metabolism, and toxic response) and oncogene expression levels in sample P5-1 were −0.78, −0.32, and −0.68, respectively. While, in cancer cells of P5-2, these *r*-values increased to −0.88, −0.65, and −0.86. This tendency was also true for aspects with a positive correlation. For instance, the *r*-values of the correlation between apoptosis and autophagy with oncogene expression levels in sample P5-1 were 0.55 and 0.46. But in P5-2, the *r*-values increased to 0.62 and 0.69.

In summary, to show the potential of the classifications of cells more succinctly based on the expressing level and expressing number of overall genes, oncogenes, or anti-oncogenes in all aspects, we generated three radar plots summarizing the mean values of the scores of the 16 classifications of cells in terms of each cellular program (Fig. S6). We also verified our findings in published datasets and obtained the same tendencies of cellular programs in a cancer cell (Figs. S7 and S8) [23,24,30]. The results showed that the cellular behaviors were dependent strongly on the expressing numbers and levels of genes, especially oncogenes and anti-oncogenes, in a hepatocellular cancer cell.

Trajectory analysis was also performed to determine the relationship between neighboring similarities among all cancer cells with different classification labels of gene expression networks (Fig. S9). The monocle trajectories manifested explicit distribution tendencies of the cancer cells labeled by different classifications of gene expression networks, especially the classifications associated with the expression patterns of total genes, oncogenes, and anti-oncogenes. The cellular programs like proliferation and differentiation also manifested clear distribution tendencies in cancer cells. More importantly, the distribution tendencies of trajectories showed perfect coincidence with the tendencies in correlation heatmap. The results of trajectory analysis further confirm our observations.

3. Discussion

In this study, to explore the potentials of cellular programs in cancer cells with gene expression patterns, we have performed analysis to count the expression levels and numbers of genes, especially oncogenes and anti-oncogenes, in individual cancer cells of

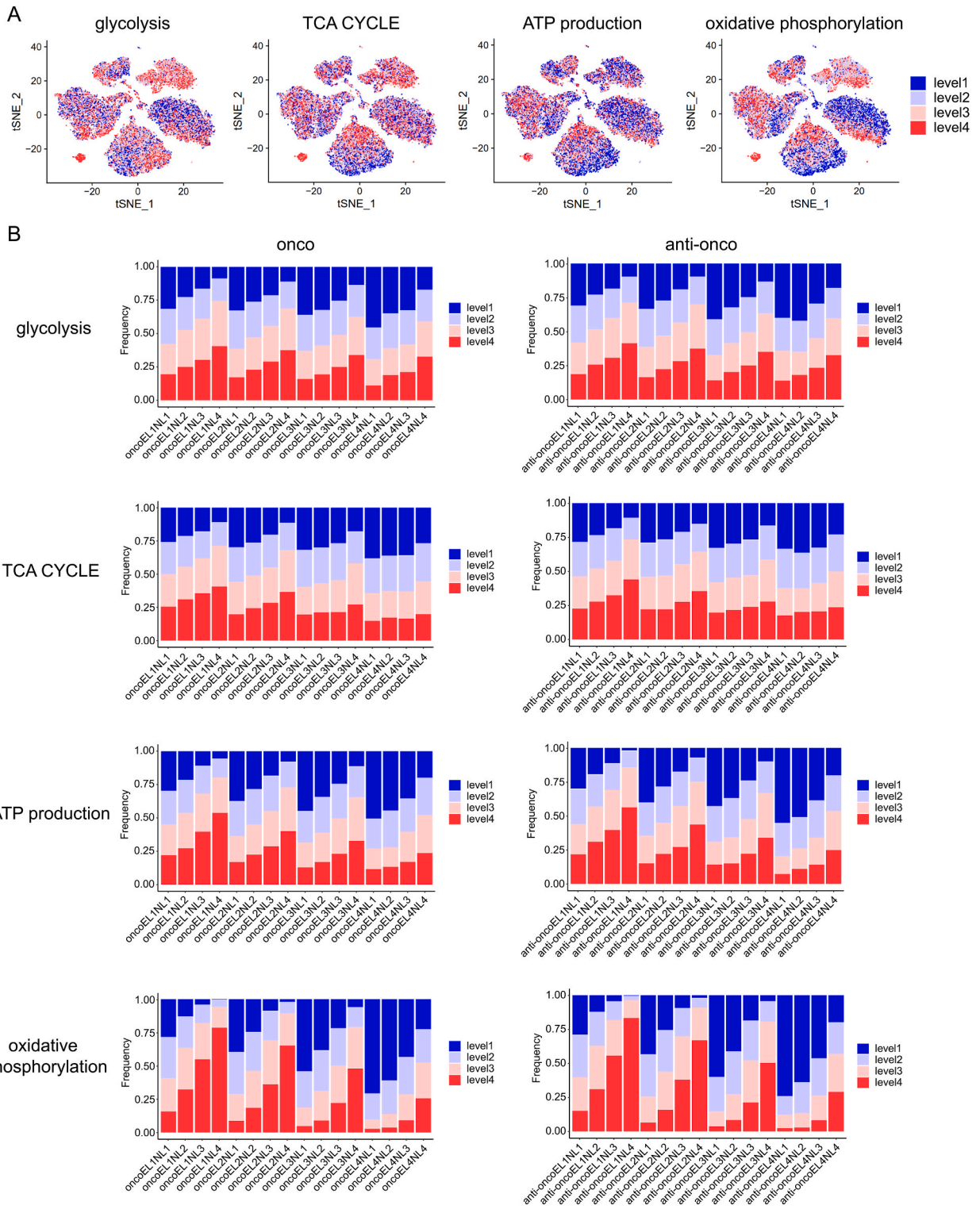
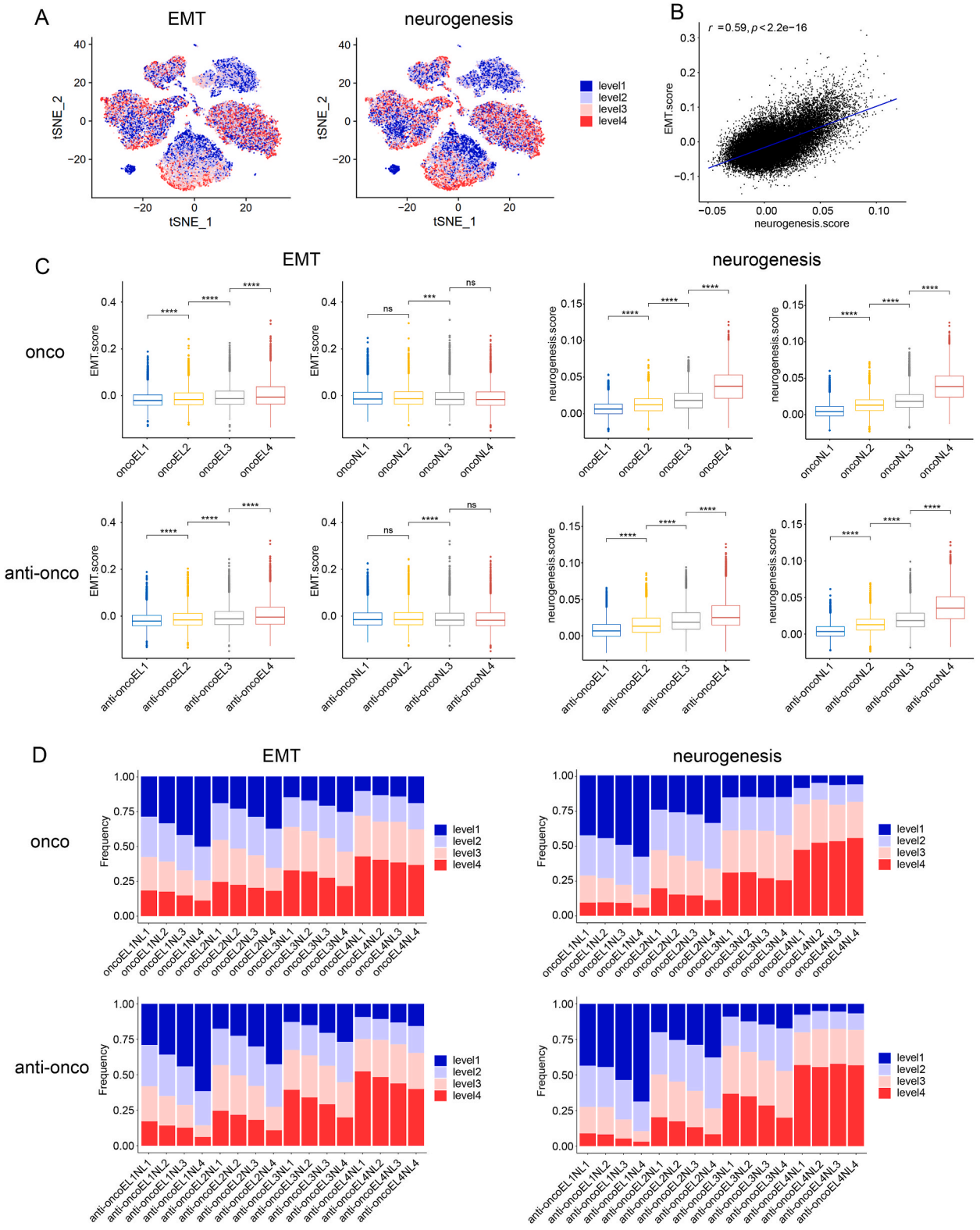


