Spatial analysis of ligand-receptor interactions in skin cancer at genome-wide and single-cell resolution

Tran M[1]*, Yoon S[2]*, Min ST[1,3], Andersen S[2,4], Devitt K[3], Lam PY[1], Purdue B[2], Raghubar A[1], Hanson SJ[5], Jones K[1], Walters S[6], Tuong ZK[3,7,8], Soyer HP[9], Frazer I[3], Nguyen Q[1#]

[1] Institute for Molecular Bioscience, The University of Queensland, Australia

[2] Genome Innovation Hub, The University of Queensland, Australia

[3] The University of Queensland Diamantina Institute, Faculty of Medicine, The University of Queensland, Australia

[4] IMB Sequencing Facility, Institute for Molecular Bioscience, The University of Queensland, Australia

[5] School of Medical Science, Menzies Health Institute, Griffith University, Australia

[6] School of Biomedical Sciences, The University of Queensland, Australia

[7] Molecular Immunity Unit, University of Cambridge Department of Medicine, MRC-Laboratory of Molecular Biology, UK

[8] Cellular Genetics, Wellcome Sanger Institute, Hinxton, UK.

[9] The University of Queensland Diamantina Institute, The University of Queensland, Dermatology Research Center, Australia

* These authors contributed equally to this work

[#]Corresponding author: Dr. Quan Nguyen: quan.nguyen@uq.edu.au

Abstract

The ability to study cancer-immune cell communication across the whole tumor section without tissue dissociation is important to understand molecular mechanisms of cancer immunotherapy and drug targets. Current experimental methods such as immunohistochemistry allow researchers to investigate a small number of cells or a limited number of ligand-receptor pairs at tissue scale with limited cellular resolution. In this work, we developed a powerful experimental and analytical pipeline that allows for the genome-wide discovery and targeted validation of cellular communication. By profiling thousands of genes, spatial transcriptomic and single-cell RNA sequencing data show genes that are possibly involved in interactions. The expression of the candidate genes could be visualized by single-molecule *in situ* hybridization and droplet digital PCR. We developed a computational pipeline called STRISH that enables us to quantitatively model cell-cell interactions by automatically scanning for local expression of RNAscope data to recapitulate an interaction landscape across the whole tissue. Furthermore, we showed the strong correlation of microscopic RNAscope imaging data analyzed by STRISH with the gene expression

values measured by droplet digital PCR. We validated the unique ability of this approach to discover new cell-cell interactions *in situ* through analysis of two types of cancer, basal cell carcinoma and squamous cell carcinoma. We expect that the approach described here will help to discover and validate ligand receptor interactions in different biological contexts such as immune-cancer cell interactions within a tumor.

Keywords: ligand-receptor, cell-cell interactions, spatial transcriptomics, single cell sequencing, single-cell genomic, RNAscope, ddPCR

Introduction

Cell-to-cell communication (CCC), underscores a dynamic cellular ecosystem that develops, evolves, and responds to environmental factors. The implications and roles of CCC have been investigated extensively, particularly in cancer, using diverse *in vitro* and *in vivo* systems, albeit at different scales and resolutions (Brücher et al. 2014). Breakthroughs arising from discoveries in CCC may have important clinical significance, such as the discovery of cancer-immune cell interactions via immune checkpoint proteins PD-1 (Programmed Cell Death 1), and its ligand PD-L1, and CTLA4 (cytotoxic T-lymphocyte-associated protein 4) (Pardoll 2012); tumor cells, tumor infiltrating lymphocytes and tumor associated myeloid cells express inhibitory PD-L1/CTLA4 ligands to engage PD-1 receptors on cytotoxic T cells and CD80/86 receptors on myeloid cells, effectively blocking immune activation against the tumor cells. The discovery has led to experimental applications of using monoclonal antibodies that specifically target this ligand-receptor interaction as a form of immunotherapy.

Although thousands of cancer patients have been cured using this strategy, often less than 30% of patients respond to a single immunotherapy, leading to the need to combine therapies, for example by using both PD-L1 and CTLA4 to target PD-1 and CD80/86 respectively (Ott et al. 2017). However, rationalization of immunotherapy agent combinations remains challenging. Therefore, research to explore and advance understanding of known and new ligand-receptor pairs in the context of tumor-immune cell interactions within a tumor is extremely important for the further development of immunotherapies. Most research so far has focused on the use of fluorescently-conjugated antibody-based methods, that are relatively low-throughput due to only being able to assess protein levels of a few target molecules at a time. Analysis of the transcriptome provides a means towards high-throughput assays; however, a comprehensive and unbiased pipeline from discovery to validation of cellular communications from transcriptomics

data within an intact tissue is still lacking. In this work, we aim to establish a pipeline to study ligand-receptor interactions of cancer and immune cells across the whole tissue section, using skin cancer as a model. Recent developments in single-cell and single-nuclei RNAseq (scRNA-seq/snRNAseq) have provided the opportunity to explore and dissect gene expression profiles at single cell level, but these techniques remain inadequate in gathering information with a spatial context. While inference methods to predict CCC using single-cell data are being developed (Browaeys et al. 2020; Efremova et al. 2020), they still lack the ability to infer spatial context of the interactions. Spatial transcriptomics (ST-seq) and RNA *in situ* hybridization (ISH) technologies overcome these limits and enable the study of (target) gene expression in intact tissue sections, maintaining tissue integrity (Salmén et al. 2018). Only recently, Ji et al. generated spatial transcriptomic data on human squamous cell carcinoma (SCC) combining scRNA-seq and multiplexed ion beam imaging analysis. In the study, integration of multi-modal data revealed ligand-receptor (L-R) networks specific to stromal and tumor cells (Ji et al. 2020).

In the current study, we use four complementary technologies to study and validate L-R interaction in skin cancer patients: ST-seq, scRNA-seq, RNA-ISH and droplet digital PCR (ddPCR). ST-seq measures gene expression in spots in microarray-based ST-seq techniques and consists of a mixture of cells which may be heterogeneous. Being able to preserve the cell morphology, ST-seq is an advanced way to study cell-cell interactions and disease mechanisms. ST-seq has been applied to studies on the landscape of tissues and diseases, such as prostate cancer (Ji et al. 2020; Berglund et al. 2018), pancreatic cancer (Moncada et al.), melanoma (Thrane et al. 2018) and allows for mapping the transcripts in the tissue sections. However, ST-seq still lacks single-cell resolution, and the number of transcriptomes (cell numbers) as well as the read quality that can be captured in each spot depend on the tissue context. These shortcomings of ST-seq can be overcome by a targeted approach such as RNA-ISH to visualize the L-R interaction at a single cell level within the region of interest. The RNAscope HiPlex assay (ACD) has been developed based on the RNA-ISH technique and improved on the signal amplification and background suppression process compared to the previous version, allowing for visualization and detection of mRNA at a higher degree. Using two adjoining spacers up to 50 base regions, the probes (spacers) specifically bind to target gene sequences, and then a pre-amplifier binds on the upper region of the spacer, maximizing fluorescence intensity. The technology allows researchers to simultaneously detect up to 12 single target genes on the same tissue section through a couple of fluorophore cleavage steps. Knowing that RNA-ISH combines sensitivity and precise localization of genes of interest in tissue, ddPCR can be used to confirm

target gene expressions in the same tissue block with extremely high sensitivity. ddPCR is the third generation of the PCR assay and does not require a standard curve for absolute quantification of target genes (Li et al. 2018; Kuypers and Jerome 2017). This technique generates millions of droplets that contain an input molecule and then partition them to be read as either a positive or negative signal according to fluorescence amplitude that is presented.

