

# 國立交通大學

生物資訊及系統生物研究所

碩士論文

同源網路模組家族網站之設計和實作：  
以同源模組研究脊椎動物的模組交互作用網路

Design and Implementation of MoNetFamily Web  
Service : using homologous modules and module-module  
interaction network in vertebrates

研究生：游尚文

指導教授：楊進木 教授

中華民國壹百零一年七月

# 同源網路模組家族網站之設計和實作：以同源模組研究脊椎動物的 模組交互作用網路

學生：游尚文

指導教授：楊進木

國立交通大學 生物資訊與系統生物所碩士班

## 摘 要

模組(module)是指一群具有高度連結並執行特定生物功能的蛋白質所組成的基本單位，因此，在了解細胞的機制與功能上，模組和模組交互作用(module-module interaction, MMI)網路扮演著不可或缺的重要角色。我們建置了同源網路模組家族網站(MoNetFamily)，可以提供使用者查詢之一個或數個蛋白質，定義出其模組、同源模組(homologous module)[或稱模組家族(module family)]、以及跨物種的模組交互作用網路，同源網路模組家族網站首先將使用者所查詢之蛋白質透過局部序列比對工具(如BLASTP)去搜尋模組模板(module template)資料庫(包含 1,785 實驗記錄模組和 1,252 結晶結構模板)並找出相似的候選模組(module candidate)，進一步，透過蛋白質交互作用(protein-protein interaction, PPI)家族(PPI family) 建構候選模組之同源模組。根據同源模組和蛋白質交互作用，我們採用超幾何分布(hypergeometric distribution)統計建構模組間的交互作用，並建立跨多個物種的模組交互作用網路。此外，同源網路模組家族網站分別在人類、老鼠、斑馬魚的模組交互作用網路中提供各個候選模組的鄰近模組和及其模組交互作用。最後，同源網路模組家族網站也針對每個模組家族進行蛋白質演化保留性分析、蛋白質交互作用組成分析、及模組家族之功能性註解。

實驗結果顯示，同源網路模組家族網站可提供圖像化模組交互作用網路(如:在人類模組網路中包含 1,278 模組和 9,678 模組交互作用)，並透過模組家族與鄰近模組提供所查詢之蛋白質的相關功能性註解。我們相信透同源網路模組家族網站建構出跨物種的同源模組以及模組交互作用網路，將有助於探討及了解模組演化及生物細胞機制。同源網路模組家族網站網址如下: <http://monetfamily.life.nctu.edu.tw>

# Design and Implementation of MoNetFamily Web Service : using homologous modules and module-module interaction network in vertebrates

Student: Shang-Wen Yu

Advisor: Dr. Jinn-Moon Yang

Institute of Bioinformatics and System Biology  
National Chiao Tung University

## ABSTRACT



A module is a fundamental unit forming with highly connected proteins and performs a certain kind of biological functions. Modules and module-module interaction (MMI) network are essential for understanding cell mechanisms and functions. The MoNetFamily web server can identify the modules, homologous modules (called module family), and MMI networks across multiple species for the query protein(s). This server first finds module candidates of the query by using BLASTP to search the module template database (1,625 experimental and 1,252 structural templates). MoNetFamily then infers the homologous modules of the selected module candidate using protein-protein interaction (PPI) families. According to homologous modules and PPIs, we statistically decided MMIs and MMI networks across multiple species. For each module candidate, MoNetFamily identifies its neighboring modules and their MMIs in module networks of *H. sapiens*, *M. musculus*, and *D. rerio*. Finally, MoNetFamily shows the conserved proteins, PPI profiles, and functional annotations of the module family. Experimental results demonstrate that the server can be useful for MMI network (e.g., 1,278 modules and 9,678 MMIs in *H. sapiens*) visualizations and query annotations using module families and neighboring modules. We believe that the server is able to provide valuable insights to determine homologous modules and MMI networks across multiple species for studying module evolution and cell mechanisms. The MoNetFamily sever is available at <http://monetfamily.life.nctu.edu.tw>.

## 誌謝

萬萬沒有想到有這一天，我能從碩士畢業，在研究所的這2年，真的學習到很多，不管是生物、程式的知識，更重要的是做事的態度，在研究的過程中，沒有實驗室一起同甘共苦的夥伴們，一起歡笑，互相鼓勵扶持，我今天也不可能畢業，我們實驗室是最棒的!

這兩年中，都要多虧我的指導教授楊進木老師在這2年期間對我的教導，在研究過程適時的導正我們的方向，不僅在研究上有老師輔助，在生活中更讓我學習到很多新的人生觀點以及面對工作的處事態度，老師的教導尚文一定謹記在心，永不會忘記!

