

出國報告（出國類別：進修）

2024 至美國德州聖安東尼奧健康科學中心 Greehey  
兒童癌症研究所學習空間基因體學研究

服務機關：臺中榮民總醫院醫學研究部

姓名職稱：康于庭高級助理研究員

派赴國家/地區：美國/聖安東尼奧

出國期間：113 年 8 月 1 日至 113 年 9 月 1 日

報告日期：113 年 10 月 1 日

## 摘要

此次出國進修至美國德州聖安東尼奧的健康科學中心 Greehey 兒童癌症研究所，學習空間基因體學技術，深入學習最新的 Visium HD 和 Xenium 空間轉錄組學儀器操作與資料分析技術。在陳一東教授團隊的指導下，我學習了如何透過 Space Ranger 和 Loupe Browser 進行解序，並將基因數據與組織影像進行整合與可視化分析。本次進修處理了來自本院的兩組乳腺癌樣本（包括三陰性乳癌及其肺轉移檢體，以及罕見性黏液囊狀乳腺癌），並使用 Seurat 進行資料前處理及後續整合分析。進修期間多次與本院病理科醫師及陳教授團隊討論，應用多種工具進行細胞類型的交叉比對與註釋，進一步提升了對空間基因體學分析的理解與應用。此外，我們也持續討論使用 StarDist、QuPath 和 Cellpose 等訓練模型進行影像分析，特別是生成基於 Visium HD 的  $2 \times 2 \mu\text{m}$  細胞核基因數據，針對不同細胞類型進行自動化的註釋與分割。此次進修不僅提升了我在空間基因體學領域的技術，也期望未來可建立簡易且快速的分析流程使空間基因體學技術可廣泛應用於臨床分析中。

**關鍵字：**空間基因體、乳癌、深度學習

## 目 次

|               |   |
|---------------|---|
| 一、 目的 .....   | 1 |
| 二、 過程 .....   | 1 |
| 三、 心得 .....   | 3 |
| 四、 建議事項 ..... | 3 |
| 五、 附錄 .....   | 4 |

## 一、 目的

在現今轉錄體學研究當中，單細胞 RNA 定序技術提供了解決許多與腫瘤以及免疫相關疾病的研究方法，透過不同細胞類型之間的基因差異與特徵，將其分類為更細緻的亞型或是可發現新的組織特定型細胞，以便釐清特定細胞在疾病發展中的貢獻與影響。雖然單細胞 RNA 定序技術可以提供細胞轉錄組的高通量分析，但卻會遺失細胞在組織上的定位以及其鄰近細胞之間交互作用的資訊。空間基因體是目前最新發展出可釐清單細胞在組織空間分布、基因表現、細胞間質分布的重要技術。傳統在空間分布上使用免疫組織化學染色或原位雜交技術進行確認，但通常需要確定目標且通量較低，亦無法同時分析不同類型細胞表達，因此研究效益不高。

本次進修至美國德州聖安東尼奧的健康科學中心 Greehey 兒童癌症研究所學習空間基因體(Spatial transcriptome)研究技術，該單位目前備有最新的空間基因體分析儀器-全自動原位影像基因分析儀 Xenium 與高解析度空間基因體分析儀器-Visium HD，研究室主持人陳一東教授也具有豐富的單細胞與空間基因體研究經驗，可助於我們快速掌握研究分析流程與技巧，以期未來能將此技術運用於更深入的研究分析更甚至使用於臨床診斷中。

## 二、 過程

### (一)學習空間基因體定序資料解序

#### 1. Space Ranger 解序軟體使用

Space Ranger 是 10X Genomics 為空間轉錄組學實驗專門開發的數據處理軟體。它的主要功能是將樣本中的基因表達數據精確地映射到樣本的空間位置信息中，幫助計算基因表達與組織結構之間的關係。利用內建的 image-processing 模組將病理 H&E 影像與 CytAssist 影像進行對齊合併，再與定序數據之 FASTQ 檔整合生成出精確的空間基因表達數據。這些數據會以多種不同解析度進行輸出，包括 2  $\mu$ m、8  $\mu$ m 和 16  $\mu$ m 的解析度，以滿足後續分析需求。