Therefore, we implemented these innovative technologies and validated our findings using automated ddPCR assay for absolute quantification on target genes in the same patients. We selected five genes that are expressed by tissue-associated immune cells - Langerin/CD207, interleukin-34 (IL34), colony-stimulating factor 1 receptor (CSF1R), THY1 (also known as CD90) and Integrin subunit alpha M (ITGAM, also known as CD11b) to study cell-cell interactions in skin cancers as a proof-of-concept. IL34 plays an essential role in innate immunity and tumor microenvironment (TME) yet the role of this ligand and its receptor, CSF1R is poorly characterized in skin cancers (Guillonneau et al. 2017). Here we report cell-to-cell interaction in two major forms of skin cancer using an experimental and bioinformatics pipeline consisting of cutting-edge ST-seq techniques along with RNA-ISH and ddPCR as validation tools. The results, while obtained from a limited sample number, demonstrate the quantitative assessment on co-expression of two target L-R pairs expressed by immune and non-immune cells in regions with distinct spatial localizations in cancerous tissue.

Results

A pipeline for studying ligand-receptor interaction from a cancer tissue section

In this work, we aim to build an experimental and analytical pipeline to comprehensively discover and validate L-R interactions at transcriptional level. To identify L-R pairs at a transcriptome-wide scale, we applied two approaches, starting with scRNA-seq and followed by ST-seq, for skin cancer samples (**Fig. 1-A**). Most ligands and receptors are expressed at relatively low level, leading to a high possibility for under-detection of those genes by using scRNA-seq and ST-seq. Therefore, we applied scRNA-seq and ST-seq at the discovery phase to detect potential ligand-receptor pairs, followed by the sensitive validation of a selected set of genes, by using single-molecule detection methods: RNAscope and ddPCR. RNAscope generates spatial information at single-cell resolution and ddPCR is approximately 1,000 times more sensitive than RNAseq. We

selected IL34-CSF1R and THY1-ITGAM as the two L-R pairs for testing the detection limit for each step of the pipeline as their interaction has been reported by previous studies.

Genome-wide analysis of ligand-receptor interactions using single-cell sequencing

To infer potential L-R pairs that are likely used as means of CCCs, we applied NicheNet pipeline (Browaeys et al, 2020) on our two scRNA-seq datasets of SCC/basal cell carcinoma (BCC) and normal samples from a patient (**Supplemental Table S1**). NicheNet combines expression data with prior knowledge on gene signaling and gene regulatory networks to predict L-R pairs used by interacting cells (sender and receiver cells). If the approach using scRNA-seq was sufficiently sensitive, we expect to detect IL34-CSF1R and THY1-ITGAM interactions. We visualized the L-R prediction of top 20 ligands and their putative receptors and the cell clusters in UMAP (**Supplemental Fig. S1-A,B**). Visual inspection of neighboring cells in UMAP space suggests distinct cell clusters that express IL34 or CSF1R (**Supplemental Fig. S1-C**). However, THY1 and ITGAM were only expressed by a few cells with no distinct patterns of expression (**Supplemental Fig. S1-D**). The results from examining THY1-ITGAM clearly showed insensitivity of using scRNA-seq alone to study cell communication. Overall, we found that scRNA-seq can be utilized to study CCC but is limited by insensitivity and missing spatial information.

Genome-wide analysis of ligand-receptor interactions using spatial transcriptomics

We postulated that CCC could be better assessed by using ST-seq, which preserves neighborhood information of interacting cells. We performed ST-seq on one tissue section from a BCC patient (patient ID-E15, **Supplemental Table S1**). The two hematoxylin and eosin (H&E) images showed two slides of tissue that were sectioned from the same tissue block and annotated with cancer nest, immune infiltration and non-cancer regions (**Fig. 1-B,C**). From the same block, one section was used (**Fig. 1-B**) for ST-seq and another for RNAscope (**Fig. 1-C**). ST-seq and RNAscope data for these two sections were subsequently used to model cell-cell communication and the results were compared (**Fig. 1-D,E**). The detailed comparisons will be discussed in a later section.