Fig. 6. The potentialities of energy metabolism in cancer cells of HCC. A. The tSNE projections show 4 levels of glycolysis, TCA CYCLE, ATP production and oxidative phosphorylation, respectively. **B.** The stacked box plots show the percentage of 4 levels of glycolysis, TCA CYCLE, ATP production and oxidative phosphorylation among cells with 16 classifications of oncogenes or anti-oncogenes, respectively.



(caption on next page)

Fig. 7. The potentialities of epithelial-mesenchymal transition and neurogenesis in cancer cells of HCC. A. The tSNE projections show 4 levels of epithelial-mesenchymal transition (EMT) and neurogenesis respectively. **B.** The scatter plot shows the correlation between EMT. score and neurogenesis. score with Pearson correlation coefficient (*r*-value) and *p*-value labeled at the top. **C.** The box plots show the differences in EMT. score or neurogenesis. score among cells defined with different NLs or ELs of oncogenes or anti-oncogenes. Statistical differences are indicated by asterisks. *p*-values were calculated by Wilcoxon test. ns: not significant. $*p < 0.05$, $**p < 0.01$, $***p < 0.001$, $****p < 0.0001$. **D.** The stacked bar charts show the percentage of 4 levels of EMT or neurogenesis among cells with 16 classifications of oncogenes or anti-oncogenes, respectively.

