Quantify the spatial distribution of immune and cancer cells within the tumour tissue in patients with primary hepatocellular carcinoma

Kangpei Tao, Nadia Guerra[1], Khalid A. Jabbar[2], Yinyin Yuan[2]

1.Sir Alexander Fleming Building, Imperial College London. 2.Institute of Cancer Research, Sutton

Abstract

This study tested the impact of the spatial distribution of immune and cancer cells on the clinical outcome of patients with primary hepatocellular carcinoma (HCC), using data from the public available TCGA dataset. To quantify such spatial distribution, we adopted an automated hematoxylin and eosin (H&E) stained image analysis program developed by our Lab. The location of lymphocytes and cancer cells in patient's tumour tissue sections were extracted. Following by a spatial statistic method called Getis-Ord hotspot analysis, we proposed three spatial scores that quantified the immune and cancer cells' spatial distribution. They were termed immune hotspot fractional, cancer hotspot fractional and cancer-immune hotspot co-localization and were associated with different biological meaning. In total, 377 patients were randomly divided into discovery (n = 188) and validation (n = 189) cohort. The prognostic value of the three spatial scores proposed, along with the lymphocyte abundance (cell-cell ratio) and infiltration (cell-space ratio) was tested using patient's overall survival data. The dependence of the prognostic factor was examined using the multivariate survival test with patient's TNM stage, α-fetoprotein level at diagnosis and history of primary risk factor being the adjusting factors. In conclusion, we found that only the immune hotspot fractional score was of strong prognostic value in the univariate survival test (discovery p = 0.027, validation p = 0.029). However, the independence of its prognostic value was significant only in the discovery cohort (p = 0.002). Patients with high immune hotspot fractional score were associated with better overall survival (discovery HR = 1.9 (1.1-3.4), validation HR = 1.8 (1.1-3)). It was the first time the spatial distribution of immune and cancer cells showing its prognostic value in HCC. It expended the boundary of cancer types where such distribution pattern was able to predict patient's clinical outcome and certainly revealed the importance to carry out spatial analysis in more cancer types. More studies needed to be done to optimize the approach in quantifying the spatial distribution and to reveal the internal driver of the distribution pattern.

Introduction

Previous study using diethylnitrosamine (DEN)-induced HCC mice model has revealed that the activation of immune system in defending tumour invasion could have an opposite effect that boosted its progression (*1*). While studies carried out in other cancer types revealed a positive implication upon such activation but in regarding to a consideration on the spatial distribution of immune cells within the tumour (*2-5*). The spatial distribution has been the interest of researchers for a long time. They were mostly focused on the immune cell's spatial distribution as it has been realized that the clinical outcome of immunotherapy is highly relied on the patient's tumour immune profile (*6*). Such profile, while still under investigation, were thought to include the patient's data in the location, density and type of immune cells within the tumour and its surrounding area (*7*). Apart from the ability in predicting the outcome of immunotherapy, it has also been established that the immune cells played an important role in tumour rejections in the early stage of cancer development which had a positive impact on the patient's performance and ultimately, the survival (*8*). The first successful attempt in correlating the spatial distribution with clinical outcome was made in the colorectal cancer where the assessment on the regional CD8+ T cell infiltration was proved to be a stronger relapse predictor than the standard TNM staging system (*7*). Since then, the similar correlation has been established in other cancer types (*9*). However, our lab has demonstrated the need to consider the spatial distribution of cancer cell as well using data from patients with estrogen receptor negative (ER-) breast cancer. In that study, we showed that the proportion of cancer-immune co-localized hotspot can predict patient's overall survival and can further stratify the patients with immune-cell rich tumour, where the proportion of immune hotspot itself was of no prediction value (*10*). We also demonstrated, in a lung cancer dataset (not published), the importance of combining the spatial distribution of multiple type of cells using multi-staining images.

HCC is the sixth most common cancer type in the world with a variety of risk factors (*11*). These risk factors displayed an unevenly distribution according to patient's regional background which made the identification of a globally applicable prognostic factor a great challenge (*12*). Many attempts have been made in the identification of a reliable prognostic biomarker, yet few had seen progress. In the quest of finding markers with prognostic value, Guerra and colleagues were among the first to report a pro-tumorigenic role of a potent immune stimulator NKG2D (Natural killer group 2 member D), which was known to associate with

tumour suppressor function, in a DEN-induced HCC mice model (*13*). NKG2D is a membrane receptor expressed on most type of lymphocytes (*14*). Killer cell lectin like receptor K1 gene (*KLRK1*) is responsible for the synthesis of NKG2D. The protein ligands that activates the NKG2D receptor in human include UL16 binding protein (ULBP) 1,2 and 3, MHC class I polypeptide-related sequence A (MICA) and MHC class I polypeptide-related sequence B (MICB) (*15*). They were either synthesized and released as soluble ligands or anchored at the cell surface in response to cellular stress (*16, 17*). Patients with primary HCC were mostly associated with a chronic inflammation background in their livers. Guerra and colleagues reported that the NKG2D-mediated innate immune response in the event of chronic inflammation was thought to propagate a positive feedback loop that would recruit more immune cells and boost tumorigenesis by releasing inflammatory cytokines and inducing the production of hepatocyte growth factor (*18*).

This drew our attention as our lab has demonstrated the importance of the spatial distribution of immune and cancer cells in breast and ovarian cancer (*10, 19, 20*). And the study based on the same discipline in HCC was still blank. Guerra's study provided evidence that the immune activity can affect the tumour progression in liver. Moreover, they found that in the NKG2D deficient mice (gene knocked out), the level of CD8+ and CD4+ T cell was reduced in the NTME (none tumour microenvironment) area compared with wide type mice while the level was the same in the TME (tumour microenvironment) area. The spatial heterogeneity displayed urged a need for a method to quantify the spatial distribution of a specific group or groups of cells. Thus, we decided to adopt the same approach used in quantifying the spatial distribution in the breast cancer to have a preliminary investigation in HCC using the public available TCGA dataset. As the method for spatial quantification was not subjected to specific cancer type. Patients tumour tissue section was used for analysis.

Restricted by the technology, previous spatial analysis carried out using patient's tissue specimen was heavily relied on the pathologist's estimation which is a laborious work and often subject to personal bias. Complete annotation on thousands of cells was also an impossible task. This hindered the researchers to extract valuable spatial information from tissue specimen. However, our lab has developed an automated cell classification and detection algorithm based on supervised machine learning, called CRImage (*21*). By utilizing the advantage of the computer-based approach, the limitation of using pathologist's estimation was prevented. The cell classification (cancer, lymphocytes, stromal and other cells) and location data can now be acquired in a high-throughput, accurate and reproducible manner. To quantify the spatial distribution, we applied a spatial statistical method that was originally designed for ecological purpose, called Getis-Ord hotspot analysis (*22*). This method examines if a given event, such as the occurrence of lymphocytes or cancer cells, was observed at a significant higher frequency in a given area. The observation in each given area was compared with the observations in its surrounding areas. Thus, whether the given area should be deemed as a 'hotspot' is independent of the global observation.

The advantage of using hotspot analysis in quantifying spatial distribution is that it can better reflect the cell clustering pattern and provide an insight in the tumour architecture and its microenvironment. Taking the estrogen receptor positive (ER+) breast cancer as an example. In the investigation of the ability of immune cell's spatial distribution to predict recurrence-free survival after endocrine therapy. Our lab has demonstrated that the density of immune cell measured using patient's tumour tissue section was not associated with such prediction value, while the spatial score generated using Getis-Ord hotspot analysis was (*20*). Using the density as the measurement of spatial distribution is a common approach used by researchers. Yet our study has revealed the limitation of such approach.

In conclusion, this study used Getis-Ord hotspot analysis to quantify the spatial distribution of lymphocytes and cancer cells in the tumour tissue sections of patients with primary HCC. The potential of the spatial score to serve as an independent prognostic factor revealed in this study solidified the importance of utilizing hotspot analysis in studying the spatial distribution. We also briefly looked into the ability of the NKG2D and its ligands' expression level in shaping the clustering pattern of lymphocytes and cancer cells within the tumour tissue area. The results indicated that the NKG2D and its ligands alone was not sufficient in explaining the pattern of spatial distribution. To fully reveal the role of immune activity in predicting patient's clinical outcome. The tissue section that covers both tumour and the surrounding area should be analyzed. And as the quantification of spatial distribution is a reflection of a combined biological events, it was conceivable that the ideal quantification should consider the spatial distribution of more cells and with a higher specificity. Thus, more work should be carried out in the future.