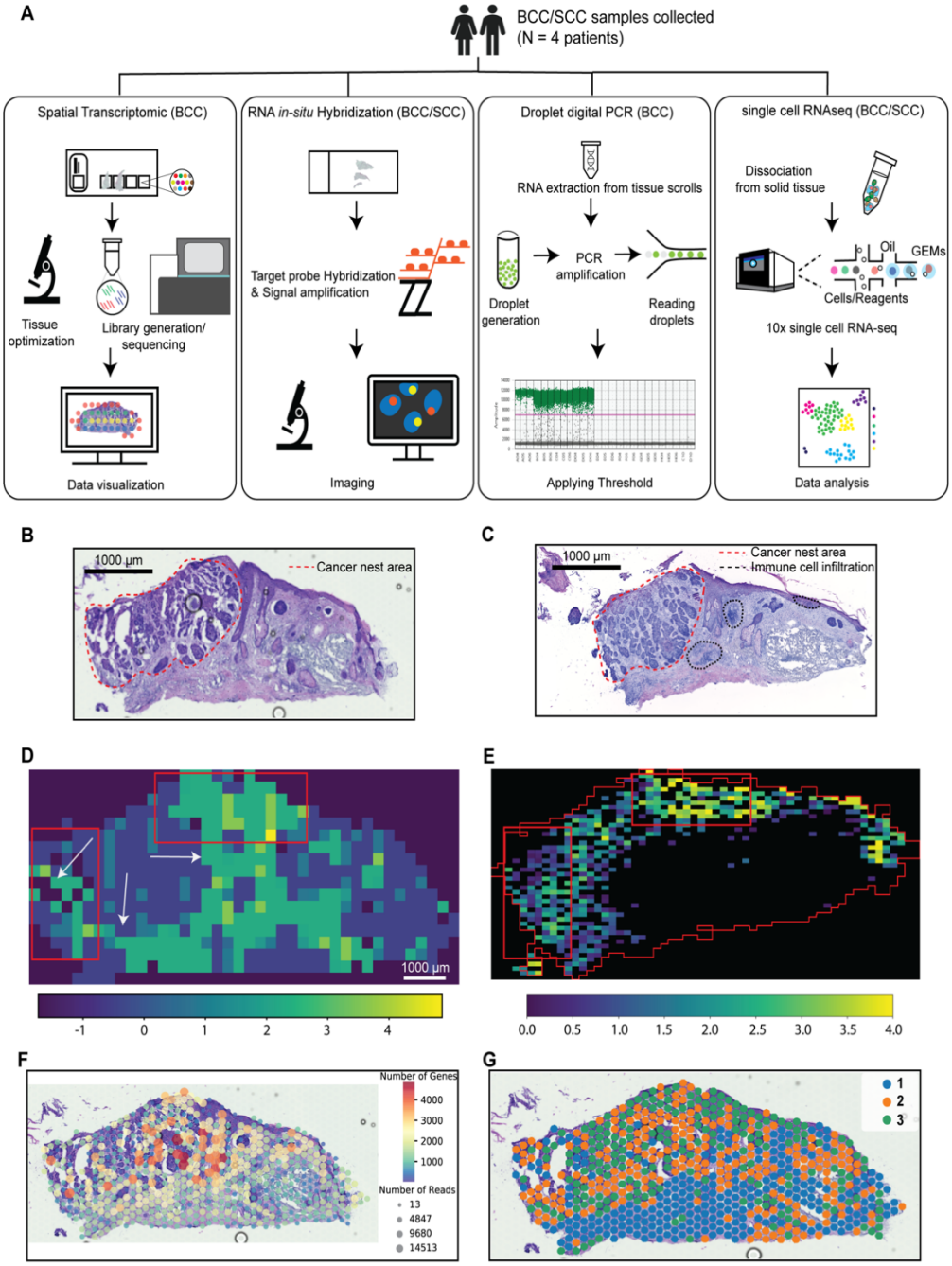

Figure 1. A pipeline to detect ligand-receptor interactions across the whole tissue section.
(A) A workflow illustrating the combination of four technologies to study ligand-receptor interactions in skin cancer tissue. The technologies include Visium Spatial Transcriptomic (ST), RNAscope Hiplex, automated ddPCR, and scRNA-seq. **(B)** The annotated H&E image of the same tissue section with ST data. **(C)** The annotated H&E for the tissue section that was used for the RNAscope. **(D)** The visualization of cell-to-cell interaction analysis using ST-seq information for a pair of ligand-receptor IL34-CSF1R. **(E)** The heatmap highlights the interaction of the same pair of ligand-receptor (IL34-CSF1R), detected by RNAscope. The ligand-receptor interaction landscape was produced by STRISH method. **(F)** The plot shows the variation of the number of transcripts that were captured in each spot across the tissue. **(G)** Spot clustering using gene expression showing distribution of clusters correlated to tissue morphology.

For the studying of CCC with ST-seq approach, we reasoned that each spot in the Visium slide contains a mixture of cells and these cells within a spot can communicate if both ligand and receptor are detected in that spot. In addition, the interaction between cells from two neighboring spots can happen if these spots display L-R co-expression. The local co-expression of L-R pairs within a spot or between neighboring spots suggests possible interactions. The result from CellphoneDB analysis pipeline showed that out of 948 possible combinations of L-R available in the ST-seq data, the sorted by descending order of significance of the interaction between CSF1R-IL34 ranked 700th while the interaction between THY1 and amb2 integrin complex (ITGAM pathway) ranked 835th. We generated the CCC heatmap (**Fig. 1-D**) from ST-seq data using stLearn as described in the methods section. Additionally, the heatmap showed high intensity of cell interactions in the left region of the hair follicle, surrounding the tumor periphery area and a few areas located in the middle left area of the tissue, highlighted by the white arrows. We observed low intensity of interaction in the center of the areas annotated to be the cancer nest, which can be explained by low number of genes and read counts (**Fig. 1-F**) and can be attributed to technical limitations of the permeabilization step. To minimize the technical issue of miss-detecting genes in highly-nucleated cancer region, we applied a normalization strategy as in stLearn that uses the correlation of tissue morphology and genes profiling to rescue genes with technical 0 values, by using the information from neighboring spots (**Supplemental Fig. 2-A**), (Pham et al, 2020). The shortcoming of the permeabilization process in the current spatial transcriptomics will likely be overcome in the future.

Unbiased clustering of gene expression using a graph-based approach on the 676 spots in the tissue section resulted in the demarcation of 3 distinct groups (**Fig. 1-G**). Compared to the H&E image (**Fig 1-B**), the distribution of cluster 2 and cluster 3 overlapped well with the cancerous areas, indicating that the gene expression of the spots around the cancer nest was uniquely distinct from the remaining areas while cluster 1 covered most of the non-cancerous areas. The feature plots of target ligand-receptor markers (**Supplemental Fig. 2-B,C**) illustrated low count values in many regions, potentially due to low depth reads and high dropout rate arising from technical limitations of ST-seq. Thus, if we were to only rely on scRNA-seq and ST-seq to study CCC in the tissue, there is a high possibility of missing information.

Single-cell resolution analysis of cell to cell interaction in skin cancer by RNAscope

To overcome the potential limitation of lack of signal in the cancer region of spatial transcriptomics and to achieve single cell resolution, we conducted RNAscope HiPlex assay to validate our

findings in Visium ST-seq and scRNA-seq analysis. The whole microscopy fluorescent images of tissue sections were captured at 40x magnification to scan for interactions at cellular resolution. Three different fluorophores (Cy3, Cy5 and Cy7) were used in two iterative wash-stain rounds to label the five distinct gene targets. Following image acquisition, the first image for THY1, IL34, and CSF1R and the second image for CD207 and ITGAM were overlaid to generate a single merged image (**Fig. 2-A**). The zoom-in images from a cancer nest area (a white dashed line) indicate a distinguishable cell to cell interaction for each pair compared to the 'No probe' control (IL34 and CSF1R in a red box and ITGAM and THY1 in a green box) (**Fig. 2-A**).

A computational pipeline, STRISH, to reconstruct tissue landscape of ligand-receptor interactions from RNAScope data

Tumor related immune infiltration cells and cancer nests are observed on both cancer tissues (**Supplemental Fig. S3-A,B; Supplemental Fig. S4-A,B**). To uncover the interaction of immune cells and cancerous cells in the whole BCC/SCC tissue section, we developed an analysis pipeline, called Spatial Transcriptomic and RNA *In-situ* Hybridization (STRISH). The pipeline uses a series of positive cells detection procedures to find regions that contain lower than 100 cells and determine the number of L-R co-expression within these regions (refer to the method section for more information). Across the tissue samples, we observed considerable interactions of IL34-CSF1R around the areas where the cancer nest is located in both BCC and SCC, particularly in the epidermal compartments (**Fig. 1-E; Fig. 2-B; Supplemental Fig. S3-A,B,C,E; S4-A,B**). In comparison with the result of using the ST-seq method to identify regions with CCC, we recorded the two areas that are consistent between two results highlighted with red boxes (**Fig. 1-D,E**).

