

# 資料採礦技術應用於肺癌患者之基因篩選及生活因素探討

紀懿芳

輔仁大學統計資訊學系應用統計碩士班

## 摘要

肺癌為國人最常見的惡性疾病之一，根據我國行政院衛生署 2011 年的統計資料顯示，在 2011 年國人十大死因排行，惡性腫瘤(癌症)連續 30 年蟬連首位，不論在男性和女性，肺癌在癌症死亡原因排名都是第一名。根據世界衛生組織 WHO (World Health Organization)之統計，全世界於肺癌死亡率與發生率皆高居全球癌症排名第一。在全球，每年癌症死亡人數約為 760 萬人，其中肺癌所導致的死亡病例每年就有約 130 萬人。

故本研究利用美國國家衛生研究中心(NCBI)資料庫所提供的 Affymetrix GeneChip Human Genome U133 Plus 2.0 Array 這組晶片所做出來的 microarray 資料為材料來進行研究，所使用的資料集為 GSE2109、GSE7307、GSE19804、GSE22047 四個資料集。使用肺癌病患之特性資料，進行關聯規則方法，找出家族遺傳、抽菸、喝酒、性別、年齡層、腫瘤期數、淋巴轉移與腫瘤轉移之間的相關性，並應用資料採礦技術建立判別模型，最後再結合 GO 基因功能分類資料庫，將篩選出的基因進行功能分類，提供生物及醫學研究方面上之參考。

關鍵字：微陣列、肺癌、資料採礦