Methods

Sample data

This study used data from the public-available TCGA dataset. The project code is TCGA-LIHC. There were 377 patients participated in this project to date. Patient's ethnicity composition is American Indian (n=2), Asian (n=161), African American (n=17), white (n=187) and not reported (n=10). Patients were randomly divided into discovery cohort (n=188) and validation cohort (n=189).

Patient's tumour tissue sections which were fresh frozen, H&E stained, 40 times magnified and scanned under the microscope were used for analysis. Patients tissue section images (n=491) were downloaded from the TCGA data portal https://portal.gdc.cancer.gov/ using the GDC data transfer tool called gdc-client, which was separately downloaded from https://gdc.cancer.gov/access-data/gdc-

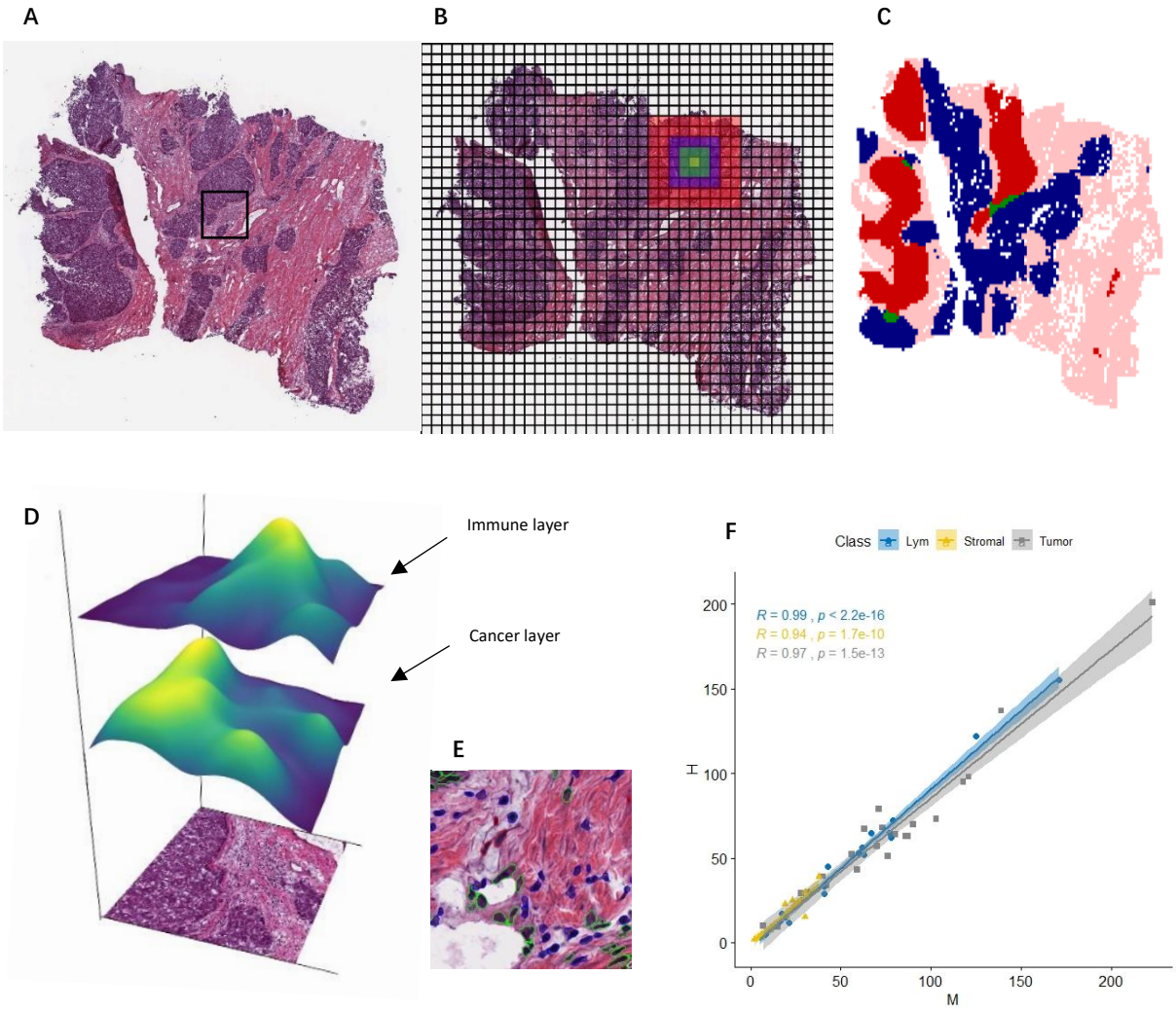

Figure 1 Illustration of the automated cell annotation and the Getis-Ord hotspot analysis. (A) An example of the HCC tumour tissue section image analysed. They were fresh frozen H&E stained tumour tissue section that is 40 times magnified under microscope. **(B)** An illustration of the rationale of Getis-Ord analysis. First, a square grid was applied to the tissue section image. The grid size can be varied where in this study we tested 50, 100 and 250 µm as the grid size. A binary tissue mask that defined the boundary of the tissue section was imported from the CRImage algorithm. So the tissue coverage of every grid can be calculated. Second, for grids with no tissue coverage or with tissue coverage less than 50%, they will be excluded from analysis. Finally, the remaining grids will be compared with their neighborhood region where there were three configurations available namely the first or the second and the fourth order. For the grid that is indicated in yellow, the first order area would be the green area. The second order area is the green plus purple area. The fourth order area is the red area plus the first and the second order area. **(C)** Result of the Getis-Ord analysis. The result was based on the spatial configuration of using 50 µm as the grid size and the fourth order as the neighborhood size. The blue area indicates the immune hotspot, the red area indicates the cancer hotspot, the green area indicates area that was co-localized by immune and cancer hotspot. **(D)** An illustration of the concept of 'hotspot'. For an H&E stained tissue section, once the location and classification data were available at a single cell level, we can use spatial statistical method to analysis the cell clustering pattern in a reproducible manner. The cancer and immune layer were the surface plot of the cell density calculated using Kernel density estimation method. This example tissue section was the square area indicated in fig. 1A. **(E)** An illustration of the result of the automated cell annotation algorithm. Using the machine learning based algorithm developed by our lab called CRImage, the cell classification and location data was generated in a high-throughput, accurate and reproducible manner. Tumour cells were lablled using green circle, lymphocytes were labelled using blue circle and stromal cells were labelled using red circle. **(F)** The detection accuracy validation test of the CRImage. The Pearson's correlation test was applied to test the concordance between the result of machine annotation (M) and human annotation (H). The correlation coefficient of tumour cells, lymphocytes and stromal cells were 0.97, 0.99, 0.94 respectively.

data-transfer-tool. Along with the image, the corresponded pathologist's report was also downloaded where the pathologist's estimation of the lymphocyte infiltration (cell-space ratio) was recorded (n=334). Patients clinical data, including the survival data was downloaded from UCSC Xena browser at https://xenabrowser.net/. The data was last updated at 2016-04-27. Gene transcription data was also downloaded from UCSC Xena browser. It was last updated at 2017-10-13. The gene expression data was measured using the Illumina HiSeq 2000 RNA Sequencing platform and followed by a log2(x+1) transformed RSEM normalization (*23*).

Because during the tissue section preparation, some patients have multiple slides in the same tissue at different locations (top, middle and bottom), and some patients have slides at normal tissue or metastatic tumour tissue. For consistency, this study only used tissue sections that were from primary tumour and top slide (n=367, discovery n=182, validation n=185).