We further assessed the performance of STRISH by measuring the interaction of another L-R pair, ITGAM-THY1. We found weaker and less wide-spread interactions of ITGAM-THY1, localized to the dermal compartments (**Fig. 2-C; Supplemental Fig. S3-A,B,D,F; Supplemental Fig. S4-A,B**). By comparing the number of tissue regions (STRISH windows) that the cell-cell interaction was found, we noticed the average count of windows for ITGAM and THY1 to locally-coexpress was 2.5 times less than the average count of windows for the pair of IL34-CSF1R. Similarly, we calculated the values of normalized local co-expression in the same local area incidents per STRISH window for IL34 and CSF1R were higher than that of ITGAM and THY1 (**Fig. 2-C**). We note that most of the STRISH-detected local co-expression of ITGAM and THY1 clustered densely around adjacent areas of immune cell infiltration (**Fig. 2-C**). In addition, ddPCR counted the markers in three tissues (**Fig. 2-D**). While the size of the patient tissues vary, THY1 appears to be the highest expressed/detected among the selected genes in all the tissues with

ITGAM being the lowest expressed/detected in the tissue samples. Overall, we found that STRISH can detect CCCs more sensitively than ST-seq and scRNA-seq.

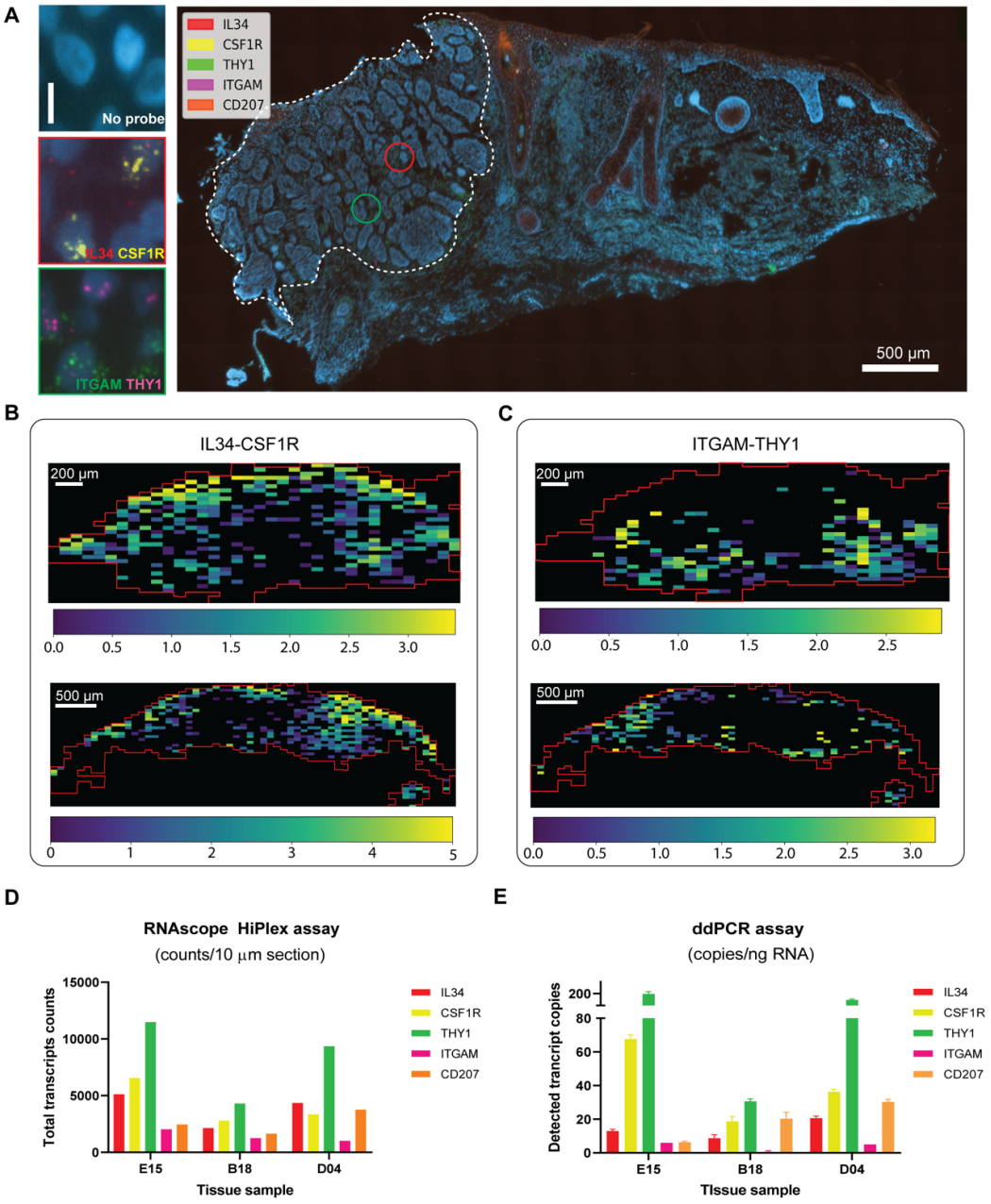

Figure 2. Detection of interactions at RNA levels in BCC patient skins by collective transcriptomic and genomic methods. (A) Target RNA molecule expression at a single cell level using RNAscope Hiplex assay defines cell to cell interaction by ligand - receptor expressing in a cancer nest area (patient ID - E15). Bar in the left panel = 10 µm. **(B)** and **(C)** Heatmaps of local co-expression in BCC patient samples, constructed by STRISH. Yellow colour indicates a high co-expression level of the pair. Samples for patients ID-B18 and ID-D04 are shown in the upper and the lower rows, respectively. The tissue histopathological annotations by H&E staining are in Supplemental Fig. S4. **(D)** and **(E)** The absolute number of the target mRNAs calculated from two assays.

Validation of ligand-receptor expression by using ddPCR

While RNAscope is expected to be able to detect single molecules in each cell, scanning through whole tissue sections with millions of cells may inadvertently lead to loss of detection due to missing spots arising from variations during slicing and preparation of tissue sections, especially for lowly expressed genes. Furthermore, it is unclear if RNAscope is appropriate for quantitative assessment of L-R interactions. ddPCR, on the other hand, is able to sensitively detect and quantify single molecules from tissues, and tissue sections, as long as there are several copies of the signal present, albeit without spatial information. To search for more evidence on the expression of ligands and receptors, we designed an automated ddPCR platform for quantitative analysis. A threshold for each targeted gene was drawn to distinguish the positive droplets from the negative droplets and the negative controls show no/less than 2 droplets detected above the threshold (**Supplementary Fig. S-6**). The transcript copy number per input RNA of each gene was presented in a bar plot (**Fig. 2-E**). The level of transcripts varies across the three BCC samples and the targeted genes. THY1 shows the highest mRNA expression in all the BCC tissues followed by CSF1R in tissues from patients E15 and D04, and CD207 in B18's tissue. ITGAM was expressed at the lowest level compared to the other genes in all samples whereas IL34 showed a moderate expression across the samples. We used Pearson's correlations to compare the quantitative analysis using RNAscope and ddPCR (**Fig. 2-D,E**) and found that overall the two assays are highly correlated, with correlation values between the two assays at 0.95, 0.89 and 0.94 for patient ID-E15, B18 and D04, respectively. In conjunction with RNA-ISH assay, our results suggest that the ddPCR confirmed the quantitativeness of using RNAscope for measuring L-R expression.