#### 2. Loupe Browser 可視化與影像整合軟體使用

Loupe Browser 是一款可單機操作的互動式軟體，適用於單細胞轉錄組學和空間轉錄組學的數據分析。它能夠執行細胞聚類、細胞類型鑑定以及基因表現的基本分析，並提供簡便的下游數據探索功能。透過 Loupe Browser，使用者可以直觀地進行數據篩選和可視化，適合快速且簡易地分析不過於複雜的樣本。未來，我們可以為較不複雜的樣本建立一套標準的單機版分析流程，便於想自行操作或學習的單位使用。此流程可以作為教學模板，幫助這些單位掌握空間基因組學的基本分析步驟。這將有助於推廣空間基因組學技術，或應用於臨床端進行自主分析和研究。

### (二)空間基因體資料分析

### 1. 初步資料分析

我們於至美國前已完成了二組 Visium HD 實驗，並獲得定序資料包含:1.三陰性乳癌(triple-negative breast cancer,TNBC) 與其肺轉移(Lung metastasis)配對，以及 2.罕見性黏液囊狀乳腺癌(Mucinous cystadenocarcinoma)，分別與本院乳房外科與病檢部合作之計劃。我們依照 Seurat v5.1.0 進行註釋細胞類型之前的初步分析工作流程包含: Setup Object、Quality control、Normalizing、feature selection、dimensional reduction、Clustering、Find differentially expressed features 等進行資料初步整理。

### 2. 細胞註釋

在細胞註釋過程中為空間基因體研究最重要的步驟之一，利用各種方式交叉比對與預測，進行細胞類型註釋，我們分別與本院病理科醫師進行細胞型態討論，以及與陳教授團隊討論使用多種細胞註釋功能包含:1.使用 published gene list 進行基因比對。2.使用 Online functions: CellMarker, SingleR 及 PanglaoDB 註釋。3.使用 ChatGPT 大型生成式 AI 進行細胞類型預測等方式完成註釋。

### 3. 下游分析

後續將進行的下游分析包含:Unsupervised clustering、Identification of spatial tissue domains、Subsetting spatial regions、Integration with scRNA-seq data、Comparing the spatial localization of different cell types 等，我們會持續與陳一東教授合作完成此次計畫。

## (三)利用影像進行模型訓練

### 1. StarDist 細胞核分割方法(Nuclei Segmentation)

Visium HD 是基於  $2 \times 2 \mu\text{m}$  正方形的網格(bins)為單位建立出  $8 \times 8 \mu\text{m}$  和  $16 \times 16 \mu\text{m}$  的基因表現資料檔，由於普遍細胞直徑大小範圍落在  $10 \sim 30 \mu\text{m}$  一般我們將會以  $8 \times 8 \mu\text{m}$  資料檔進行後續分析。然而，許多免疫細胞的大小可能更小，因此我們需要提取  $2 \times 2 \mu\text{m}$  資料檔，以獲得更高的解析度。為了更精確地表達細胞的不規則多邊形狀態，我們與陳教授團隊合作討論，如何應用 StarDist，這是一款基於 Python 的深度學習工具，利用卷積神經網絡 (CNN) 進行影像處理，專門用於利用顯微鏡影像進行細胞核分割分析。透過 StarDist，我們期望能將  $2 \times 2 \mu\text{m}$  網格(bins)重新生成精細的細胞核基因表達資料檔，以進行更深入的後續分析。

### 2. QuPath 與 Cellpose 細胞核分割及機器學習模型訓練:

陳教授團隊擅長影像分析，因此也使用 QuPath 和 Cellpose 工具進行精確的細胞核分割，通過將 Cellpose 的分割結果匯入 QuPath，我們可以進一步進行細胞分類和機器學習模型訓練，針對不同類型的細胞進行自動化分類。

### 三、心得

美國德州聖安東尼奧的健康科學中心是德州知名的醫學與生物科學研究機構之一，擁有多個頂尖研究中心，其中包括我此次進修的 Greehey 兒童癌症研究所，以及 Texas Biomedical Research Institute 和 Barshop Institute 等。這些研究中心涵蓋了從兒童到老年化的癌症、基因體學、神經科學及其他重大疾病的多領域研究。其中 Greehey 兒童癌症研究所內設有 Genome Sequencing Facility（基因組定序設施），該設施類似於我們的精準醫學中心，配備了多種先進的定序儀器，以及單細胞和空間基因體學相關的分析設備。這使得 Greehey 研究所不僅能進行臨床分析檢驗，亦擁有豐富的基礎研究能力，成為結合臨床與研究的研究中心。

