

# 問都給你問

## 1. 有使用過目前方法以外的方法嗎？

- 當初我們所使用的分群分法為 single linkage 階層式分群法，且在後續的研究時，也有使用過 K-means 作為分群方式，但最後我們選擇 complete linkage 作為我們最終的分群方法，因為 2.

## 2. 為什麼使用complete linkage?

- 相對於 single linkage 來說，使用 complete linkage 和 average linkage 會使得結果更有關聯性，能夠在對於分類時考慮得更加周到。我們也從 complete linkage 和 average linkage 中做得比較，我們發現 complete linkage 所分類出之結果可從中看出疾病之間合理的相關性，也發現了一些有趣的結果。

## 3. 為甚麼採用分群方式分類疾病？

- 使用分類的主要原因是為了瞭解各個疾病之間的共病關聯性，因此我們將各個疾病的患病情況視為特徵，透過計算疾病之間的correlatioZ得知其之間的相關性。利用由相關性轉換的距離矩陣去做階層式分群，階層式分群能夠根據疾病的特徵相似度將它們分為不同的群組。

## 4. 為甚麼是分五群？

- 最主要的原因是因為資料集的人數(?)，最一開始採用single linkage的階層式分群法時，分成了七群。因此換成現在的方式時，也從七群開始，視情況增減群數，最後分為五群能夠得到最佳的分群結果與人數。而若分為六或七群，其結果也只是以分為五群時的結果再使人數已經很少的群組分成更多微小的群組，這樣一來，這些人數過少的群組由於人數過少而無法為我們的分析提供實質的幫助。

## 5. 有做Kmeans，那為何不用？

- k-means分群法在初始化時需要指定結果的分類群數，但階層式不用，階層式可以依照結果來決定分群群數。而我們所想要的結果是在分群過

後，為各個群組有各自的主題或是高相關性，所以最後選擇階層是分群

## 6. 差異甲基化分析與內部控制基因位點中需要的低差異，是否矛盾，兩者的差異是否一樣

- 不一樣
- 差異甲基化分析的主要目的是比較不同條件或組別間特定基因位點的甲基化差異。通過統計分析，我們可以過濾出在統計上有顯著差異的基因位點。
- 而內部控制基因位點的特點是，無論在何種時間、空間或實驗方法下，它都應該展示出高度的一致性和穩定性。這意味著這些位點的甲基化狀態在各種條件下都是相對固定的，因此它們可以作為一個基準或參考。

## 7. 過濾後的DMPs已有顯著的差異性，為什麼還適合最為候選內部控制基因位點的參考

- 首先，我們遵循了標準的差異甲基化分析通用流程。這意味著我們的分析方法是根據已經公認且被廣泛應用的標準方法進行。這不僅確保了我們的結果的可靠性，而且意味著我們的結果可以與其他使用相同方法的研究進行比較。
- 我們在多次重覆的實驗中觀察到這些DMPs都展示了高度的穩定性。這種一致性為我們提供了信心，認為這些DMPs可以作為DNA甲基化實驗的內部控制。畢竟，內部控制的主要特性就是它們應該在不同的實驗或條件下保持相對的穩定性。
- 更重要的是，我們觀察到這些DMPs在每次實驗中的表達量都是相對穩定的。這確保了這些DMPs的甲基化水平不會因為實驗中的微小變化而受到太大的影響，從而提供了一個穩定的基線，用於比較其他可能的變異性或差異性。
- 結合這三點，我們認為都符合理想內部控制所需要的特性

## 8. 為什麼用Accuracy作為評判標準

- 在資料集當中，我們所挑選的疾病發生率幾乎是常態分佈，這意味著罹患疾病和未罹患疾病的分布相對均勻。在這樣的情境下，Accuracy是一個合理且有代表性的指標，因為模型的隨機猜測不太可能獲得很高的Accuracy。

- 如果他們問起為什麼沒有使用其他指標作為評判標準 (ex.

### Precision, Recall, F1 Score)

- 雖然Precision, Recall, 和 F1 Score能夠提供模型對於是否罹患疾病預測效能的更多細節，但我們認為Accuracy足以描述我們模型的整體效能，並能夠概括模型在所有情境下的預測能力。
- 我們希望結果報告能夠簡單、直觀地傳達給大眾，而Accuracy是一個容易理解的指標。它直接告訴我們模型正確預測的百分比，對於非專業背景的聽眾來說，這是一個很直觀的描述。