題目

利用 RankNet 排序結果推論突變發生之先後關係

摘要

突變會造成等位基因頻率(allele frequency)的改變，等位基因頻率越高代表突變發生的時間越早。RankNet 是利用機器學習的方式，來解決排序上的問題，將兩兩錯排的機率降到最低。我們依照等位基因頻率的高低，利用 RankNet 的演算法，找出突變發生的先後順序。由於單一樣本不足以支持不同基因突變之間的因果關係，我們希望藉由多個樣本來推測這些重複發生的演化關係。我們提出的方法先利用近鄰連結法(Neighbor-joining method)來建構一顆無根的樹，再從這棵樹中找出使排序錯誤最低的樹根，排序的結論是利用 RankNet 計算得到。這篇研究的目標是利用不同樣本之間的等位基因頻率，來推測突變發生的先後順序及突變的因果關係，藉此幫助我們探索癌細胞的進化過程。

Title

Inference of Mutation Order through the RankNet Algorithm

Abstract

Allele frequency is the relative frequency of a variant at a particular locus, and larger cellular occupancy of a mutation is associated with earlier mutations. RankNet is a machine learning algorithm that aim to reduce pairwise ranking errors. We use it to infer the right order of mutation occurrence. Since the evolutionary order of gene mutations by only one sample is not informative to any causal relationship, we use multiple samples to infer the causal relationship of those mutations. We proposed a procedure to solve this problem. Our first step is to use Neighbor joining method to constructed an unrooted tree according to the allele frequencies of mutations. We then find the optimal tree by choosing a root with minimum rank error rate. The aim of this thesis is to use the mutation allele frequencies from multiple samples to infer the the right order of mutation occurrence and reconstruct the most likely phylogenetic tree for the recurrent mutations of cancer cells.