摘要

解碼基因為人類帶來了期望，因為可藉此了解遺傳及環境引起的人類進化、疾病和 DNA 內部的構造功能。人類基因組序列宣示將在 2005 年完成基因定序，為加速如此龐大的基因鹼基序列定序尋找基因，需仰賴電腦科技運算的基因預測程式，因此許多程式陸續地被開發。然而，基因預測工具相當多，目前這些程式都無法相當正確的尋找基因，還未能完全真正解決許多基因結構預測的問題。

本研究結合統計特徵的計算分析與候選 exon 區段計分演算兩種方式進行基因尋找。此方法的主要目的是定義 exon 序列及基因序列邊界，而其邊界是藉由比較基因序列區段的鹼基組成統計特徵，以及邊界的函數進行劃分。

本研究之基因預測的程式整合了電腦運算及統計技術，可在配合實驗室研究下幫助生物學家分析新定序的序列，並致力克服人類基因組複雜的結構，迅速搜尋基因，能以統計量化數據窺得序列全面的特徵，不需複雜的統計模式進行推估，仰賴訊號正確性，未來希望調查 exon 中鹼基出現頻率規則使得基因預測在實驗研究過程中化繁為簡，達成生物學家生物實驗研究先機，對於醫學研究影響深遠。

Abstract

Decoding gene bringing an expectation to understand how the heredity and environment affect human evolution, human disease and DNA structure and function. Human Genome Project publicizes that entire human gene sequencing will be finished in 2005; however the progress is uncertain so that many gene prediction programs are designed continuously to speed up gene finding in massive amount of genomic sequences. Although gene prediction programs are developed for the complex structure of human genomic sequences, but a accuracy program still be needed.

The research combined statistical characteristic analysis and exon region score algorithm to find gene region. The purpose of our method is to define exon region and gene region boundary. The boundary is generated by comparing base composition of gene region and scoring function of boundary.

The program integrates computer science and statistical analysis which tries to solve the problem of complex human genome structure could help biologists to find gene rapidly, the sequences are numeral to analysis global structure of genomic sequences without complex statistical model to depend on signal to estimate.

Future work is hopefully to help biologists to investigate exon frequency and discover exon composition rule to predict gene automatically.

誌謝

在論文完成的過程中，首先感謝指導教授莊振村老師在資訊方法上面臨困問題時給予鼓勵，並且不斷的啟發自己去思考新的創意與靈感，影響我追求學問與研究的精神。此外，相當感謝口試委員王旭正博士及周志賢博士，百忙之中抽空審閱論文，給予指教，才能完成一份完善的研究論文。

本論文的研究需要相當多的分子生物知識，在充實分子生物的知識時，感謝吳禮字老師及林妍如老師以其專業知識爲不懂的地方耐心解惑。

最後在完成論文的討論過程中，感謝鎮嘉對於方法上的建議與意見，並且對於完成論文過程中不斷給予支持與鼓勵，也感謝再我完成論文的期間所有關心過我的朋友與同學，最後感謝我的父母提供一個舒適安穩的環境，使我無後顧之憂的完成碩士學位，因此將這份榮耀獻給我的家人。

研究生 黃梅芬 謹致

中華民國九十三年六月

目錄

圖目錄

表目錄