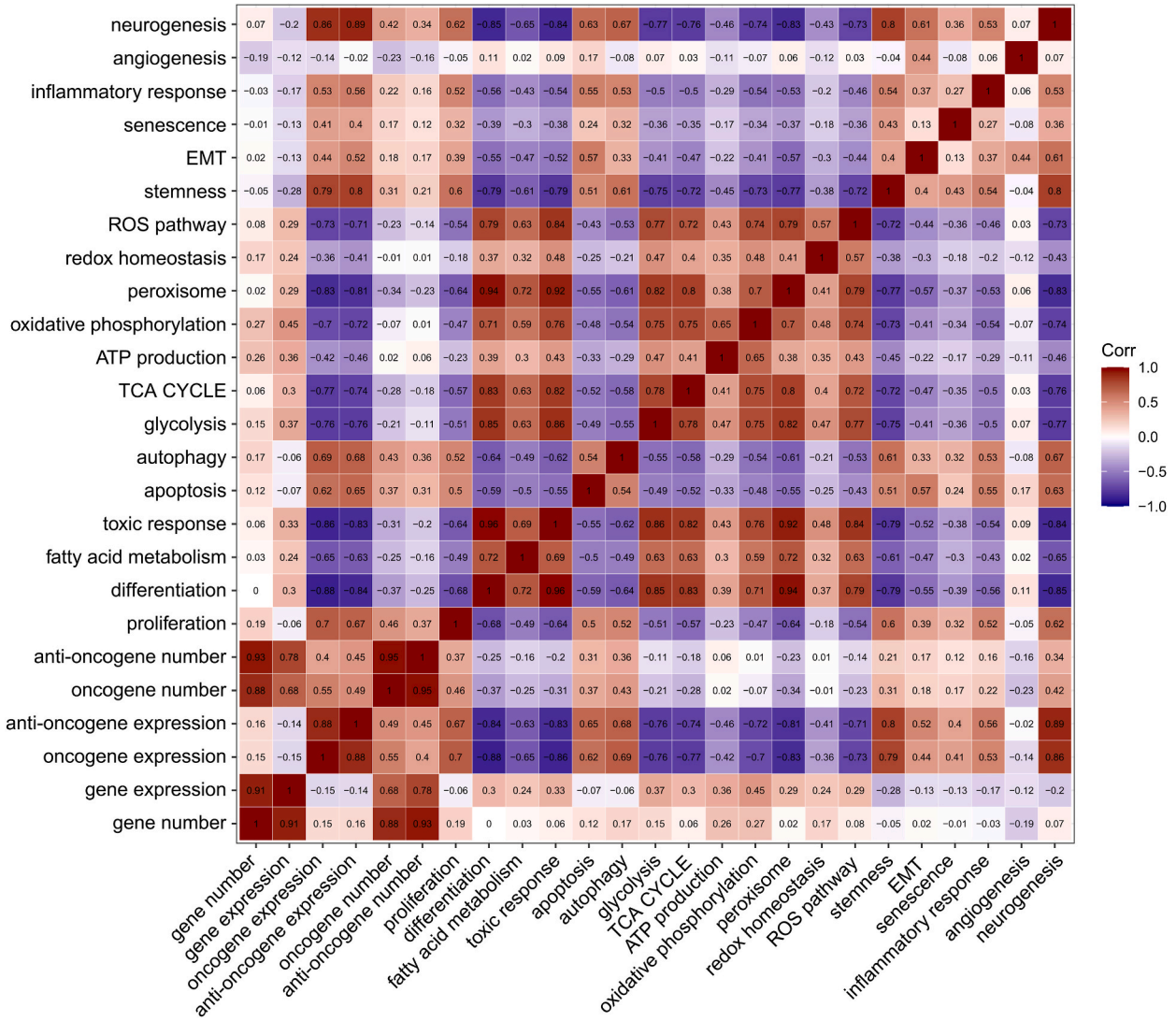


Fig. 8. The correlations among all cellular programs in cancer cells of HCC. The heatmap shows the correlations among all the cellular programs detected in cancer cells of sample P5-2. All Pearson correlation coefficients (*r*-values) are shown.

HCC and have divided all cells into 4 expression levels (ELs) and 4 number levels (NLs) of genes to form a classification pattern of cancer cells. The classifications of cancer cells display nicely the cellular potentials underlying cellular programs and cellular states in a cancer cell. The ease and efficiency of the data processing methods are possible to be widely used in single-cell RNA sequencing data and are very powerful in areas evaluating the cellular potentials of cell identities and states. The conclusions obtained in this article are relatively limited because we are unable to visualize the protein networks in a cancer cell to determine its actual cellular identity and state. However, the results conveyed in this article provide basic data of potential activations in a cell. The data can be used to develop a deeper understanding of how cancer progresses and undergoes evolutionary processes.

Accumulate data have demonstrated that genetic and epigenetic mutations and alterations that produce oncogenes with dominant gain of functions and anti-oncogenes with recessive loss of functions drive cancer initiation and progression. The underlying mechanism is that the expressing alterations of oncogenes and anti-oncogenes caused by genetic and epigenetic mutations and alterations

transform the inherent characteristics of a cell [31]. From an evolutionary view, DNA mutations or gene expressions that bring growth advantage to cells should be retained during population proliferation, while some others are expressed as genetic drift. With this traditional thinking, we examined the expression of some classical driver genes of HCC or genes highly associated with HCC. To be astonished, the expression levels and positive ratios of these genes were generally not high in cancer cells of HCC. The results urged us to look for the underlying genes to drive the formation and development of cancer cells. To this end, we analyzed all oncogenes and anti-oncogenes sequenced in individual cancer cells. The results showed strong inter-sample heterogeneity and intra-sample heterogeneity. We measure the gene expression in each cancer cell of HCC and identify that the expression of oncogenes and anti-oncogenes is always linked to the expression of overall genes in an individual cancer cell. The heterogeneous expression of oncogenes or anti-oncogenes in cancer cells represents the heterogeneities of overall activations of gene expression in cancer cell population in cancer tissues. Gene transcription occurs on naked DNA despite the complexity of eukaryotic transcription machinery in a cell [32]. The basic underlying mechanisms to achieve distinct functions of chromatin are changes in the accessibility of DNA to the transcription machinery leading to the activation or repression of target genes. In hepatocellular cancer tissues, transcriptional activation is extremely different among cancer cells. The transcription phenotypes reflect the chromatin structures in the cancer cells. It is reasoned that the large numbers of genes transcribed mean that a cancer cell carries more euchromatin structures for overall transcriptions within chromatins. To agree with the point, we identify if a cancer cell expresses more total genes, it expresses more oncogenes and anti-oncogenes. However, we find that the expressing levels of oncogenes and anti-oncogenes are not associated with the expressing levels of total genes in a cancer cell. The fact indicates that expression readouts of oncogenes and anti-oncogenes are more tightly controlled in a cancer cell.

