透過 Paralogs 概念建置的 SWOP 系統來找出疾病的路徑

Disease Pathway Finder Using Web-based SWOP (SideWays Observation via Paralogs) System by the Concept of Paralogs

中文摘要

近十年來， microarray 技術的進步，讓我們可以發現不同的基因在病理學上或是生理學上表現不同；但是到目前為止仍然沒有證據可以證明他們之間的關係。在這篇研究當中，我們發展了一個全新方法，透過 paralogs 的概念、microarray 基因表現的資料、基因網路和 GO 的資料來找到疾病和正常路徑之間交叉的基因。首先透過生物反應(biological reaction)，如白質交互作用(protein-protein interactions)、蛋白質調控(protein-DNA regulations)、磷酸化(phosphorylations)和甲基化(methylations)把基因組(genome)連結成一個巨大的基因網路。接下來根據每個基因在 GO 的特性把整個網路切成成千上萬條路徑，也因為如此就可以把單獨的基因表現量換成路徑的表現量。最後搜尋疾病和正常的路徑之間的 paralogous 基因，也許可能會在兩組之間找到相同的基因，而這相同的基因可能和 paralogous 基因有關。在這套系統中，使用者可以上傳人類、老鼠以及大鼠的 microarray 資料，透過系統的分析後所得到的 MI 就可以找到兩組有 paralogous 基因的兩組路徑，而且可以以圖形化的方式呈現兩組路徑，而其中相同的基因則會特別的標示出來。SWOP 系統透過 paralogs 和複雜的 microarray 資料所找到疾病和正常路徑兩組之間相同基因，或許可以拿來研究藥物所要抑制的基因。

英文摘要

From the viewpoint of evolution, gene duplication produces two functionally redundant, paralogous genes and thereby frees one of them from selective constraints. Divergent evolution has made tumor counterpart cells survived easier. Recently, microarray emerges as a nice approach to find distinct individual genes that differentially expressed between physiological and pathological cells, where, however, no clues to show the relationship among these gene cluster. Here in the research, we developed a brand new approach that combines the concept of paralogs, microarray gene expression data, gene network, gene ontology to obtain intersectional genes where the point that disease has gone astray from normal pathways. Firstly, we exerted the biological reactions, such as protein-protein interactions, protein-DNA regulations, phosphorylations, methylations, etc. to the genomes to build up network connectivity. Secondly, we partitioned the network into hundreds of thousands of pathways according to the GO cellular component properties of each gene, where helps to calculate the differentially expressed pathways rather than just differentially expressed genes. Finally, We searched for paralogous genes appeared in the contrary

differentially pathways of normal and disease status, respectively. It is very likely to find an intersectional gene between different pathways, because this intersectional gene has higher probability to associate with both these paralogous genes in different situation.

Users are allowed to upload their microarray data of human, rat or mouse. The system would identify differentially expressed pathways between normal and disease tissues by calculating with mutual information methods. The matched counterpart pathways that associated paired paralogs would be identified and visualized in a graph. The intersectional genes would be therefore spotted and highlighted in the graph. This SWOP system facilitates to transform complex microarray data into simplified intersectional genes between disease and normal pathways by using the duplicate paralogs as which indicate the shortest evolutionary time where disease go wrong from normal pathways. In addition, it helps to design drug targets to inhibit the disease pathways.