CRImage: an automated cell classification and detection algorithm

CRImage is an algorithm based on supervised machine learning (*21*). This program was used to acquire cell classification and cell location data. It was validated on another two TCGA lung cancer datasets, namely TCGA-LUSC and TCGA-LUAD. The tissue preparation procedure among the three datasets are of the same. The target cell's morphology feature is also similar among the three datasets. Thus, we left the parameters in the CRImage unchanged in processing TCGA-LIHC tissue section images.

The detection accuracy is validated by comparing with the results from human annotation. In total, over 60 million cells were processed through the program. For validation, three patient's tissue sections were randomly selected. Seven 500 by 500 μm regions that contained a mixture of tumour cells, lymphocytes and stromal cells detections were selected from each of the three images. The 21 image tiles were manually annotated. In total, there were 1402 tumour cells, 937 lymphocytes and 336 stromal cells annotated. The results were compared with the machine counts generated by CRImage. A Pearson's correlation plot was used to test the concordance between the two.

Getis-Ord hotspot analysis

The cell location and classification data acquired from CRImage was served as inputs to the hotspot analysis. As indicated in *fig. 1B*, the images were first divided into small square regions. A binary tissue map which specifies the boundary of tissue in each image was also inputted into the hotspot analysis. This helps us to exclude regions that contained no tissue. Regions that had a tissue coverage less than a half were also excluded from analysis (*10*). There are two customizable parameters, the first is the grid size which defined the size of each region, the second is the neighborhood size which defined the area that each region was comparing to (*10*). There are numbers of considerations in deciding the two parameters, as we want to reflect the cells clustering pattern within each image in a way that best correlates to the clinical outcome of patients. A grid size that is too large would lose its ability in reflecting the spatial distribution pattern, where a grid size that is too small would significantly increase the computational workload and potentially introduce more noise detections in the output. The neighborhood size is denoted by the level of order, as indicated in *fig. 1B*, where the first order means the regions that is adjacent to the region and its four vertexes and the second order means the regions that is adjacent to the first order region and its four vertexes plus the first order region itself, so on so forth (*10*).

In consistent to the previous successful implementation of Getis-Ord hotspot analysis in the ER- breast cancer, here we adopted the same spatial configuration tested in that study (*10*). The grid size of 50, 100 and 250 μm and the neighborhood size of the first order, the second order and the fourth order were combined into different spatial configurations. When the grid size was set as 250 μm, only the neighborhood size of the first order is considered as the radius of the neighborhood should not be exceeding twice the cell-cell communication distance (250 μm) (*24*).

Based on the result of Getis-Ord hotspot analysis, three different spatial score was used to summarize the spatial distribution of immune and cancer cells, namely *fi*, *fc* and *fci*. They were proposed and validated in the previous study (*10*). These spatial scores were calculated in the different way based on the biological significance they represented. *fi* is the immune hotspot fractional, which represents the proportion of immune hotspots that are also cancer hotspots (*10*). *fc* is the cancer hotspot fractional, which represents the proportion of cancer hotspots that are also immune hotspots (*10*). *fci* is the cancer-immune hotspot colocalization, which represents the proportion of the cancer-immune co-localized hotspots in the whole area (*10*).

Survival test

Three parameters *fi*, *fc* and *fci* under 7 spatial configurations plus the lymphocyte abundance (cell-cell ratio) and the lymphocyte infiltration (cell-space ratio) were tested in the discovery cohort using univariate survival test to find the cut-off point that yield the most favorable p-values. The R function called 'surv_cutpoint' was used to find the best cut-off point (*25*). The same cut-off point found in the discovery cohort was tested in the validation cohort to examine the reproducibility of the findings. Cut-off points that fell out of the 15% - 85% range of its original data will be deemed as invalid.

Patient's history of primary risk factor, pathological TNM stage score and the level of α-fetoprotein at diagnosis were used as the existing major prognostic factors in the multivariate survival test (*12*). According to a previous review on the existing prognostic factors in HCC, patients with the level of α-fetoprotein higher than 25 μg/L were associated with poor survival (*12*). The value of these three prognostic factors were recorded in the clinical data, however the records were not complete in some patients (history n=353, TNM stage n=354, α-fetoprotein n=282). To test the prognostic value of these three factors, univariate survival test was done in both cohorts. They were used to examine whether the prognostic value of the spatial score, if any of them could prove to have such value in the univariate survival test, was independent or not. The univariate and multivariate survival test were based on the cox proportional hazard model. Survival curve was plotted using Kaplan-Meier method.

Other statistical methods

All statistical works were carried out in R (*26*). Data normality was examined using QQ plot. For comparing two unpaired groups of continues data that is not normally distributed, the Mann-Whitney U-test was used. For paired two groups of continues data that is not normally distributed, the Wilcoxon test was used. To test the correlation between two groups of continues data, the

Neighborhood size	*First order*			*Second order*			*Fourth order*		
Grid size (μm)	fi	fc	fci	fi	fc	fci	fi	fc	fci
50	x	x	x	0.0005	0.002	0.0007	0.027	0.026	0.027
100	x	x	0.025	0.0009	0.0019	0.0005	x	x	x
250	/	/	/	-	-	-	-	-	-
Lym abundance	0.02								
Lym infiltration	x								

x denoted the p-value is greater than 0.05, '/' denoted the configuration is not applicable in some images

Table 1 Select spatial scores with prognostic value in the discovery cohort. An R function 'surv_cutpoint' was used in the discovery cohort under each spatial score to find the cut-off point that yield the minimum p-value. The cut-off point should fall within the 15 – 85% range of its data, outliers will be deemed as invalid cut-off point and excluded. This table showed the minimum p-value found under each spatial score. x denoted the prognostic value was not significant for that spatial score in the discovery cohort. '/' denoted the spatial configuration was not applicable in some patient's tissue section images, spatial scores associated with these spatial configurations will be excluded. Where a p-value is significant means the corresponded spatial score was associated with prognostic value in the discovery cohort. The method is based on cox proportional hazard model. All three spatial scores under the spatial configuration of '50_4', '50_2' and '100_2' and the cancer-immune co-localization under '100_1' plus the lymphocyte abundance were found to be of prognostic value in the discovery cohort.

Configuration	*Discovery cohort min pval*	*Validation cohort pval*
50_4 fi	0.027	0.029
50_4 fc	0.026	x
50_4 fci	0.027	x
50_2 fi	0.0005	x
50_2 fc	0.002	x
50_2 fci	0.0007	x
100_1 fci	0.025	x
100_2 fi	0.0009	x
100_2 fc	0.0019	x
100_2 fci	0.0005	x
Lym abundance	0.02	x

x denotes the p-value is greater than 0.05

Table 2 Immune hotspot fractional score under the spatial configuration of 50 μm grid size and the fourth order of neighborhood size was selected as the most suitable spatial score. In consistent to the previous study, we first tested the same spatial configuration that was showed to be the most suitable spatial configuration in the previous study as shown in the top three rows. Method was based on the cox proportional hazard model. The cut-off point found in the discovery cohort was applied to the validation cohort to test the reproducibility of the prognostic value. The immune hotspot fractional was showed to associate with prognostic value in both discovery and validation cohort (dis p = 0.027, val p = 0.029). We also tested the remaining spatial configuration with prognostic value revealed in the discovery cohort as shown in table 1. None of them could reproduce their prognostic value in the validation cohort (p > 0.05).

Pearson's correlation test was used when the data was normally distributed, the Spearman's rank correlation test was used when the data was not normally distributed. To test the correlation between discrete and continues data, the Jonckheere-Terpstra (JT) trend test was used. Any observation is deemed as significant when the p-value is less than 0.05.

Results

Accurate automated cell classification enables spatial mapping of lymphocytes and cancer cells

The way that CRImage annotated different type of cells was illustrated in *fig. 1E*. Tumour cells, lymphocytes and stromal cells were labelled using green, blue and red color respectively. For objects that can't be determined, a white label will be applied. Along with the annotation on the tissue section, CRImage also returns the coordinates and the classification data of each cell it annotated.

The concordance between the human annotation and the machine output is shown in *fig. 1F*. They are all strongly correlated with p-value less than 0.05. The correlation coefficient is 0.97 for tumour cells, 0.99 for lymphocytes and 0.94 for stromal cells. The lymphocyte abundance can be

calculated using the output data of CRImage, the distribution of the lymphocyte abundance is shown in *fig. 2D*, the median is 0.160 the interquartile range is 0.114 – 0.198.

