# Bioinformatics: 2017 Final Project

在一個 transcriptome 的研究中，我們發現肝癌細胞株中 A 基因 mRNA 有大量表現的情形。為了進一步研究 A 基因大量表現所代表的意義，請以 5 人為一組進行討論分析試著回答以下的問題：

1. 請找 A 基因在染色體上的位置。請簡單描述此基因的大小及結構（exon，intron，splicing site，等）。已知此基因有不同 RNA 產物，請比較此基因的不同 RNA 產物，找出 A 基因 mRNA 產生的原因？如果是 alternative splicing，請說明不同 RNA 產物 exon 的組成。如果是其他原因， 請說明。
2. 請在資料庫中搜尋此基因在不同組織中的表現情形？如何偵測此基因的表現狀況? *(Hint:利用 primer or probe design in RT-PCR or Northern blot，並試推測是否有 intronic read through…,etc )。*利用 GEO 資料庫，是否可以找出此基因可能與癌症或其他疾病相關的證據？有那些資料顯示其此基因表現量在何種情況下會發生變化？
3. 利用資料庫資料，分析此基因 RNA 的 UTR，並分析 UTR 來推測其是否具有可能的調控因子，此外並請找出是否有 microRNA 會 target 到該基因。 分析的此基因的 5'-flanking sequence（-1000bp）並推測其可能參與調控之轉錄因子 (promoter, transcription factor…)？ 已知 A 基因 mRNA 在正常肝細胞中不表現但在肝癌細胞株中卻大量表現，由 UTR 及 5'-flanking sequence 的分析，你可以解釋這個現象嗎？
4. 請分析 A 蛋白質的特性（IP、subcellular localization、functional domain、homolog proteins in other species…..）。
5. 已知 A 基因具有 guanine nucleotide exchange factor （GEF）的特徵，並可能參與在訊息傳遞路徑中。請利用資料庫或其他物種網路資源找出可能與 A 蛋白質相關的作用分子 (protein? DNA? RNA?....）。 並試著利用蛋白質交互作用及 pathway analysis 工具來預測 A 蛋白質參與的交互作用網路。 你可以預測 A 蛋白質大量表現對在肝癌細胞的影響嗎？

You will have approximately four weeks to finish your project. ***There is no restriction to the tools you use. You can apply any computational tools (online or desktop programs) to help you to complete your project. But you have to report the name and reference for the program you used in this project.*** After you have completed your project, upload the results to your web directory before 1/3/2018. There will be ~30 minutes project presentation /discussion for each group at the end of the semester. If you have any questions regarding this project, you are welcome to post your question on the forum, or ask any of the teaching assistants. Good luck and have fun!