Gene expression patterns or gene expression networks (GENs) provide system level explanations of developmental and physiological functions in the terms of the genomic gene expressions [33,34]. During embryonic development, GENs generate and guide the system-wide spatial development of specific cellular functions [35]. GENs also determine the main events of postembryonic development, including organogenesis and formation of adult parts and cell types. Beyond that, GENs control a vast array of physiological capabilities and modes of response to environmental fluctuations and challenges. Thus, gene expression plays a key role in a wide variety of cores of biological processes, ranging from organismal development, tissue homeostasis, and cell differentiation to cellular stress responses. Basic patterns of gene expression can be used to define these core biological processes and cellular behaviors in a cell. A specific cellular state, which is defined by the cell's phenotypes and functional characteristics, ultimately represents the readout of specific gene expression patterns. We identify that the overall expressing numbers of genes including oncogenes and anti-oncogenes highly associate with many major cellular processes, including proliferation, differentiation, apoptosis, autophagy, energy metabolism, EMT, and neurogenesis in individual cancer cells. The facts indicate that the overall transcriptional activations also transform the activation of genes involved in these cellular processes. However, we find that the overall expressing levels of genes are weakly linked to these major cellular processes. The facts suggest that the expression levels of genes that control a specific cellular process are tightly controlled in a cancer cell.

Phenotypically, the modified characteristics convert the cellular behaviors that drive a cell to become a cancer cell for cancer initiation or a cancer cell to gain more forces for cancer progression, expansion, invasion, and distant metastasis. The converted cellular behaviors of a cell include increasing abilities of proliferation, resisting cell death, altered autophagy, reassessing senescence, activating invasion and metastasis, conducting EMT, gaining self-renewal and stemness, promoting inflammation, reprogramming energy metabolism, and inducing angiogenesis and neurogenesis [36,37]. In this study, we apply the single-cell sequencing datasets of hepatocellular cancer cells to define mRNA expressing networks. Using the mRNA patterns, we have visualized the inherent biological capacities of a hepatocellular cancer cell and provide hallmarks to grade the potential behaviors of a cancer cell in cancer cell population of human hepatocellular carcinoma. Our results show that gene expression networks to conduct several phenotypic behaviors are coexisted in an individual cancer cell in human hepatocellular carcinoma. When counting the cells with different gene expression labels, we identify that cancer cells with the gene expression of oncogenes and anti-oncogenes at EL1NL1 and EL4NL4 consistent of the highest percentage of cell populations. It is reasonable to speculate that, there always exists a fraction of cells with the lowest expressing level and the least expressing number of oncogenes and anti-oncogenes, and another fraction of cells with the highest expressing level and the most expressing number of oncogenes and anti-oncogenes. Based on the gene expressing networks, EL1NL1 may exhibit higher hepatocytic signatures, energy metabolism characteristics, and redox potential, which makes it closer to non-malignant cells, while the EL4NL4 labeled cell population may have higher potential for almost all cellular programs. Although the number of rest classifications based on the expression of oncogenes and anti-oncogenes is relatively small, they also account for a significant proportion of cancer cells. We speculate that the identities and states of cells in those classified populations are more phenotypically variable compared to cells labeled with EL1NL1 or EL4NL4. It seems that those cells are more likely in transition states to undergo a specific cellular program.

In our data, as expected, the hepatocellular differentiation programs are oppositely associated with most major cellular processes including proliferation, apoptosis, autophagy, stemness, EMT, and neurogenesis in a cancer cell, further confirming the credibility of our data. Based on the gene expressing networks, many cellular processes are identified in a cancer cell. During the last few decades, evidence shows that altering the stoichiometry of lineage specifying transcription factors coexisting in a cell can induce cell fate changes in stem cells [38–40]. Now we show that gene expressing patterns control many cellular processes co-existing in a cancer cell. The discoveries have profoundly influenced mechanisms of how cancer cells undergo their processes. It might give rise to a conception that a cancer cell expresses balanced combinations of synergistic, agonistic, and antagonistic genes and undergoes a given biological process by a quick transmutation in gene expression networks resulting from the simultaneous activation of a new gene expression program and the extinction of the old one. Rapid transmutations in the expressions of these genes could be produced by random instabilities in a cancer cell or by surrounding cells including cancer cells, endothelial cells, neural cells, fibroblasts, inflammatory

cells, and other elements in the external microenvironment of a cancer cell. The finally outcoming gene expressing networks control a cancer cell to perform a specific cellular process in cancer tissues. Further experiments using single-cell mass cytometry and absolute quantification to capture the temporal dynamics of protein networks in an individual cancer cell might reveal which biological progress a cancer cell undergoes and provide evidence to solve the mechanisms of how a cell becomes a cancer cell or how a cancer cell is progressing in cancer tissues.

4. Materials and methods

4.1. Human sample collection

We enrolled 5 male patients who were pathologically diagnosed with hepatocellular carcinoma (HCC) and underwent curative resection at West China Hospital of Sichuan University. The study was approved by the Ethics Committee on Biomedical Research of West China Hospital, and all patients provided written informed consent. Their ages ranged from 55 to 69. For patient No.5 (P5), his two surgical samples were collected in less than a year (primary sample: P5-1; recurrent sample: P5-2). All patients were HBV-positive based on the HBsAg test. All 6 samples were diagnosed as moderate differentiation of hepatocellular carcinoma.

4.2. Single-cell isolation and sorting

All cancer tissues were collected and stored in an ice-cold RPMI 1640 medium and then immediately transferred for dissociation. On arrival, samples were placed in a sterile Petri dish on ice. With necrotic areas being removed, the rest tumor samples were cut into approximately 1 mm^3 pieces and added the premixed cell dissociation solution including 1 mg/mL collagenase I, 1 mg/mL collagenase II, 1 mg/mL collagenase IV, 0.25 U/mL dispase, 1 U/mL DNase I dissolved in 1640 medium. The digestive samples were incubated in a 37 °C water bath for 30–40 min to obtain a single cell suspension. The digestion reaction was terminated when most of the cells were single cells detected under the microscope. Finally, the single-cell suspensions were centrifuged at 500 *g* and 4 °C for 5 min.

After washing with the 1640 medium, the cell pellet was resuspended for FACS sorting. FITC anti-human CD45 antibody (555482, BD Biosciences) was used to exclude $CD45^+$ immune cells. After being stained for 20 min at room temperature, DAPI was added to exclude dead cells. The stained cells were washed with 1640 medium once and then sorted using BD Melody. 2×10^5 cells were used for further library construction. Once sorting is complete, the cells should be kept on ice.