The result of the Getis-Ord hotspot analysis is illustrated in *fig. 1C.* The blue area indicated the immune hotspot, the red area indicated the cancer hotspot, the green area indicated area that the immune and cancer hotspot was co-localized.

Select the most suitable spatial score

The prognostic value of spatial scores under all available spatial configurations and the lymphocyte abundance and infiltration were tested in an attempt to find the cut-off point that yield the minimum p-value in the univariate survival analysis in discovery cohort. The result was shown in *table 1*. The spatial configuration using 50 μm and 100 μm as the grid size, the second order as the neighborhood size and the 50 μm with the fourth order were associated with prognostic value in the discovery cohort. What's more, the cancer-immune co-localization under the configuration of 100 μm grid size and the first order neighborhood size plus the lymphocyte abundance also revealed their prognostic value in the discovery cohort.

Next, the reproducibility of these spatial scores' prognostic value was tested in the validation cohort using the same cut-off point. The same spatial configuration used in the previous successful implementation was tested first. As listed in the top three rows in *table 2*, the immune hotspot fractional score under the spatial configuration using 50 μm as the grid size and the fourth order as the neighborhood size (50_4 *fi*) has proved to be of prognostic value in both discovery and validation cohort. The spatial score under other spatial configurations were also tested as listed in the rest part in *table 2*. However, none of these spatial scores could reproduce their prognostic value in the validation cohort, leaving the 50_4 *fi* being the only spatial score that showed prognostic value in both discovery and validation cohort. It is in consistent with the previous study in terms of the spatial configuration used (*10*). The cut-off point found for 50_4 *fi* is 0.0633. And this will be referred as the immune hotspot fractional score in the following text.

The distribution of the immune hotspot fractional score is shown in *fig. 2E*. The median is 0.032, the interquartile range is 0.003 – 0.110. The cut-off point is located between the median and the upper quartile of the data. The lymphocyte abundance profile in patients with high or low immune hotspot fractional score was investigated, the result is shown in *fig. 2C*. The Mann-Whitney U-test showed no significant difference was observed between the high and the low group (p = 0.31).

Test on the correlation between the immune hotspot fractional score and the three major prognostic factors

The result of JT trend test revealed that in both discovery and validation cohort the immune hotspot fractional score was not significantly correlated with the trend of three major prognostic factors used in the multivariate survival analysis (JT > 0.05) as shown in *fig. 2A 2B.*

Test on the prognostic value of the immune hotspot fractional score

The Kaplan-Meier survival plot for the immune hotspot fractional on the whole cohort is shown in *fig. 3A*. Higher value of immune hotspot fractional is associated with better overall survival. It is in consistent with the result in the univariate survival test in both cohorts. The results of the univariate and multivariate test are listed in *table 3*. Due to the missing value for some parameters recorded in the TCGA dataset, a noticeable change in the cohort size is observed that affecting not just each single parameter in the univariate survival test but also, even more extensively due to mismatching, the cohort size in the multivariate test. The size of the discovery and validation cohort in the multivariate test has been reduced to n=132 and n=127 respectively. The prognostic value of the immune hotspot fractional, when adjusted in the multivariate test using TNM stage, history of primary risk factor and the level of α-fetoprotein, was still significant in the discovery cohort but not in the validation cohort. For TNM stage, stage I was served as reference. Stage IV was associated with poor clinical outcome in both cohorts through the univariate and multivariate survival test. However, it's worth noticing that the number of patients with TNM stage score of IV is limited. The result for Stage III was similar to that of the Stage IV, except Stage III failed to show its prognostic value in both cohorts during the multivariate survival test. Patient's history of primary risk factor failed to show its prognostic value in any survival test performed in this study. However, an interesting trend that patients with risk factor may associated with a better clinical outcome was found through all four tests, though not significant. Low level of α-fetoprotein at diagnosis was associated with better overall survival. The trend was significant in discovery cohort across both univariate and multivariate survival test, while not in the validation cohort.

However, if looking at the whole cohort level, immune hotspot fractional was associated with significant prognostic value in the multivariate test (p = 0.001 HR = 2.29 (1.38 – 3.8)) as shown in *fig. 3B*.

Patients with low immune hotspot fractional may be further stratified by lymphocyte abundance, although not significant

As the lymphocyte abundance would potentially influence the behavior of lymphocytes clustering inside the tumour. We investigated the correlation between lymphocyte abundance and the immune hotspot fractional and associated such correlation with survival to see its further implication on patient's clinical outcome. As mentioned before, no significant difference is observed in the lymphocyte abundance between immune hotspot fractional high and low group. A further scatter plot was done to