Discussion

In this study, we presented a sensitive, high-resolution, and genome-wide approach to discover and validate cell-cell communication across the whole spatial landscape of a tumor tissue section. Specifically, we combined spatial transcriptomics and scRNA-seq with RNA-ISH and ddPCR to study CCC in two non-melanoma skin cancers. This allowed us to quantitatively and visually assess the expression of target transcripts while maintaining the physiological spatial information in undissociated tissue sections. We demonstrated the novel approach to study the cell-to-cell interaction apart from scRNA-seq with ST-based method and refined the target-specific signal with RNAscope. The communication between detected cells with RNAscope technology was

additionally validated with ddPCR. We observed the improvement in smoothing effect of the gene expression differences between neighboring spots which facilitated denoising technical variation from spatial sequencing. Although the combination of ST-seq and scRNA-seq does provide insight into the RNA expression of the cells, the variation in the number of genes per spot poses the challenge to find the hotspot within a tissue that likely to have high cell interaction in the tissue.

In this study, we investigated the local co-expression of IL34-CSF1R and THY1-ITGAM using RNA-ISH multiplex assay to understand spatial context in disease. The signaling interaction between CSF1R and CSF1/IL34 is well known to regulate macrophage differentiation. IL34, CSF1 and their receptor co-expressed within the immune cell infiltrated area and regulated the different downstream signaling pathways in breast cancer (Zins et al. 2018). While the expression of both ligands is related to multiple tumors, IL34 is specifically expressed in the skin and central nervous system whereas CSF1 expression is more systemic (Baghdadi et al. 2018; Sasmono et al. 2003). CSF1R is known to be expressed on CD1a/CD207 Langerhans cells in human epidermis and stratified epithelial (Lonardi et al. 2020), where it can be found in high abundance in the basal and squamous layers of the epidermis, and is involved in anti-tumoral immune responses (Pogorzelska-Dyrbuś and Szepietowski 2020). In skin, differentiation and long-term maintenance of LCs *in situ* requires CSF1R signaling via IL34 (Ginhoux et al. 2006; Wang et al. 2012). We observed that the co-expression of IL34-CSF1R L-R pair is high in/near cancer nest area in all the patient samples analyzed. While IL34's role is canonically regarded as a niche factor for Langerhans cells in skin, increased expression has been implicated with poorer prognosis in non-small lung carcinoma due to its action on regulatory T cells (Baghdadi et al. 2017) where it is induced in cancer cells under chemotherapeutic condition, hinting its role as a biomarker that correlates with malignancy in some diseases (Baghdadi et al. 2016). In the context of skin cancer, it remains unknown if Langerhans cells and regulatory T cells are involved with tumor control or clearance and it is not known if they are associated with any prognostic significance.

THY1 is one of the cell surface markers for T-lymphocytes and mesenchymal stromal cells (Uder et al. 2018). The interaction of THY1 and ITGAM, an integrin molecule marking myeloid cell populations such as monocytes, macrophages and polymorphonuclear granulocytes, is shown to be involved in leukocyte migration in injured or inflammatory tissue, leading to an initiation of the function of leukocytes in host defenses (Wetzel et al. 2004). Furthermore, THY1 was determined to be responsible for tumor suppressive function in nasopharyngeal carcinoma and ovarian cancer (Lung et al. 2005; Abeysinghe et al. 2004) which suggested to be a putative candidate marker to study cell-cell interaction in cancerous diseases. While the co-expression of ITGAM-

THY1 is not as clear as the one of IL34-CSF1R with respect to tissue morphology related to cancerous areas, it is correlated with regions of infiltrated immune cells rather than cancer nests. This is concordant with the literature, suggesting that this L-R pair may be involved in migration of leukocytes to damaged cells and initiate host defense.

In our study, the expression of the selected molecules in skin cancers are supported by RNAscope Hiplex assay, demonstrating the advantages of the assay. RNAscope detection is more suitable than IHC in precancerous dysplastic nodules, helping early monitoring and diagnosis for patients with high risk diseases (Bakheet et al. 2020). Additionally, the absolute quantification of the target genes from ddPCR analysis supports the possibility of quantification using RNAscope assay. The tissue section for the ddPCR assay was sequentially collected after the section for RNAscope, assuring to keep an identical morphology between the two sections. While RNAscope assay only quantifies from a microscopic image that captures the fluorescent transcript on a cell that is stained with DAPI, ddPCR provides the total target transcript number in the input RNA sample. The direct comparison between the two assays may not be ideal but the ddPCR assay successively assists the result from the RNAscope analysis and reduces quantitative variability that RNAscope may carry. The combination of RNAscope and ddPCR can be considered as a powerful approach for target gene expression analysis when spatial information is also required.

Here, we propose a complete pipeline from bench work to bioinformatic analysis to study L-R in cancer. Our novel STRISH method scans through the whole tissue section to detect local co-expression of L-R pairs, rather than colocalization in traditional methods used in analysis of immunofluorescence images. STRISH, therefore, allows us to model L-R autocrine and paracrine interaction landscape across the whole tissue section. The workflow demonstrates the feasibility to discover new L-R pairs by genome-wide approaches, which are less sensitive but cover all genes, followed by targeted validation by high-resolution, sensitive RNAScope imaging and ddPCR. The pipeline generates the evidence for cell-to-cell interaction in tumors using collective methodologies on a same cancer patient specimen. The combination of different technologies overcomes the limitations of any individual method. Our findings on the mapping of more than one pair of ligand-receptor interactions in cancer tissues may shed a light on the development of a promising combination immunotherapy approaches.

Methods

Collection of human samples

The six human biopsies collected were determined to be BCC and SCC by histopathology. The shaved biopsies were each sampled from different patients who were presented at the Dermatology Department at the Princess Alexandra Hospital in 2019. The shaved biopsies were collected from various sun-exposed areas such as calf, cheek, ankle, and back. The patients were advised of the research protocol and written informed consent was obtained prior to the collection of the samples. Samples were de-identified upon receipt before experiments were allowed to proceed. The study was approved by The University of Queensland Human Research Ethics Committee (HREC-11-QPAH-477, The University of Queensland clearance no. 2012000052). The patient IDs and the analysis performed in this study are shown in **Fig. 1-A** and **Supplemental Table S1**.

Tissue preparation and cryosectioning

Human biopsy samples were snap-frozen in liquid nitrogen and transferred to the -80℃ biobank freezer for long-term storage. To prepare tissues for cryosectioning, fresh-frozen tissue samples were multiplexed and embedded in ice-cold optimal cutting temperature (OCT) compound (Tissue-Tek, Sakura Finetek U.S.A) in a standard cryomold (Tissue-Tek, Sakura Finetek U.S.A) and frozen on powdered dry ice. After solidification, the OCT embedded tissue block was stored either at -80℃ freezer until cryosectioning or used immediately for cryosectioning. Block was equilibrated in cryostat (Leica CM3050) for half an hour, with cryostat chamber temperature adjusted to -18 to -20℃ and specimen temperature set to -16℃. Embedded tissues were cryosectioned using R35 low-profile blades (Feather®, Japan) at 10µm thickness and transferred onto SuperFrost® Ultra Plus slides (Thermo Scientific, U.S.A.) for RNAscope Hiplex assay or Visium slides (10X Genomics, U.S.A.). Slides were kept in chilled slide mailers on dry ice and transferred to the -80℃ freezer in a sealed box prior to analysis.