4.3. Single-cell library construction and sequencing

According to the Singleron GEXCOPE protocols, the single cell suspension with a concentration of 2×10^5 cells/mL was loaded into a microfluidic chip. Cells randomly fell into the microwell. Subsequently, the prepared magnetic beads were loaded into the microfluidic chip in the same way. Once the lysis buffer was added to the chip, the released RNA from each cell was connected to a poly(T) sequence on beads which fell into the same microwell. To be noted, every bead has a unique cell barcode, and each poly(T) has its unique molecular identifier (UMI). After collecting the magnetic beads, GEXSCOPE Single-Cell RNA Library Kit was applied for reverse transcription. Using PCR to amplify the cDNA, products were purified and fragmented. After being measured by Qubit and Agilent Fragment Analyzer, the library was sequenced on an Illumina NovaSeq platform with 150 bp paired end reads.

4.4. Single-cell RNA-seq data processing and quality control

Raw reads from scRNA-seq were processed to generate gene expression matrixes using an internal pipeline. Briefly, raw reads were first processed with fastQC v0.11.4 and fastp to remove low-quality reads, and then were processed with cutadapt to trim the poly-A tail and adapter sequences. Cell barcode and UMI were extracted. After that, STAR v2.5.3a [41] was used to map reads to the reference genome GRCh38 (ensembl version 92 annotation). UMI counts and gene counts of each cell were acquired with featureCounts v1.6.2 software and used to generate expression matrix files for subsequent analysis.

All the downstream analyses were performed with the gene expression matrix using R v4.0.2. Before analyses, cells with <200 genes or $>20\%$ mitochondrial genes were filtered out.

4.5. Dimension reduction, clustering, and annotation

After filtering, the Seurat R package v4.1.1 was used for dimension reduction and clustering. The NormalizeData and ScaleData functions of Seurat were applied to normalize and scale all gene expressions. Then 3000 highly variable genes (HVGs) of each cell were generated with the FindVariableFeatures function for PCA analysis. After PCA, graph-based clustering at a resolution of 0.6 was performed with the top 30 principal components (PCs) and visualized using t-distributed Stochastic Neighbor Embedding (tSNE). Differentially expressed genes (DEGs) between different clusters were identified with the FindMarkers function of Seurat (Wilcoxon rank-sum test, avg_$\log_2$FC > 0.5). The classic marker genes were used to annotate the cell types of each cluster. The expressions of classic marker genes among all cell types were visualized using the VlnPlot function of Seurat.

4.6. Single-cell copy number variation (CNV) analysis

The copy number variation for each cell in HCC cancer cells was estimated with the inferCNV package v1.2.1 [42]. We took a set of immune cells from one sample as a reference and predicted the copy number changes of the cancer cells in the same sample based on the gene expression level. A sliding window including 100 genes was used to average the expression of these genes. By matching the genes with the human chromosomes, we obtained the overall inferred copy number variations of each cell. Every horizontal line in the heatmap represents one cancer cell. The red regions in the heatmap represent higher copy numbers compared to the reference and the blue regions represent lower copy numbers.

4.7. Identification of heterogeneities among cancer cells

To identify the heterogeneous expression patterns and signaling pathways in cancer cells derived from different samples, all the cancer cells were extracted to form a new object. After DEGs were identified among cancer cells from different samples ($avg\_log_2FC > 0.5$), the DotPlot function of Seurat was applied to exhibit the top 3 marker genes of each sample's cancer cells. Besides, all the DEGs belonging to each sample's cancer cells were shown in the heatmap by the DoHeatmap function. The UCell R package v2.1.0 and the irGSEA R package v1.1.2 were applied to further analyze the heterogeneities of cellular characters and signaling pathways using the whole MSigDB gene sets. The irGSEA.score function was used to calculate the enrichment score of all gene sets. The irGSEA.heatmap function was performed to visualize the gene sets with significant differences among cancer cells derived from different samples.

4.8. Classification defining of cancer cells based on expressed oncogenes, anti-oncogenes, and overall genes

First, the oncogenes and anti-oncogenes were collected from 3 databases: OncoKB [43], ONGene [44], and TSGene [45]. (https://www.oncokb.org/cancerGenes, http://www.ongene.bioinfo-minzhao.org/, https://bioinfo.uth.edu/TSGene/). Subsequently, the genes sequenced within each sample (nFeatures of all cells from this sample) were intersected with established oncogene sets and the anti-oncogene sets, respectively. The obtained oncogenes and anti-oncogenes were used for subsequent scoring. The oncogene. score and anti-oncogene. score of each cancer cell were calculated using the AddModuleScore function of the Seurat R package v4.1.1 to present the expression level of oncogenes or anti-oncogenes in a cell. Based on the quartiles of oncogene. scores or anti-oncogene. scores of all cells within one sample, four classifications of expression levels (EL) were defined. Then, the number of oncogenes or anti-oncogenes that had an expression higher than 1 was counted in every cancer cell (gene expression smaller than this threshold were ignored). Based on the quartiles of oncogenes. number and anti-oncogenes. number, four classifications of number levels (NL) were defined. The final 16 classification ids of oncogenes or anti-oncogenes were obtained by combining the EL and NL labels of each cell. As for the overall genes, the "nFeature_RNA" (represents the measured total number of genes) and "nCount_RNA" (represents the measured total number of mRNA molecules) information of each cell was used. Based on the quartiles of "nFeature_RNA" of all cells within one sample, we divided the cancer cells into 4 classifications with different NL labels. NL1 means cells with expressing gene numbers lower than Q1 (the lower quartile value of all numbers). NL2 means cells with the expressing gene numbers between Q1 and Q2 (the median value). NL3 means cells with the expressing gene numbers between Q2 and Q3 (the upper quartile value). NL4 means cells with expressing gene numbers higher than Q3. Based on the quartiles of "nCount_RNA" of all cells within one sample, 4 classifications of EL were defined. The definition way of ELs was the same as the definition way of NLs. The final classification ids of overall genes were obtained by combining EL and NL labels of each cell. After defining all the classifications of the cells within every sample, we merged the cells and conducted further analysis.