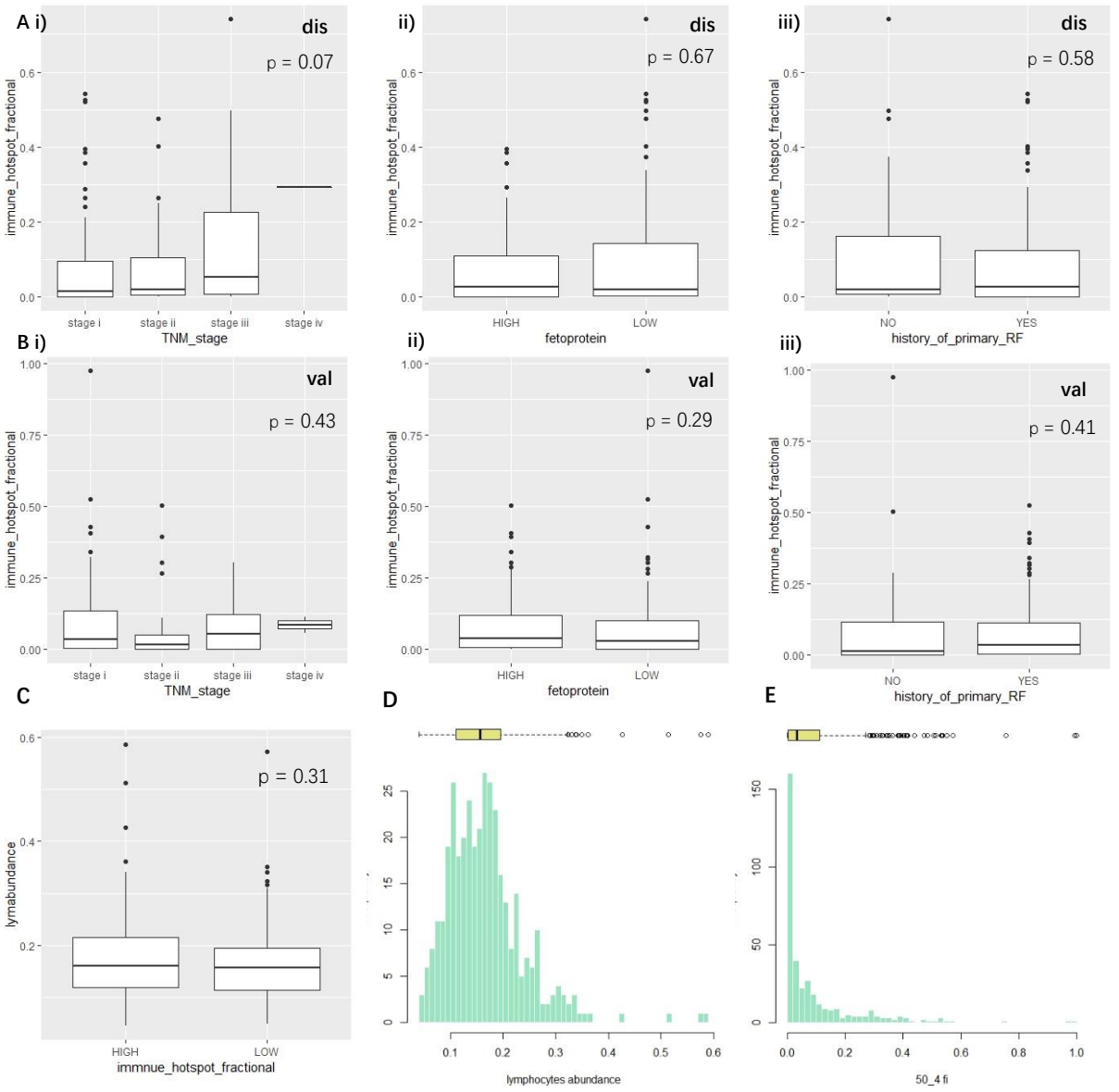

Figure 2 Test the immune hotspot fractional and the lymphocyte abundance's distribution and correlation in the whole cohort and test the correlation between immune hotspot fractional and the three major prognostic factors in discovery and validation cohort. (A) The immune hotspot fractional score was not correlated to the trend of three major prognostic factors in discovery cohort. The method is based on JT trend test. *(i)* correlation with TNM stage (p = 0.07). *(ii)* correlation with α-fetoprotein level at diagnosis (p = 0.67). *(iii)* correlation with patient's history of primary risk factor (p = 0.58). **(B)** The immune hotspot fractional score was not correlated to the trend of three major prognostic factors in validation cohort. The method is based on JT trend test. *(i)* correlation with TNM stage (p = 0.43). *(ii)* correlation with α-fetoprotein level at diagnosis (p = 0.29). *(iii)* correlation with patient's history of primary risk factor (p = 0.41). **(C)** No significant difference was observed in the level of lymphocyte abundance between patients with high and low immune hotspot fractional. Patients were divided into two groups using their immune hotspot fractional score. The cut-off point was the same to the one used in survival analysis. The lymphocyte abundance was compared between the two group, a Man-Whitney U test was used (p = 0.31). **(D)** Patient's distribution pattern of lymphocyte abundance. The median is 0.160, the interquartile range is 0.114 – 0.198. **(E)** Patient's distribution pattern of immune hotspot fractional score. The median is 0.032, the interquartile range is 0.003 – 0.110.

visualize the distribution of the lymphocyte abundance and the immune hotspot fractional. As no cut-off point was found that could divide the patients into two groups with distinctive clinical outcome using their lymphocyte abundance. The median was used to divide the lymphocyte abundance into high and low group as shown in *fig. 4A*. A Kaplan-Meier survival plot combined the status of both lymphocyte abundance and the immune hotspot fractional is shown in *fig. 4B*. The difference in the abundance of lymphocytes didn't further stratify patients in the high immune hotspot fractional group. However, in patients with low immune hotspot fractional, those with lower lymphocyte abundance, although not significant (p=0.11), may associate with worse overall survival, where the survival curve of patients with low immune hotspot fractional but high lymphocyte abundance looked closer to the survival curve of high immune hotspot fractional group.

The gene transcription level of NKG2D and its ligands within the tumour tissue region was not correlated with the density of lymphocytes or the immune hotspot fractional score

		Univariate analysis			*Multivariate analysis*		
		p-value	HR	n	p-value	HR	n
Discovery							
fi	High (ref)			64			48
	Low	0.027	1.9 (1.1-3.4)	114	0.002	3.5 (1.6-7.5)	84
TNM stage	I (ref)			87			76
	II	0.18	1.6 (0.8-3.3)	40	0.13	1.9 (0.8-4.1)	29
	III	0.002	2.6 (1.4-4.7)	43	0.073	2.1 (0.9-4.8)	26
	IV	0.005	19.4 (2.4-154.4)	1	0.002	35.6 (3.7-341.6)	1
History of primary RF	No (ref)			41			30
	Yes	0.12	0.6 (0.4-1.1)	132	0.212	0.63 (0.3-1.3)	102
α-fetoprotein	High (ref)			55			50
	Low	***	0.3 (0.2-0.5)	86	***	0.3 (0.2-0.6)	82
Validation							
fi	High (ref)			75			52
	Low	0.029	1.8 (1.1-3)	106	0.186	1.6 (0.8-3.2)	75
TNM stage	I (ref)			85			71
	II	0.82	1.1 (0.5-2.2)	44	0.792	0.9 (0.3-2.3)	32
	III	0.006	2.3 (1.3-4.3)	35	0.214	1.7 (0.8-3.7)	22
	IV	0.03	3.8 (1.1-12.8)	4	0.029	5.6 (1.2-26.6)	2
History of primary RF	No (ref)			49			30
	Yes	0.055	0.6 (0.4-1)	126	0.203	0.6 (0.3-1.3)	97
α-fetoprotein	High (ref)			74			67
	Low	0.435	1.3 (0.68-2.4)	65	0.188	1.6 (0.8-3.2)	60

**** denoted the p-value is less than 0.001*

Table 3 Result of the survival analysis of the immune hotspot fractional score in discovery and validation cohort. This table showed the result of univariate and multivariate survival analysis in discovery and validation cohort. Patient's pathological TNM score, history of primary risk factor and the α-fetoprotein level at diagnosis were used in the multivariate analysis. Their prognostic value was also examined in the univariate survival analysis. Due to the missing record in some patients in the TCGA dataset. The cohort size in the univariate and multivariate survival analysis was different. The number of patients in each cohort was labelled under section 'n'. The method was based on the cox proportional hazard model. The result showed that the immune hotspot fractional is an independent prognostic factor in the discovery cohort (p = 0.002, HR = 3.5(1.6 – 7.5)) but not in the validation cohort (p = 0.186, HR = 1.6 (0.8 – 3.2)). Method was based on the cox proportional hazard model.

The Spearman's rank correlation test showed that the gene transcription data of *KLRK1, ULBP1, ULBP2, ULBP3, MICA* and *MICB* was associated with weak correlation with the density of lymphocytes ($0.1 < r < 0.33$) as shown in *fig. 5A*. Although the p-value indicated the correlation observed is significant, based on the appealing of the plot, it could be largely caused by the number of datapoints rather than a true correlation. As the tumour tissue section used in this study mostly resembled a TME area, our finding was revealed similarity with the report from Guerra and colleagues, where they claimed the level of CD8+ T cell was the same in mice with different expression level of NKG2D. The correlation test between the gene transcription profile and the immune hotspot fractional score also revealed a similar situation as observed in the density of lymphocytes. The correlation was near random as shown in *fig. 5B* ($0.0081 < r < 0.17$).

Discussion

The result has shown that the immune hotspot fractional score under the spatial configuration of 50 μm grid size and the fourth order of neighborhood size is a potent prognostic factor in patients with primary HCC. Although the independence of its prognostic value was insignificant in the validation cohort, its performance had proved the value of investigating the spatial distribution. Especially, it was the spatial distribution, not the lymphocyte abundance or infiltration, that associated with prognostic value. Patients with high immune hotspot fractional score were associated with better overall survival. Although the finding that the TNM stage IV was associated with significant prognostic value in both univariate and multivariate survival analysis across two cohorts fits the fact that TNM stage is a powerful prognostic factor. Still, it should be noticed that the number of patients with TNM stage IV in the multivariate survival analysis was just above zero (discovery n = 1, validation n = 2). It raised concerns that whether there was enough statistical power to substantiate the finding. Moreover, as