Multiplexed RNA *in-situ* hybridization of cancer tissue sections

The following target probes to detect ligand-receptor interaction were designed by ACD probe design Team and used for the RNAscope HiPlex assay (ACD Cat. No. 324110): THY1 (ADV430611T2), IL34 (ADV313011T3), CSF1R (ADV310811T4), CD207 (ADV809521T7), and

ITGAM (ADV555091T8). The assay was performed as described in the manufacturer's user manual (ACD, 324100-UM). Briefly, a 10μm thickness tissue slide sectioned from the OCT embedded BCC or SCC tissue block was used for the assay and a consecutive section was made for a negative control. The (frozen) sections were fixed with freshly made 4% PFA for an hour followed by a dehydration process in ethanol and then were digested with protease IV for 30 minutes at room temperature. The slide was stained with a mixture of the 5 probes to allow them to hybridize with RNAs. The negative control slide was stained with RNAscope HiPlex 12 Negative control Probe that is provided in the kit. Consequently, a specific signal was amplified with high efficiency using RNAscope HiPlex Amp 1–3 reagents. After a couple of wash with a provided washing buffer, the sections were then stained with RNAscope HiPlex Fluor T1–T4 reagent and were counterstained with DAPI followed by mounting with a ProLong Gold Antifade Mountant (Fisher Scientific).

The images were captured by Axio Z1 slide scanner (Zeiss) with an appropriate adjustment of each fluorescent intensity. The first round of images was performed using 4 filters including DAPI for nuclei, Cy3 for THY1, Cy5 for IL34, and Cy7 for CSF1R. For the high resolution of an image, a 40x objective was used and the Z-stack interval was set up to 1.5μm resulting in 9 of Z-slices for each slide. Completing the first round of image, the fluorophores on the slide were cleaved for the second round of imaging process. The sections were stained with RNAscope Fluoro T5 – T8 reagent and images were captured using 3 filters including DAPI for nuclei, Cy5 for CD207 and Cy7 for ITGAM. The parameters for the microscope were set up the same as the first round. The images were further processed by ZEN software (version 3.2) for manual stitching and adjusting contrast/brightness.

RNAscope cell-cell interaction analysis

In order to determine the interaction of cell L-R across the whole tissue, we performed the positive cell detection using QuPath software (Bankhead et al. 2017) on the RNAscope stitched image produced by the Zen software. We developed STRISH to scan through the images and find the target region in the window strategy.

First we applied cell detection in the broad area with width and height dimensions were set to 10% of the whole scan image (**Supplemental Fig. S6-A,B**). With each window that cells are found, we subsequently narrowed down the searching area to a smaller window perimeter of 2.5% and 1.25% of the image until the number of cells found in each window are less than 10 cells (**Supplemental Fig. S6-C**). Otherwise the window was discarded and removed. From the

result of cell detection, we scanned for the cells within the window which express either markers of our scope, IL34 and CSF1R or THY1 and ITGAM by running another positive cell detection. We constrained the measure of the cells positive with each marker including IL34, CSF1R, ITGAM or THY1 to the number of cells with the same window to make all the windows comparable. The STRISH local co-expression level of each pair was calculated by normalizing the total count of each marker in the target pair in a window with the count of cells detected within the same window area.

For the interaction of ITGAM and THY1, as the signal of the genes were captured in separated imaging rounds with individual stitching in post process, some variants are introduced. To overcome the unaligned tissue layout, we performed image registration to map one image to the other (**Supplemental Fig. S6-D**). The image registration was performed using SITK library (Yaniv et al. 2018; Lowekamp et al. 2013). Upon the registration, the same analysis process was applied to detect the CCCs in one image and finally generate the heatmap of local co-expression is generated (i.e. **Supplemental Fig. S6-E**). In addition to image registration, we drew tissue contours around the heatmap and crop the non-tissue padding using skimage package (van der Walt et al. 2014) to remove unnecessary outliers that are caused by autofluorescence outside of the tissue.

Our code for detecting cells local coexpression and heatmap generation with STRISH is publicly available on github https://github.com/BiomedicalMachineLearning/SkinSpatial.

Visium spatial transcriptomics sequencing

Tissue sections cryosectioned at 10µm thickness were transferred to chilled Visium Tissue Optimization Slides (3000394, 10x Genomics, USA) and Visium Spatial Gene Expression Slides (2000233, 10x Genomics, USA), and allowed to adhere by warming the back of the slide. Tissue sections were dried for one min at 37℃, fixed in chilled 100% methanol for 30 minutes and stained with haematoxylin and eosin for 5 minutes and 2 minutes as per Visium Spatial Tissue Optimization User Guide (CG000238 Rev A, 10x Genomics) or Visium Spatial Gene Expression User Guide (CG000239 Rev A, 10x Genomics). Brightfield histology images were captured using a 10x objective on an Axio Z1 slide scanner (Zeiss). Brightfield images were exported as high-resolution tiff files using Zen software. This H&E staining and imaging protocol was used to stain all skin sections for histopathological annotation in this study.

For tissue optimization experiments, fluorescent images were taken using DMi8 inverted microscope (Leica Microsystems) at 10x objective with a TRITC filter at 7% laser power and

1000ms exposure time. Fluorescent images were stitched and exported as tiff files using the LasX program. For gene expression samples, tissues were permeabilized for 15 minutes, as optimally determined using the Tissue Optimization Kit. Gene expression libraries were prepared according to the Visium Spatial Gene Expression User Guide (CG000239, 10x Genomics) with slight modification to fragmentation reaction and SPRI bead ratios in response to reduced average fragment length of full-length cDNA. Libraries were loaded at 2pM and sequenced using a 150-cycle High output reagent kit (Illumina), on a NextSeq500 instrument house at the IMB Sequencing Facility. Sequencing was performed using the following protocol: Read1 - 28bp, Index1 - 10bp, Index2 - 10bp, Read2 - 120bp.

Visium sequencing and analysis

The Visium raw sequencing data in BCL format was converted to 110,782,035 FASTQ reads using bcl2fastq/2.17. The reads were trimmed by cutadapt/1.8.3 to remove sequences from poly-A tails and template-switching-oligos. The H&E tissue image was resized to. We used SpaceRanger V1.0 to map FASTQ reads to the cellRanger human reference genome and gene annotation for GRCh38-3.0.0. On average, for each spot we mapped 93,963 reads and detected 1,205 genes. The count matrix of the Visium data was preprocessed to remove genes that expressed in less than three cells, followed by the normalization, log transformation and scaling. For L-R analysis with CellPhoneDB, we followed the default pipeline from Efremova et al. 2020 and used a curated database v2.0.0. For spot clustering, we performed gene expression normalization using spatial morphological information from the H&E image using our stLearn software (Pham et al.). The normalization reduced the technical limitation in detecting lowly expressed genes. Secondly, a neighborhood graph of spots was built based on a reduced dimensional space, followed by the application of Louvain community detection to group similar spots into clusters. The plot for cell-cell interaction of a pair IL34 and CSF1R analysis is produced by the stLearn package (Pham et al.).