4.9. Evaluation of multiple cellular potentialities

To evaluate the potentialities associated with cell states and cell identities in every single cancer cell, the scores of many cellular programs and cellular behaviors were calculated using the AddModuleScore function of the Seurat R package. The explored potentialities include cell proliferation, differentiation, apoptosis, autophagy, energy metabolism, ROS pathway, redox homeostasis, peroxisome, stemness, senescence, EMT, inflammatory response, angiogenesis, and neurogenesis. The corresponding gene sets were downloaded from the MSigDB (https://www.gsea-msigdb.org/gsea/msigdb/), and all the involved gene sets were listed in the supplementary material. All analyzed genes were binned based on averaged expression, and the control genes were randomly selected from each bin [46]. The score of each cell was determined based on the average expression level of all genes in a given gene set. Then, based on the quartiles of the scores, 4 levels of each cellular program were defined. Cells with scores below the Q1 (the lower quartile value) were defined as level1. Cells with scores between Q1 and Q2 (the median value) were defined as level2. Cells with scores between Q2 and Q3 (the upper quartile value) were defined as level3. Cells with scores higher than Q3 were defined as level4. To exclude background interference between samples, we scored and defined 4 levels within each sample separately according to their respective quartiles. After defining all the cancer cells from each sample in terms of all aspects of cellular programs, we merged the cancer cells and performed further analysis.

4.10. Pearson correlation analysis

The Pearson correlation analysis was conducted to study the pairwise correlation of all variables. First, the metadata from every cancer cell including the nFeature_RNA, nCount_RNA, and all the scores calculated above was extracted to form a matrix. Then, with

the cor function of the ggcorrplot R package v0.1.4, a matrix of correlation coefficient (*r*-value, which measures the strength and direction of linear relationships between pairs of continuous variables) between every two cellular programs was generated. The matrix of *p*-values of correlation between every two cellular programs was generated with the cor_pmat function. The matrixes of *r*-values were used to generate the correlation heatmaps using the ggcorrplot function. All the matrixes of correlation coefficient (*r*-values) and *p*-values were shown in the supplementary material.

4.11. Trajectory analysis

Pseudotime analysis was performed using the Monocle R package v2.20.0. First, the newCellDataSet function was used to create a monocle object based on the established Seurat object. To be more detail, the RNA expression matrix and the metadata (which contains all scores and all classification labels of each cell) were transferred from the primary Seurat object to the monocle object. By calculating the average expression of all genes in all cancer cells within one sample and ranking them, we obtained the top 3000 genes in terms of average expression. The setOrderingFilter function was performed by setting the top 3000 genes as ordering_genes (a parameter of the setOrderingFilter function). Then, standard processes of monocle were performed. Trajectory plots of all cancer cells were generated with the plot_cell_trajectory function. By changing the color_by parameter, all the classification labels were projected on the trajectories respectively.

4.12. Immunofluorescence staining

Human hepatocellular carcinoma tissues were fixed in formalin for 1 h and then dehydrated with 30% sucrose overnight. When the dehydration was completed, the tissues should be embedded in O.C.T Compound and stored at −80 °C before slicing. Using a freezing microtome, the embedded tissues were cut into 8 μm slides for staining. The sections were first rinsed with 1 × PBST buffer 2 times for 5 min to remove the O.C.T around the tissues. After that, 0.5% Triton X-100 in PBS was used for permeabilization. The permeabilization was performed for 15 min at room temperature. Then the sections were blocked with 4% BSA in PBST for 1 h at room temperature. After that, sections were incubated with primary antibodies overnight at 4 °C: anti-KLF6 (207452-T10, Sino Biological, dilute at 1/50), anti-ELF3 (A6371, Abclonal, dilute at 1/100), anti-JUN (24909-1-AP, Proteintech, dilute at 1/100). On the next day, the sections were washed with 1 × PBST buffer 3 times for 5 min and then incubated with corresponding fluorescently conjugated secondary antibodies for 1 h at room temperature. After rinsing again, DAPI was added to the sections to dye the nuclear. Images were acquired using a ZEISS Scope A1 microscope.

4.13. Statistical analysis

All statistical analyses were implemented in R software. For the comparisons between two subpopulations of cancer cells, the ggpubr package v0.4.0 was used (Wilcoxon test). ns: not significant, $*p < 0.05$, $**p < 0.01$, $***p < 0.001$, $****p < 0.0001$.

Significance

Gene expressing networks of multiple biological processes coexist and are strongly related to the expressing number and levels of genes, especially oncogenes and anti-oncogenes, in an individual hepatocellular cancer cell.

Author contribution statement

Jin Zhao: Performed the experiments; Analyzed and interpreted the data; Wrote the paper.
Ran Lu, Chen Jin: Contributed reagents, materials, analysis tools or data.
Siying Li: Performed the experiments; Analyzed and interpreted the data.
Yulin Chen, Qiaorong Huang, Xue Li, Wentong Meng: Performed the experiments.
Hong Wu, Tianfu Wen: Conceived and designed the experiments.
Xianming Mo: Conceived and designed the experiments; Wrote the paper.

Funding statement

Xianming Mo was supported by The National Natural Science Foundation of China {81972592}.

Data availability statement

Data associated with this study has been deposited at the China National GeneBank DataBase under the accession number CNP0004283.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to

influence the work reported in this paper.

Acknowledgement

The Ethics Committee of West China Hospital, Sichuan University approved this study. We thank the participants and their families for their kind cooperation, generosity, and patience. The authors thank Yifan Mo, Kennedy Krieger Institute, for manuscript editing. This work was supported by the National Natural Science Foundation of China (81972592) and the 1.3.5 project for disciplines of excellence of West China Hospital (ZYGD20007, ZYJC18011).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heliyon.2023.e18305.

