



統計科學研究所

INSTITUTE OF  
STATISTICAL SCIENCE



統計所學術演講



中研院統計所

## 學術演講

講題：Tukey計畫口頭報告「A data-driven approach to the property estimation and structural generation of molecules via deep learning」

講者：陳璿宇副研究員、袁新盛副研究員、  
楊振翔副研究員、謝叔蓉研究員  
(中央研究院統計科學研究所)

時間：2022年12月19日(星期一)，10:30-12:00

地點：統計所B1演講廳

### Abstract

在子計畫一中，以一組443例肺癌病人的基因表達微陣列數據，納入所有基因的基因表達進行影像轉換，以卷積神經網路方法，找出肺癌高風險病人基因表達圖譜。結果顯示，在不同模型條件設定下訓練組準確率可達100%，然而驗證組準確率為0.3%-0.76%。子計畫二與四研發了選變數的方法，對十種標靶及化療藥物的預測AUC比最先進的深度學習機器還好(六個藥的測試AUC 0.75~1.0)。對於不符合我們假設的資料集，我們導出了預測閾值，使得七個藥的預測準確度達到0.70~0.93。子計畫三完成了The Cancer Genome Atlas資料之整合分析，發現了癌症中基因體與轉錄體之間數以萬計的關聯，將其組織為階層式的結構，並在數百筆外界資料上驗證了這些關聯。此外，我們也開發了一種演算法，結合單細胞與大量RNA測序資料來評估檢體之轉錄體異質性。現有方法多以單細胞RNA測序資料為基準，估計大量RNA測序資料檢體中族群成分，我們則先從大量RNA測序資料估計族群成分，再於單細胞RNA測序資料中驗證。

※ 茶會：上午10:10開始。