Single molecule droplet digital PCR

Frozen scrolls were adjacently sectioned (110-120 µm total thickness) from the same OCT block that was used for Visium and RNAscope assays. Individual BCC tissues from the block (3 in total) were isolated into separate tubes and snap-frozen with dry ice. Total RNA was extracted using RNeasy MinElute Cleanup kit (Qiagen) according to the manufacturer's instructions. RNA integrity was determined by Agilent RNA 6000 Pico kit and concentration measured by Qubit

(Thermo Fisher). cDNA was synthesized using Superscript™ IV VILO™ master mix with ezDNase™ enzyme (Invitrogen). In parallel, a No-RT control using equal RNA input was also generated to confirm lack of gDNA contamination.

The ddPCR was carried out on the QX200 platform (Bio-Rad) according to the manufacturer's instructions. Each triplicate reaction contained 1x ddPCR SuperMix for Probes no dUTP (Bio-rad), 1x target primer/probe mix conjugated with FAM or HEX (PrimePCR assay, Bio-Rad), cDNA and dH_2O. The controls consisted of reaction mixture containing dH_2O instead of cDNA or No-RT template from cDNA synthesis. Greater than 10,000 droplets were generated in each well by an automated droplet generator (range = 10,381 – 19,788). Subsequently, PCR amplification was performed in a C1000 Touch Thermal Cycler using an optimized program. The reaction was run at 95℃ for 10 minutes, 40 cycles of 94℃ for 30 seconds, 57.5℃ for 30 seconds and a final incubation at 98℃ for 10 minutes. Results from the amplification were read using a QX200 Droplet Reader followed by data analysis with the QuantaSoft analysis software. Context sequences for the target probes are provided in **Supplemental Table S2**. The absolute transcript number for each target gene was determined by the software after manually setting the threshold for defining positive droplets. The mean of each triplicate was then calculated to give detected transcripts per microliter, from which values for transcript copies/ng RNA input were calculated.

Single cell RNA sequencing and analysis

Fresh shaved suspected SCC and BCC lesions and a 4mm punch biopsy of non-sun exposed skin from the same patient were collected in DMEM for immediate tissue dissociation. The tissue sections were rinsed in PBS and incubated in 8 mg/mL Dispase at 37°C for 1h, and were minced before a 3 minutes incubation in 0.25% Trypsin at 37°C. To collect single cells, the suspension was filtered through a 70µm cell strainer. Cells were collected in PBS containing Fetal Calf Serum. The 10x Genomics Chromium scRNA-sequencing followed the manufacturer’s instructions, using the Single Cell 3' Library, Gel Bead and Multiplex Kit (version 2, PN-120233; 10x Genomics). Cell numbers in each reaction were optimized to capture approximately 3,000 cells. The single-cell transcriptome libraries were sequenced on an Illumina NextSeq500, using a 150-cycle High Output reagent kit (NextSeq500/550 version 2, FC-404-2002; Illumina) as follows: 98 bp (read 2), 8 bp (I7 index), and 26 bp (read 1 - cell barcodes and UMI).

The BCL file was converted to a FASTQ file using bcl2fastq/2.17. We used CellRanger/3.0.2 for mapping to Homo_sapiens.GRCh38p10 reference. Using Seurat/3.2.0 in R/3.6.3, we removed cells with fewer than 200 or more than 5000 genes and cells with over 20%

of all reads mapped to mitochondrial genes. The processed expression matrix was scaled to 10,000 reads/cell, log normalized, and only the top 2500 most variable genes were kept for PCA dimensionality reduction. The top 50 PCs were used for building a nearest neighbor graph, followed by Louvain graph-based clustering in Seurat at the resolution of 0.5.

Using scRNA-seq data to infer L-R interaction that may differentiate normal and cancer tissues, we applied NicheNet ligand-receptor prediction. We ranked the list of potential ligand activity by descending Pearson values and pulled out the top 20 ligand activities. Using top 20 upstream ligands, we filtered the prebuilt ligand receptor network to find the corresponding upstream receptor and plotted the heatmap of potential interaction in **Supplemental Fig. S1-A**.

Acknowledgements

We would like to thank the Dermatology Department of the Princess Alexandra Hospital for their ongoing efforts and support to recruit patients and collect tissue for our research. In particular, we thank A/Prof. Helmut Schaider and Christine Shen. We thank the histopathology facility at the Translational Research Institute, the microscopy facility at the School of Biomedical Sciences, and the sequencing facilities at the Institute for Molecular Bioscience. This work has been supported by the Australian Research Council (ARC DECRA DE190100116), the University of Queensland, and the Genome Innovation Hub. We especially thank the patients for donating their tissue for our research.

Disclosure Declaration

The authors declare no competing interests.

References

Abeysinghe HR, Pollock SJ, Guckert NL, Veyberman Y, Keng P, Halterman M, Federoff HJ, Rosenblatt JP, Wang N. 2004. The role of the THY1 gene in human ovarian cancer suppression based on transfection studies. *Cancer Genet Cytogenet* **149**: 1–10.

Baghdadi M, Endo H, Takano A, Ishikawa K, Kameda Y, Wada H, Miyagi Y, Yokose T, Ito H, Nakayama H, et al. 2018. High co-expression of IL-34 and M-CSF correlates with tumor progression and poor survival in lung cancers. *Sci Rep* **8**: 418.

Baghdadi M, Endo H, Tanaka Y, Wada H, Seino K-I. 2017. Interleukin 34, from pathogenesis to clinical applications. *Cytokine* **99**: 139–147.

Baghdadi M, Wada H, Nakanishi S, Abe H, Han N, Putra WE, Endo D, Watari H, Sakuragi N, Hida Y, et al. 2016. Chemotherapy-Induced IL34 Enhances Immunosuppression by Tumor-Associated Macrophages and Mediates Survival of Chemoresistant Lung Cancer Cells. *Cancer Res* **76**: 6030–6042.

Bakheet AMH, Zhao C, Chen J-N, Zhang J-Y, Huang J-T, Du Y, Gong L-P, Bi Y-H, Shao C-K. 2020. Improving pathological early diagnosis and differential biomarker value for hepatocellular carcinoma via RNAscope technology. *Hepatol Int* **14**: 96–104.

Bankhead P, Loughrey MB, Fernández JA, Dombrowski Y, McArt DG, Dunne PD, McQuaid S, Gray RT, Murray LJ, Coleman HG, et al. 2017. QuPath: Open source software for digital pathology image analysis. *Sci Rep* **7**: 16878.