References

[1] C.M. Perez-Mesa, Cancer of the breast. Gross and histologic pathlogy, Major Probl. Clin. Surg. 5 (1979) 157–196.
[2] N. McGranahan, C. Swanton, Clonal heterogeneity and tumor evolution: Past, present, and the future, Cell 168 (2017) 613–628, https://doi.org/10.1016/j.cell.2017.01.018.
[3] M.A. Dawson, T. Kouzarides, Cancer epigenetics: from mechanism to therapy, Cell 150 (2012) 12–27, https://doi.org/10.1016/j.cell.2012.06.013.
[4] M. Greaves, C.C. Maley, Clonal evolution in cancer, Nature 481 (2012) 306–313, https://doi.org/10.1038/nature10762.
[5] M. Gerstung, C. Jolly, I. Leshchiner, S.C. Dentro, S. Gonzalez, D. Rosebrock, et al., The evolutionary history of 2,658 cancers, Nature 578 (2020) 122–128, https://doi.org/10.1038/s41586-019-1907-7.
[6] Y. Li, N.D. Roberts, J.A. Wala, O. Shapira, S.E. Schumacher, K. Kumar, et al., Patterns of somatic structural variation in human cancer genomes, Nature 578 (2020) 112–121, https://doi.org/10.1038/s41586-019-1913-9.
[7] S.C. Dentro, I. Leshchiner, K. Haase, M. Tarabichi, J. Wintersinger, A.G. Deshwar, et al., Characterizing genetic intra-tumor heterogeneity across 2,658 human cancer genomes, Cell 184 (2021) 2239–2254 e2239, https://doi.org/10.1016/j.cell.2021.03.009.
[8] R. Vegliante, I. Pastushenko, C. Blanpain, Deciphering functional tumor states at single-cell resolution, EMBO J. 41 (2022), e109221, https://doi.org/10.15252/embj.2021109221.
[9] G.L. Brien, D.G. Valerio, S.A. Armstrong, Exploiting the Epigenome to control cancer-promoting gene-expression programs, Cancer Cell 29 (2016) 464–476, https://doi.org/10.1016/j.ccell.2016.03.007.
[10] M. Cesana, M.H. Guo, D. Cacchiarelli, L. Wahlster, J. Barragan, S. Doulatov, et al., A clk3-hmga2 alternative splicing Axis impacts human hematopoietic stem cell molecular identity throughout development, Cell Stem Cell 22 (2018) 575–588 e577, https://doi.org/10.1016/j.stem.2018.03.012.
[11] B. De Rybel, B. Moller, S. Yoshida, I. Grabowicz, P. Barbier de Reuille, S. Boeren, et al., A Bhlh Complex controls embryonic vascular tissue establishment and indeterminate growth in arabidopsis, Dev. Cell 24 (2013) 426–437, https://doi.org/10.1016/j.devcel.2012.12.013.
[12] A.A. Connor, S. Gallinger, Pancreatic cancer evolution and heterogeneity: integrating Omics and clinical data, Nat. Rev. Cancer 22 (2022) 131–142, https://doi.org/10.1038/s41568-021-00418-1.
[13] J. Pan, Y. Hu, S. Sun, L. Chen, M. Schnaubelt, D. Clark, et al., Glycoproteomics-based signatures for tumor subtyping and clinical Outcome prediction of high-grade serous ovarian cancer, Nat. Commun. 11 (2020) 6139, https://doi.org/10.1038/s41467-020-19976-3.
[14] J.R. Conway, F. Dietlein, A. Taylor-Weiner, S. AlDubayan, N. Vokes, T. Keenan, et al., Integrated molecular drivers coordinate biological and clinical states in melanoma, Nat. Genet. 52 (2020) 1373–1383, https://doi.org/10.1038/s41588-020-00739-1.
[15] Z. Xu, K. Hu, P. Bailey, C. Springfeld, S. Roth, R. Kurilov, et al., Clinical impact of molecular subtyping of pancreatic cancer, Front. Cell Dev. Biol. 9 (2021), 743908, https://doi.org/10.3389/fcell.2021.743908.
[16] T. Baslan, J. Hicks, Unravelling Biology and shifting Paradigms in cancer with single-cell sequencing, Nat. Rev. Cancer 17 (2017) 557–569, https://doi.org/10.1038/nrc.2017.58.
[17] Q. Zhang, Y. He, N. Luo, S.J. Patel, Y. Han, R. Gao, et al., Landscape and dynamics of single immune cells in hepatocellular carcinoma, Cell 179 (2019) 829–845 e820, https://doi.org/10.1016/j.cell.2019.10.003.
[18] D. Barkley, R. Moncada, M. Pour, D.A. Liberman, I. Dryg, G. Werba, et al., Cancer cell states Recur across tumor types and form specific Interactions with the tumor microenvironment, Nat. Genet. 54 (2022) 1192–1201, https://doi.org/10.1038/s41588-022-01141-9.
[19] Y.C. Cohen, M. Zada, S.Y. Wang, C. Bornstein, E. David, A. Moshe, et al., Identification of resistance pathways and therapeutic targets in relapsed multiple myeloma patients through single-cell sequencing, Nat. Med. 27 (2021) 491–503, https://doi.org/10.1038/s41591-021-01232-w.
[20] M. Furuta, M. Ueno, A. Fujimoto, S. Hayami, S. Yasukawa, F. Kojima, et al., Whole genome sequencing Discriminates hepatocellular carcinoma with intrahepatic metastasis from multi-centric tumors, J. Hepatol. 66 (2017) 363–373, https://doi.org/10.1016/j.jhep.2016.09.021.
[21] R. Saeki, H. Nagai, S. Kaneko, M. Unoura, N. Yamanaka, E. Okamoto, et al., Intratumoral genomic heterogeneity in human hepatocellular carcinoma detected by restriction Landmark genomic scanning, J. Hepatol. 33 (2000) 99–105, https://doi.org/10.1016/s0168-8278(00)80165-8.
[22] R. Xue, L. Chen, C. Zhang, M. Fujita, R. Li, S.M. Yan, et al., Genomic and transcriptomic profiling of combined hepatocellular and intrahepatic cholangiocarcinoma reveals distinct molecular subtypes, Cancer Cell 35 (2019) 932–947 e938, https://doi.org/10.1016/j.ccell.2019.04.007.
[23] L. Ma, M.O. Hernandez, Y. Zhao, M. Mehta, B. Tran, M. Kelly, et al., Tumor cell biodiversity drives microenvironmental reprogramming in liver cancer, Cancer Cell 36 (2019) 418–430 e416, https://doi.org/10.1016/j.ccell.2019.08.007.
[24] Y. Sun, L. Wu, Y. Zhong, K. Zhou, Y. Hou, Z. Wang, et al., Single-cell Landscape of the Ecosystem in early-relapse hepatocellular carcinoma, Cell 184 (2021) 404–421 e416, https://doi.org/10.1016/j.cell.2020.11.041.
[25] N.N. Danial, S.J. Korsmeyer, Cell death: critical control points, Cell 116 (2004) 205–219, https://doi.org/10.1016/s0092-8674(04)00046-7.
[26] K. Kinouchi, P. Sassone-Corsi, Metabolic Rivalry: circadian homeostasis and tumorigenesis, Nat. Rev. Cancer 20 (2020) 645–661, https://doi.org/10.1038/s41568-020-0291-9.
[27] J.P. Thiery, H. Acloque, R.Y. Huang, M.A. Nieto, Epithelial-mesenchymal transitions in development and disease, Cell 139 (2009) 871–890, https://doi.org/10.1016/j.cell.2009.11.007.
[28] R. Lu, C. Fan, W. Shangguan, Y. Liu, Y. Li, Y. Shang, et al., Neurons generated from carcinoma stem cells support cancer progression, Signal Transduct. Targeted Ther. 2 (2017), 16036, https://doi.org/10.1038/sigtrans.2016.36.
[29] S.T. Schafer, F.H. Gage, Nerve cells from the brain invade prostate tumours, Nature 569 (2019) 637–638, https://doi.org/10.1038/d41586-019-01461-7.
[30] L. Ma, S. Heinrich, L. Wang, F.L. Keggenhoff, S. Khatib, M. Forgues, et al., Multiregional single-cell dissection of tumor and immune cells reveals stable lock-and-key features in liver cancer, Nat. Commun. 13 (2022) 7533, https://doi.org/10.1038/s41467-022-35291-5.
[31] D. Hanahan, Hallmarks of cancer: new dimensions, Cancer Discov. 12 (2022) 31–46, https://doi.org/10.1158/2159-8290.CD-21-1059.