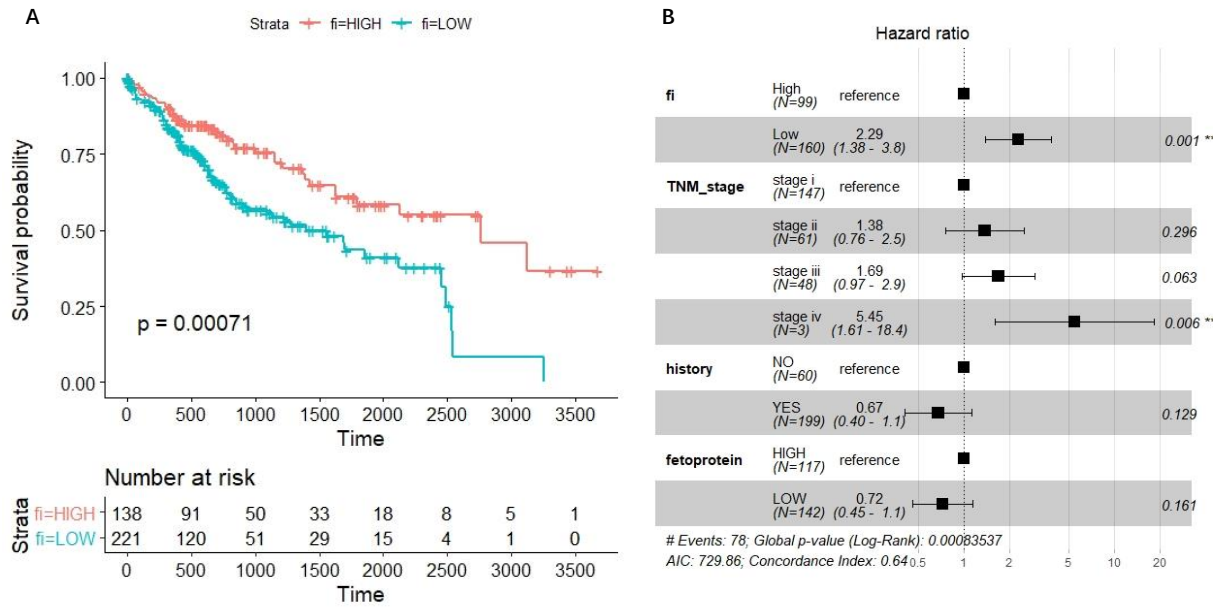

Figure 3 Survival analysis on the whole cohort. (A) The Kaplan-Meier survival curve of immune hotspot fractional score in the whole cohort of patients (time in days). Patients with high immune hotspot fractional score were associated with better survival (p = 0.00071). **(B)** The multivariate survival analysis on the whole cohort. Immune hotspot fractional (p = 0.001, HR = 2.29 (1.38 – 3.8)) and TNM stage iv (p = 0.006, HR = 5.45 (1.61 – 18.4)) were shown as independent prognostic factor when looking at the whole cohort. The method was based on the cox proportional hazard model.

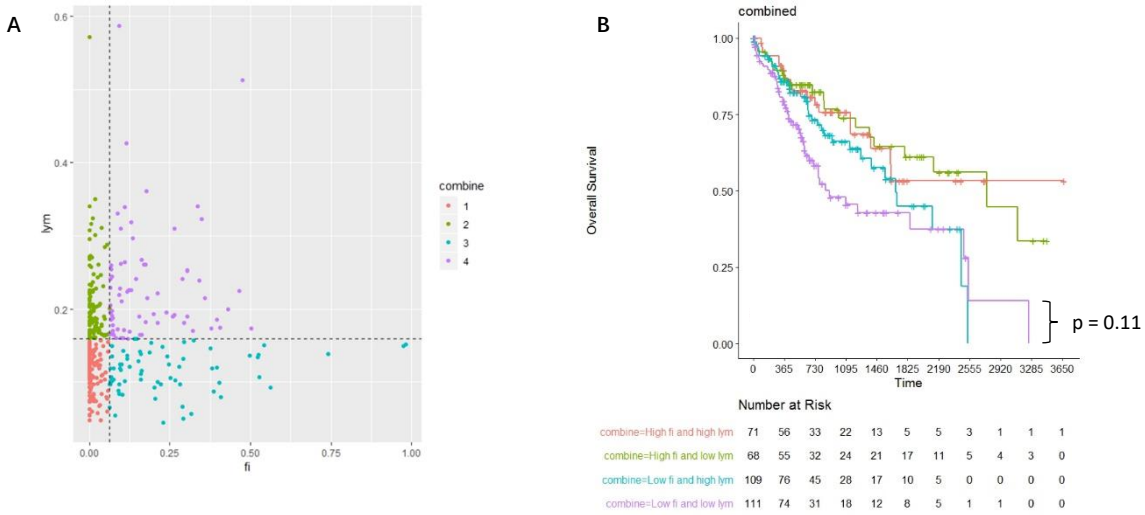

Figure 4 Lymphocyte abundance could further stratify patients with low immune hotspot fractional score although not significant. (A) Using scatterplot to illustrate the correlation between patient's lymphocyte abundance (lym) and immune hotspot fractional score. Immune hotspot fractional was divided by the cut-off point used in the survival analysis. The lymphocyte abundance was divided by median as no cut-off point was found to associated with significant prognostic value. **(B)** Survival plot was done to test if patient's lymphocyte abundance can further stratify patients with high and low immune hotspot fractional score (time in days). In patients with high immune hotspot fractional score, the trend of survival curve of patients with high and low lymphocyte abundance was close. In patients with low immune hotspot fractional score, those with a lower lymphocyte abundance was associated with worse survival. However, this difference was not significant (p = 0.11). Still, judged by the appealing of the curve, patients with low lymphocyte abundance and immune hotspot fractional score faced an unfavourable overall survival in the first four years.

shown in *fig. 2A (i) 2B(i),* the immune hotspot fractional status of these three patients were all 'high', which means their predicted clinical outcome using TNM staging and immune hotspot fractional score was completely opposite with no patients to divide the stage IV using immune hotspot fractional score. Thus, there are possibilities that the finding in the multivariate survival analysis was due to pure chance. More patients with TNM stage IV should be recruited into the study. Also, it might be able to explain that the independence of immune hotspot fractional's prognostic value was restored when looking at the whole cohort. For the level of α-fetoprotein, the different findings in its prognostic value in the discovery and validation cohort might be due to an unbalanced grouping. As the result of the multivariate survival analysis in the whole cohort revealed no significant prognostic value was found for the α-fetoprotein level.

In this study, an interesting tendency was found in the prognostic value of the patient's history of primary risk