Berglund E, Maaskola J, Schultz N, Friedrich S, Marklund M, Bergenstråhle J, Tarish F, Tanoglidi A, Vickovic S, Larsson L, et al. 2018. Spatial maps of prostate cancer transcriptomes reveal an unexplored landscape of heterogeneity. *Nat Commun* **9**: 2419.

Browaeys R, Saelens W, Saeys Y. 2020. NicheNet: modeling intercellular communication by linking ligands to target genes. *Nat Methods* **17**: 159–162.

Brücher BLDM, Björn L D, Jamall IS. 2014. Cell-Cell Communication in the Tumor Microenvironment, Carcinogenesis, and Anticancer Treatment. *Cellular Physiology and Biochemistry* **34**: 213–243. http://dx.doi.org/10.1159/000362978.

Efremova M, Vento-Tormo M, Teichmann SA, Vento-Tormo R. 2020. CellPhoneDB: inferring cell-cell communication from combined expression of multi-subunit ligand-receptor complexes. *Nat Protoc* **15**: 1484–1506.

Ginhoux F, Tacke F, Angeli V, Bogunovic M, Loubeau M, Dai X-M, Stanley ER, Randolph GJ, Merad M. 2006. Langerhans cells arise from monocytes in vivo. *Nat Immunol* **7**: 265–273.

Guillonneau C, Bézie S, Anegon I. 2017. Immunoregulatory properties of the cytokine IL-34. *Cell Mol Life Sci* **74**: 2569–2586.

Ji AL, Rubin AJ, Thrane K, Jiang S, Reynolds DL, Meyers RM, Guo MG, George BM, Mollbrink A, Bergenstråhle J, et al. 2020. Multimodal Analysis of Composition and Spatial Architecture in Human Squamous Cell Carcinoma. *Cell* **182**: 497–514.e22.

Kuypers J, Jerome KR. 2017. Applications of Digital PCR for Clinical Microbiology. *J Clin Microbiol* **55**: 1621–1628.

Li H, Bai R, Zhao Z, Tao L, Ma M, Ji Z, Jian M, Ding Z, Dai X, Bao F, et al. 2018. Application of droplet digital PCR to detect the pathogens of infectious diseases. *Biosci Rep* **38**. http://dx.doi.org/10.1042/BSR20181170.

Lonardi S, Scutera S, Licini S, Lorenzi L, Cesinaro AM, Gatta LB, Castagnoli C, Bollero D, Sparti R, Tomaselli M, et al. 2020. CSF1R Is Required for Differentiation and Migration of Langerhans Cells and Langerhans Cell Histiocytosis. *Cancer Immunol Res* **8**: 829–841.

Lowekamp BC, Chen DT, Ibáñez L, Blezek D. 2013. The Design of SimpleITK. *Front*

Neuroinform **7**: 45.

Lung HL, Bangarusamy DK, Xie D, Cheung AKL, Cheng Y, Kumaran MK, Miller L, Liu ET-B, Guan X-Y, Sham JS, et al. 2005. THY1 is a candidate tumour suppressor gene with decreased expression in metastatic nasopharyngeal carcinoma. *Oncogene* **24**: 6525–6532.

Moncada R, Wagner F, Chiodin M, Devlin JC, Baron M, Hajdu CH, Simeone DM, Yanai I. Integrating single-cell RNA-Seq with spatial transcriptomics in pancreatic ductal adenocarcinoma using multimodal intersection analysis. http://dx.doi.org/10.1101/254375.

Ott PA, Hodi FS, Kaufman HL, Wigginton JM, Wolchok JD. 2017. Combination immunotherapy: a road map. *J Immunother Cancer* **5**: 16.

Pardoll DM. 2012. The blockade of immune checkpoints in cancer immunotherapy. *Nature Reviews Cancer* **12**: 252–264. http://dx.doi.org/10.1038/nrc3239.

Pham D, Tan X, Xu J, Grice LF, Lam PY, Raghubar A, Vukovic J, Ruitenberg MJ, Nguyen Q. stLearn: integrating spatial location, tissue morphology and gene expression to find cell types, cell-cell interactions and spatial trajectories within undissociated tissues. http://dx.doi.org/10.1101/2020.05.31.125658.

Pogorzelska-Dyrbuś J, Szepietowski JC. 2020. Density of Langerhans Cells in Nonmelanoma Skin Cancers: A Systematic Review. *Mediators Inflamm* **2020**: 8745863.

Salmén F, Ståhl PL, Mollbrink A, Navarro JF, Vickovic S, Frisén J, Lundeberg J. 2018. Barcoded solid-phase RNA capture for Spatial Transcriptomics profiling in mammalian tissue sections. *Nat Protoc* **13**: 2501–2534.

Sasmono RT, Oceandy D, Pollard JW, Tong W, Pavli P, Wainwright BJ, Ostrowski MC, Himes SR, Hume DA. 2003. A macrophage colony-stimulating factor receptor-green fluorescent protein transgene is expressed throughout the mononuclear phagocyte system of the mouse. *Blood* **101**: 1155–1163.

Thrane K, Eriksson H, Maaskola J, Hansson J, Lundeberg J. 2018. Spatially Resolved Transcriptomics Enables Dissection of Genetic Heterogeneity in Stage III Cutaneous Malignant Melanoma. *Cancer Res* **78**: 5970–5979.

Uder C, Brückner S, Winkler S, Tautenhahn H-M, Christ B. 2018. Mammalian MSC from selected species: Features and applications. *Cytometry Part A* **93**: 32–49. http://dx.doi.org/10.1002/cyto.a.23239.

van der Walt S, Schönberger JL, Nunez-Iglesias J, Boulogne F, Warner JD, Yager N, Gouillart E, Yu T, scikit-image contributors. 2014. scikit-image: image processing in Python. *PeerJ* **2**: e453.

Wang Y, Szretter KJ, Vermi W, Gilfillan S, Rossini C, Cella M, Barrow AD, Diamond MS, Colonna M. 2012. IL-34 is a tissue-restricted ligand of CSF1R required for the development of Langerhans cells and microglia. *Nat Immunol* **13**: 753–760.

Wetzel A, Chavakis T, Preissner KT, Sticherling M, Haustein U-F, Anderegg U, Saalbach A. 2004. Human Thy-1 (CD90) on Activated Endothelial Cells Is a Counterreceptor for the Leukocyte Integrin Mac-1 (CD11b/CD18). *The Journal of Immunology* **172**: 3850–3859.

http://dx.doi.org/10.4049/jimmunol.172.6.3850.

Yaniv Z, Lowekamp BC, Johnson HJ, Beare R. 2018. SimpleITK Image-Analysis Notebooks: a Collaborative Environment for Education and Reproducible Research. *J Digit Imaging* **31**: 290–303.

Zins K, Heller G, Mayerhofer M, Schreiber M, Abraham D. 2018. Differential prognostic impact of interleukin-34 mRNA expression and infiltrating immune cell composition in intrinsic breast cancer subtypes. *Oncotarget* **9**: 23126–23148.