此次到 Greehey 兒童癌症研究所進行進修，讓我深刻體會到美國在單細胞核空間基因體學研究的蓬勃發展，我觀察到陳教授的團隊在基因體學研究上的分工規畫、持續學習與進步，並且他們也積極發展各種演算法來應對不同的研究需求，團隊還制定了詳細、快速的分析流程，以便靈活應對各類研究計劃，這種高效、專業的工作方式非常值得我們借鑒。陳教授也提到目前國外也都面臨研究人力不足的問題，但若能活用深度學習，包含機器學習或是大型生程式語言模型，都可以加速資料的分析速度，以及解決複雜的資料運算。

在這次學習空間基因體分析的時間裡，我們也遇到了許多困難，包含發現帶到美國的資料品質並不好，導致後續資料分析的困難度提升，因此我們也與美國端 10X Genomics 的 engineer 進行討論以解決後續資料分析問題，原廠工程師也提供了我們許多意見，在未來進行空間基因體實驗時可以調整的部分，期望我們在後續的研究中會更加順利。

回國後，剛好遇到美國國衛院范揚政老師來指導，因此我向范老師報告了我的分析資料，並得到了寶貴的指導與建議。同時，謝部主任也邀請我在醫研部內進行進修心得的分享，分享完後許多同仁都對於空間基因體又有更多的了解及興趣，相信這次經驗不僅促進了內部的學術交流，也為未來跨科室合作、推動空間基因體技術的應用奠定了良好的基礎。

最後，此次進修行程是我初次訪美，十分感謝蕭自宏研究員推薦以及醫院給予機會，還有榮興基金會的資助。此次行程讓我獲得十分寶貴的經驗，不只在工作上，生活上亦是！院部也在我決定要出國進修之際，皆十分關心我的狀態以及準備情形，包含在出國前師培科持續確認我的準備情形、謝育整部主任對於我到美國學習過程的關心，還有陳一銘科主任在生活上的關懷，讓初次進修的我都覺得十分得安心。在美國期間也感謝陳一東教授團隊的指導以及照顧，未來我們也將持續合作，希望能在基因體學研究中獲得更多知識與技能，並且能將所學回饋於醫院，成為本院的研究助力。

### 四、建議事項

- (五) 持續發展空間基因體相關研究:空間基因體不只有基因體學，亦有蛋白質體學，但因設備較為昂貴需要院方支持才能持續進行下去，未來亦可與病檢部共同發展，使空間基因體學逐漸應到臨床端。
- (六) 強化病理專科人才: 空間基因體分析高度依賴病理科醫師的專業判斷與觀察，尤其在初期挑選檢體時，病理科醫師的協助對於檢體特性的準確判定至關重要，並將直

接影響後續研究的發展方向。然而，病理科醫師目前面臨業務繁重人力短缺的現實情況，因此希望院方能提供更多資源和支持，協助培養病理專科人才，使其能一同參與研究當中，可提升研究效率並促進跨領域合作。

- (七) 招募生物資訊分析人才: 為了進一步提升本院在基因體學領域的生物資訊分析能力和規模，本院目前在相關領域的人力資源相對有限。若要成功將空間基因體學技術應用於臨床，迫切需要招募更多具備研究經驗的專業生物資訊人才，這將有助於推動研究技術的臨床轉化。
- (八) 提升本院大型語言模型能量: 在當前的基因體學或遺傳學研究中，越來越多地依賴於語言模型（如大型生成式語言模型，LLMs）來處理和解釋大量生物數據。儘管醫院目前已開設了一些基礎的 AI 課程，但較偏向於基本理論與一般應用。希望院內可開設針對基因組學和遺傳學應用的 AI 進階課程，涵蓋例如大型語言模型（LLMs）、深度學習、機器學習在基因數據分析中的實際應用。

## 五、 附錄



美國德州聖安東尼奧的健康科學中心 Greehey 兒童癌症研究所外觀



Greehey 兒童癌症研究所內部空間



陳一東教授所屬研究室內部