[32] N.A. Woychik, M. Hampsey, The Rna polymerase Ii machinery: structure Illuminates function, Cell 108 (2002) 453–463, https://doi.org/10.1016/s0092-8674(02)00646-3.
[33] N. Festuccia, R. Osorno, V. Wilson, I. Chambers, The role of pluripotency gene regulatory network components in mediating transitions between pluripotent cell states, Curr. Opin. Genet. Dev. 23 (2013) 504–511, https://doi.org/10.1016/j.gde.2013.06.003.
[34] C.S. Vink, F.J. Calero-Nieto, X. Wang, A. Maglitto, S.A. Mariani, W. Jawaid, et al., Iterative single-cell analyses define the transcriptome of the first functional hematopoietic stem cells, Cell Rep. 31 (2020), 107627, https://doi.org/10.1016/j.celrep.2020.107627.
[35] A. Kirjavainen, P. Singh, L. Lahti, P. Seja, Z. Lelkes, A. Makkonen, et al., Gata2, Nkx2-2 and Skor2 Form a Transcription Factor Network Regulating Development of a Midbrain Gabaergic Neuron Subtype with Characteristics of Rem-Sleep Regulatory Neurons, Development, 2022, p. 149, https://doi.org/10.1242/dev.200937.
[36] D. Hanahan, R.A. Weinberg, Hallmarks of cancer: the next generation, Cell 144 (2011) 646–674, https://doi.org/10.1016/j.cell.2011.02.013.
[37] P. Kreuzaler, C.J. Watson, Killing a cancer: what are the Alternatives? Nat. Rev. Cancer 12 (2012) 411–424, https://doi.org/10.1038/nrc3264.
[38] T. Graf, Transcription factor stoichiometry drives cell fate: single-cell Proteomics to the rescue, Cell Stem Cell 24 (2019) 673–674, https://doi.org/10.1016/j.stem.2019.03.002.
[39] C.G. Palii, Q. Cheng, M.A. Gillespie, P. Shannon, M. Mazurczyk, G. Napolitani, et al., Single-cell Proteomics reveal that quantitative changes in Co-expressed lineage-specific transcription factors determine cell fate, Cell Stem Cell 24 (2019) 812–820 e815, https://doi.org/10.1016/j.stem.2019.02.006.
[40] M. Ye, H. Iwasaki, C.V. Laiosa, M. Stadtfeld, H. Xie, S. Heck, et al., Hematopoietic stem cells expressing the myeloid Lysozyme gene Retain long-term, multilineage repopulation potential, Immunity 19 (2003) 689–699, https://doi.org/10.1016/s1074-7613(03)00299-1.
[41] A. Dobin, C.A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, et al., Star: Ultrafast universal Rna-seq aligner, Bioinformatics 29 (2013) 15–21, https://doi.org/10.1093/bioinformatics/bts635.
[42] A.P. Patel, I. Tirosh, J.J. Trombetta, A.K. Shalek, S.M. Gillespie, H. Wakimoto, et al., Single-cell rna-seq Highlights intratumoral heterogeneity in primary glioblastoma, Science 344 (2014) 1396–1401, https://doi.org/10.1126/science.1254257.
[43] D. Chakravarty, J. Gao, S.M. Phillips, R. Kundra, H. Zhang, J. Wang, et al., Oncokb: a precision Oncology Knowledge Base, 2017, JCO Prec. Oncol. (2017), https://doi.org/10.1200/PO.17.00011.
[44] Y. Liu, J. Sun, M. Zhao, Ongene: a literature-based database for human oncogenes, J. Gen. Genom. 44 (2017) 119–121, https://doi.org/10.1016/j.jgg.2016.12.004.
[45] M. Zhao, P. Kim, R. Mitra, J. Zhao, Z. Zhao, Tsgene 2.0: an updated literature-based Knowledgebase for tumor suppressor genes, Nucleic Acids Res. 44 (2016) D1023–D1031, https://doi.org/10.1093/nar/gkv1268.
[46] I. Tirosh, B. Izar, S.M. Prakadan, M.H. Wadsworth 2nd, D. Treacy, J.J. Trombetta, et al., Dissecting the multicellular Ecosystem of metastatic melanoma by single-cell rna-seq, Science 352 (2016) 189–196, https://doi.org/10.1126/science.aad0501.