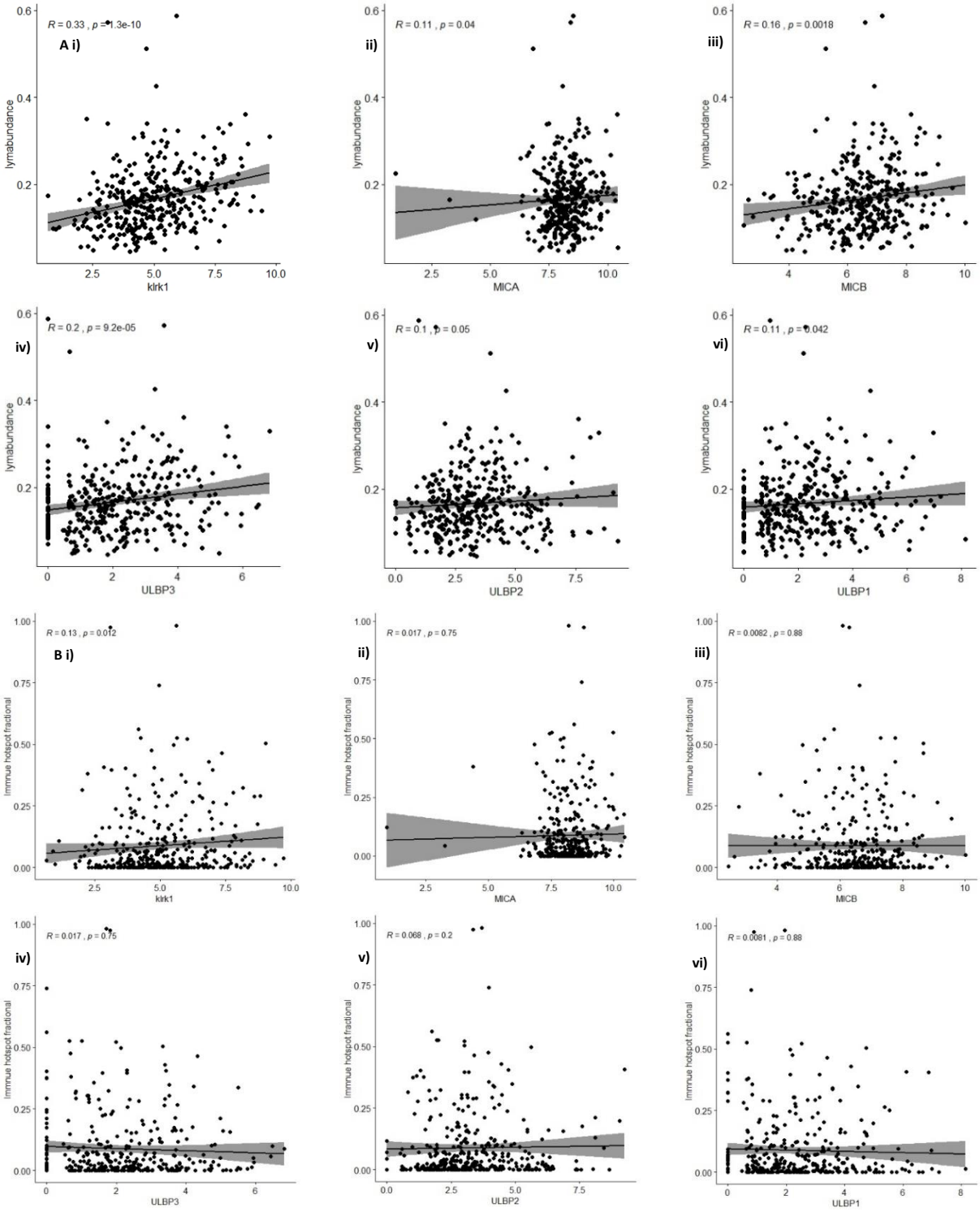

Figure 5 Test on the immune hotspot fractional score and lymphocyte abundance's correlation with NKG2D and its ligands gene transcription data using Spearman's rank correlation test. (A) The patient's lymphocyte abundance had a weak correlation with NKG2D and its ligands' gene transcription data measured within the same tissue block (0.1 < r < 0.33). *(i)* correlation with *KLRK1* (r = 0.33). *(ii)* correlation with *MICA* (r = 0.11). *(iii)* correlation with *MICB* (r = 0.16). *(iv)* correlation with *ULBP3* (r = 0.2). *(v)* correlation with *ULBP2* (r = 0.1). *(vi)* correlation with *ULBP1* (r = 0.11). **(B)** The patient's immune hotspot fractional score was not correlated with NKG2D and its ligands' gene transcription data measured within the same tissue block (0.0081 < r < 0.17). *(i)* correlation with *KLRK1* (r = 0.13). *(ii)* correlation with *MICA* (r = 0.017). *(iii)* correlation with *MICB* (r = 0.0082). *(iv)* correlation with *ULBP3* (r = 0.017). *(v)* correlation with *ULBP2* (r = 0.068). *(vi)* correlation with *ULBP1* (r = 0.0081).

factor. Patients with a history of risk factor were actually associated with longer overall survival. Although not significant, this finding was consistent across all survival analysis done in this study. This seemingly counter-intuitive finding might be explained by the diversity of patient's background. As mentioned earlier, such diversity greatly contributed to the complexity of HCC, it also made the identification of a globally applicable prognostic factor a

near impossible task. The difference in the prevalence of being threatened by different kinds of risk factors that caused by the distinction in genetic compositions and living habits contributed most to the diversity of patient's backgrounds. However, since the site of hospitals involved in TCGA projects was all located in the United States, it brought about another possibility to this diversity – migrants. It has been pointed out that the migrants had a different incidence in some disease as their living habits were changed (*27*). The genetic composition could also be altered as a result of mating with the local residence. However, the TCGA's recording in patient's ethnicity was too general to aid a proper grouping by patient's background. This is quite unfriendly to researchers in a disease type with dependency to the patient's background, especially at the beginning stage of identifying a novel multidisciplinary based prognostic factor. Thus, a new project with simplex composition of patients and with emphasizing on the image data collection should be proposed in the future for a thorough investigation.

The biological significance of quantifying the spatial distribution started with the understanding of cancer evolution. It has been established that the intratumor heterogeneity existed widely across most type of cancer (*6*). Two different regions within the same tumour site could be treated as two different tumour under such heterogeneity (*6*). The way that cancer interacts with the microenvironment it's located in is complicated. The ability of tumour cells to passively react to the microenvironment, in favors of its progression or metastasis, by altering its functional mutation profile was revealed by a study using multiregion sequencing (*28*). In that study, it was found that more than a half of all somatic mutation was not detectable across all tumour region (*28*). Tumour cells can also actively modify the microenvironment in its favor by means like immune evasion (*6*). The behavior of cancer cell in its progression is like an adaption that follows the Darwinian selection theory (*28*). Back to the heterogeneity of the spatial distribution, the tissue section image is like a snapshot of a certain timepoint in the life of cancer development, the spatial distribution of cancer and immune cells could reveal valuable information about the intension of cancer and the immune system. The ability to decipher such information is of great value in providing individualized therapy regimen.

As mentioned above the Getis-Ord hotspot analysis is a measurement method for ecological purpose. One straightforward example of its usage is for the police force to discover areas with high crime rate, so they could assign their resources accordingly. This usage scenario shares a lot of similarity with what we want to achieve in interpreting tumour microenvironment. The area that was deemed as cancer hotspot where cancer cells were clustered together could indicate the existence of a cancer nest. And the cancer hotspot fractional, which reflected the area that been penetrated by high density of lymphocyte cluster, could indicate the proportion of cancer hotspot that is currently under the effect of immune system. While the immune hotspot, denoted by the clustering of lymphocytes, may have other indications. An immune hotspot that was not co-localized with a cancer hotspot could suggest a successful previous rejection of cancer development, where the selective pressure forced cancer cells to progress through other directions. Or it could be aggregated in response to a false decoy released by the cancer cell, as part of the immune evasive mechanism. However, this possibility was questionable as the immune evasion will ultimately lead to immune suppression, and the histopathological feature of an immune suppressive area, as reported in a previous study (*6*), was distinctive to the feature of a hotspot. The immune hotspot that was co-localized with cancer hotspot could be deemed as favorable to the patient's outcome. As it represented a desirable concentration of immune activity. As the statistical meaning of the immune hotspot fractional was the proportional area of immune hotspot that is co-localized with cancer hotspot. The biological meaning of this spatial score can be interpreted as the level of immune cells infiltrated into the tumour area and aggregated with high density of cancer cells. Or in an abstract terms, it denoted the proportion of immune hotspot that was actually fighting the enemy. The biological implication of these spatial scores might be too general as we were limited by the knowledge of which cell group that contributed most to the hotspot. This question should be addressed in future study. As previously discovered in NKG2D deficient mice, the presence of anti-tumor immune activity in the liver favored a tumorigenesis that may increase the tumour burden (*13*). Yet, our study revealed a favorable role of immune activity in terms of patient's overall survival by analyzing the spatial distribution. Our finding actually helped to explain the seemingly controversial findings in human, as high score of immune hotspot fractional not just indicated the intensity of immune activity, but also suggested less friendly fire.

The internal factors that drive the clustering pattern of the cancer and immune cells during the cancer development were still unknown. Whether it was contributed solely by a single gene and its associated pathway or by a collective of mechanisms at a higher level was still under investigation. The key finding that leads to this study – NKG2D was shown to only have a weak correlation with the density of lymphocytes or the lymphocytes and cancer cells' clustering pattern in the patient's TME area with a near random correlation in the latter. However, since the NKG2D's expression in hepatocytes was only restricted in lymphocytes. It can be argued that the correlation between the expression of *KLRK1* and the lymphocyte abundance as shown in *fig. 5A (i)* was actually significant, because the way that TCGA measured the gene transcription data can't reflect the expression heterogeneity between cell types. To make things more complicated, Thomas Spies and colleagues claimed the expression of NKG2D receptor was observed on the surface of cancer cell in breast, colon, ovarian and prostate cancer (*29*). This claim was controversial against the common thinking in the field. What's more, the expression of surface NKG2D was absent

in most tumour cell lines and in almost all mice tumour model (*29*). There is no evidence for NKG2D expression in liver cancer cell. Thus, we could assume that there is a significant positive correlation between the *KLRK1* gene expression level and the lymphocyte abundance within the TME area in patients with primary HCC. However, a further investigation using lymphocytes as the sole source in testing the NKG2D-associated gene transcription data must be carried out to substantiate such correlation.

Algorithm-wide, CRImage was not the only program available for extracting cell classification and location data from tissue section staining images. And in TCGA dataset, tissue sections that were treated by fresh froze were also not the only source of histopathological images. For TCGA-LIHC dataset specifically, there was another source of such images that were processed for patient's diagnostic purpose. These slides were termed diagnostic slides in TCGA dataset and were treated with formalin and embed in paraffin. Thus, they can preserve much more morphological feather than the fresh frozen tissue sections. Plus, tissue sections and diagnostic sections represented the spatial information collected in two different time points. The tissue section was processed using the tumour mass that been removed from patients. Obviously, any analysis done using tissue section was of little value in deciding individualized therapy regimen based on the fact of tumour heterogeneity. If the spatial score acquired from diagnostic slides was of the same power in prognosing, we would be able to better stratify the patients at the beginning of the treatment. Thus, the same spatial analysis approach should also be applied on the diagnostic slides. However, the CRImage was not designed to handle image other than fresh frozen treated ones. Thus, such analysis was not done in this study. This represented the major limitation of this study, and the importance in analyzing diagnostic slides should be highlighted in future study.

In conclusion, this study examined the prognostic value of spatial score which represented the spatial distribution of the immune and cancer cells in the tumour tissue section in patients with primary HCC. The immune hotspot fractional was found to associate with strong prognostic value in the TCGA dataset which was composed by a mixture of patients with different backgrounds. It was the first time that the spatial distribution of lymphocytes and cancer cells were found to have the ability to predict patient's clinical outcome in HCC. And the performance of the immune hotspot fractional score proved itself a promising candidate in the competition of a new standard in stratifying patients with primary HCC. Future study is required to finalize the investigation of the spatial score's value in predicting patient's outcome as this study still had areas unchecked. And the biological significance of the spatial heterogeneity was still awaited to be examined in a more systemic way.

Acknowledgments

The author wants to thank to Dr Nadia Guerra, for her detailed and valuable investigation in the mice model which inspired this study. Also, many thanks to Dr Yinyin Yuan and the group member in the lab for their generous help and patience in the guiding the technical details in the bioinformatic analysis.

Abbreviations

HCC, hepatocellular carcinoma; DEN, diethylnitrosamine; H&E, hematoxylin and eosin; TCGA, the cancer genome atlas; UCSC, University of California, Santa Cruz; ER, estrogen receptor; NKG2D, natural killer group 2 member D; *KLRK1*, killer cell lectin like receptor K1 gene; ULBP, UL16 binding protein; MICA, MHC class I polypeptide-related sequence A; MICB, MHC class I polypeptide-related sequence B; MHC, the major histocompatibility complex; TME, tumour microenvironment; *fi,* immune hotspot fractional; *fc,* cancer hotspot fractional; *fci,* cancer-immune hotspot co-localization; JT, Jonckheere-Terpstra.

References

1. N. Guerra *et al.*, NKG2D-deficient mice are defective in tumor surveillance in models of spontaneous malignancy. *Immunity* **28**, 571-580 (2008).
2. J. Galon *et al.*, Type, density, and location of immune cells within human colorectal tumors predict clinical outcome. *Science* **313**, 1960-1964 (2006).
3. Y. Issa-Nummer *et al.*, Prospective validation of immunological infiltrate for prediction of response to neoadjuvant chemotherapy in HER2-negative breast cancer--a substudy of the neoadjuvant GeparQuinto trial. *PLoS One* **8**, e79775 (2013).
4. S. Loi *et al.*, Prognostic and predictive value of tumor-infiltrating lymphocytes in a phase III randomized adjuvant breast cancer trial in node-positive breast cancer comparing the addition of docetaxel to doxorubicin with doxorubicin-based chemotherapy: BIG 02-98. *J Clin Oncol* **31**, 860-867 (2013).
5. S. M. Mahmoud *et al.*, Tumor-infiltrating CD8+ lymphocytes predict clinical outcome in breast cancer. *J Clin Oncol* **29**, 1949-1955 (2011).
6. J. Galon, D. Bruni, Approaches to treat immune hot, altered and cold tumours with combination immunotherapies. *Nat Rev Drug Discov* **18**, 197-218 (2019).
7. B. Mlecnik *et al.*, Histopathologic-based prognostic factors of colorectal cancers are associated with the state of the local immune reaction. *J Clin Oncol* **29**, 610-618 (2011).
8. K. Ogasawara, J. Benjamin, R. Takaki, J. H. Phillips, L. L. Lanier, Function of NKG2D in natural killer cell-mediated rejection of mouse bone marrow grafts. *Nat Immunol* **6**, 938-945 (2005).
9. J. Galon, H. K. Angell, D. Bedognetti, F. M. Marincola, The continuum of cancer immunosurveillance: prognostic, predictive, and mechanistic signatures. *Immunity* **39**, 11-26 (2013).
10. S. Nawaz, A. Heindl, K. Koelble, Y. Yuan, Beyond immune density: critical role of spatial

heterogeneity in estrogen receptor-negative breast cancer. *Mod Pathol* **28**, 766-777 (2015).

11. A. P. Venook, C. Papandreou, J. Furuse, L. L. de Guevara, The incidence and epidemiology of hepatocellular carcinoma: a global and regional perspective. *Oncologist* **15 Suppl 4**, 5-13 (2010).
12. J. A. Marrero, M. Kudo, J. P. Bronowicki, The challenge of prognosis and staging for hepatocellular carcinoma. *Oncologist* **15 Suppl 4**, 23-33 (2010).
13. S. Sheppard *et al.*, The immunoreceptor NKG2D promotes tumour growth in a model of hepatocellular carcinoma. *Nat Commun* **8**, 13930 (2017).
14. D. H. Raulet, N. Guerra, Oncogenic stress sensed by the immune system: role of natural killer cell receptors. *Nat Rev Immunol* **9**, 568-580 (2009).
15. R. A. Eagle, J. Trowsdale, Promiscuity and the single receptor: NKG2D. *Nat Rev Immunol* **7**, 737-744 (2007).
16. L. L. Lanier, NKG2D Receptor and Its Ligands in Host Defense. *Cancer Immunol Res* **3**, 575-582 (2015).
17. N. Nausch, A. Cerwenka, NKG2D ligands in tumor immunity. *Oncogene* **27**, 5944-5958 (2008).
18. S. Sheppard, A. Ferry, J. Guedes, N. Guerra, The Paradoxical Role of NKG2D in Cancer Immunity. *Front Immunol* **9**, 1808 (2018).
19. A. Heindl *et al.*, Microenvironmental niche divergence shapes BRCA1-dysregulated ovarian cancer morphological plasticity. *Nat Commun* **9**, 3917 (2018).
20. A. Heindl *et al.*, Relevance of Spatial Heterogeneity of Immune Infiltration for Predicting Risk of Recurrence After Endocrine Therapy of ER+ Breast Cancer. *J Natl Cancer Inst* **110**, (2018).
21. Y. Y. Failmezger H, Rueda O, Markowetz F, CRImage: CRImage a package to classify cells and calculate tumour cellularity. R package version 1.30.0. . (2018).
22. O. J. Getis A, The analysis of spatial association by use of distance statistics. *Geogr Anal*, 24 (1992).
23. S. Anders, P. T. Pyl, W. Huber, HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166-169 (2015).
24. K. Francis, B. O. Palsson, Effective intercellular communication distances are determined by the relative time constants for cyto/chemokine secretion and diffusion. *Proc Natl Acad Sci U S A* **94**, 12258-12262 (1997).
25. A. K. a. M. Kosinski, survminer: Drawing Survival Curves using 'ggplot2'. (2018).
26. R. C. Team}, R: A Language and Environment for Statistical Computing. (2018).
27. P. Vineis, C. P. Wild, Global cancer patterns: causes and prevention. *Lancet* **383**, 549-557 (2014).
28. M. Gerlinger *et al.*, Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *N Engl J Med* **366**, 883-892 (2012).
29. A. C. Benitez *et al.*, Expression, signaling proficiency, and stimulatory function of the NKG2D lymphocyte receptor in human cancer cells. *Proc Natl Acad Sci U S A* **108**, 4081-4086 (2011).