這趟學習旅程中，衷心感謝我們這組的組員峻宇學長、怡瑋學姊、宇書學長、怡馨學姊，教導學弟尚文相關的生物知識及一起討論，朝著目標邁進，過程雖然艱苦，但在過程中有實驗室朋友的陪伴及歡笑，讓尚文能有動力持續朝目標邁進。尤其感謝峻宇組長對我的細心教導，讓我在研究上能了解自己的缺失因此更加進步。也謝謝實驗室凱程學長、志達學長、一原學長、章維學長在研究過程中旁敲側擊的點醒我們，讓我們研究能更紮實茁壯。

更重要的是在研究所過程中，因為有父母在一旁的默默支持，讓尚文可以心無旁騖的專心於研究上，沒有您們的支持及鼓勵，也不會有今天的尚文，謝謝爸爸媽媽!真的很愛您們，還有感謝我女友貞穎，在這一路上的噓寒問暖，陪伴著我度過這2年研究生生活。

能來到 BioXGEM 這實驗室認識大家，真的是尚文的福氣，也謝謝這實驗室帶給尚文的一切，讓尚文能更上一層樓，在未來的旅程中，尚文會將您們永謹記在心，感謝您們。

總目錄

摘要 .....	I
ABSTRACT .....	II
誌謝 .....	III
表目錄 .....	V
圖目錄 .....	VI
壹、緒論 .....	1
一、研究背景 .....	1
二、研究動機 .....	4
三、論文總覽 .....	5
貳、研究材料與方法 .....	7
一、同源模組 (Homologous module) .....	10
二、模組交互作用網路 (Module-module interaction network) .....	14
三、蛋白質模組註解 .....	16
參、結果與討論 .....	17
一、同源模組的合理性 .....	17
二、模組交互作用網路的合理性 .....	21
三、範例分析 .....	29
(一)、JAK2-PAFR-TYK2 模組 .....	29
(二)、TRAP-SMCC mediator 模組 .....	35
肆、結語 .....	38
一、總結 .....	38
References.....	44

## 表目錄

表一、建構和驗證所使用的資料庫列表 .....	6
表二、JAK2-PAFR-TYK2 模組的鄰近模組功能註解 .....	39



## 圖目錄

圖一、舉例並描述如何藉由細胞網路中的模組來探討細胞生化機制及交互作用網路之形成 .....	2
圖二、模組與模組之間的交互作用是構成生化機制反應的主要骨幹 .....	3
圖三、同源網路模組家族網站架構圖 .....	8
圖四、模組與模組交互作用之蛋白質交互作用組成分析圖 .....	9
圖五、透過蛋白質—蛋白質交互作用家族(PPI family)定義同源模組之架構圖 .....	12
圖六、KEGG MODULE 的拓模相似度統計數據圖 .....	13
圖七、JAK2-PAFR-TYK2 模組和 Crystal structure of the hexameric human IL-6/IL-6 alpha receptor/gp130 模組內部及兩模組間之蛋白質交互作用組成 .....	15
圖九、模組交互作用網路生化途徑剖析圖 .....	23
圖十、模組交互作用網路節點分支分佈圖 .....	24
圖十一、GO terms(BP、CC)在人類、老鼠、斑馬魚的模組交互作用網路註解分析 .....	26
圖十二、透過模組交互作用網路描述細胞增殖的行為 .....	28
圖十三、細胞增殖行為中 8 個模組間的詳細蛋白質交互作用組成 .....	28
圖十四、在人類、老鼠、斑馬魚的 JAK2-PAFR-TYK2 模組與其鄰近模組 .....	33
圖十五、人類 IL-6 蛋白質(chain B of 1p9m)與老鼠、斑馬魚之同源蛋白質的序列比對圖 .....	34
圖十六、同源網路模組家族網站搜尋 TRAP-SMCC mediator 模組之網頁介紹 .....	37
圖十七、人類、老鼠、斑馬魚之 TRAP-SMCC mediator 模組與鄰近模組 .....	37

# 壹、緒論

## 一、研究背景

生物體網路(蛋白質與蛋白質交互作用網路)在了解生物體過程中一直都是十分重要的議題。過去研究指出蛋白質模組(module;以下簡稱模組)是由一群具有高度連結並執行特定生物功能的蛋白質所組成。在相同時間和空間尺度下,分子間精確地聚集且協同作用對於生物程序是不可或缺的,例如細胞週期和轉錄作用[1]。因此,藉由模組的概念將有助於更進一步了解生物體網路的形成(圖一)。

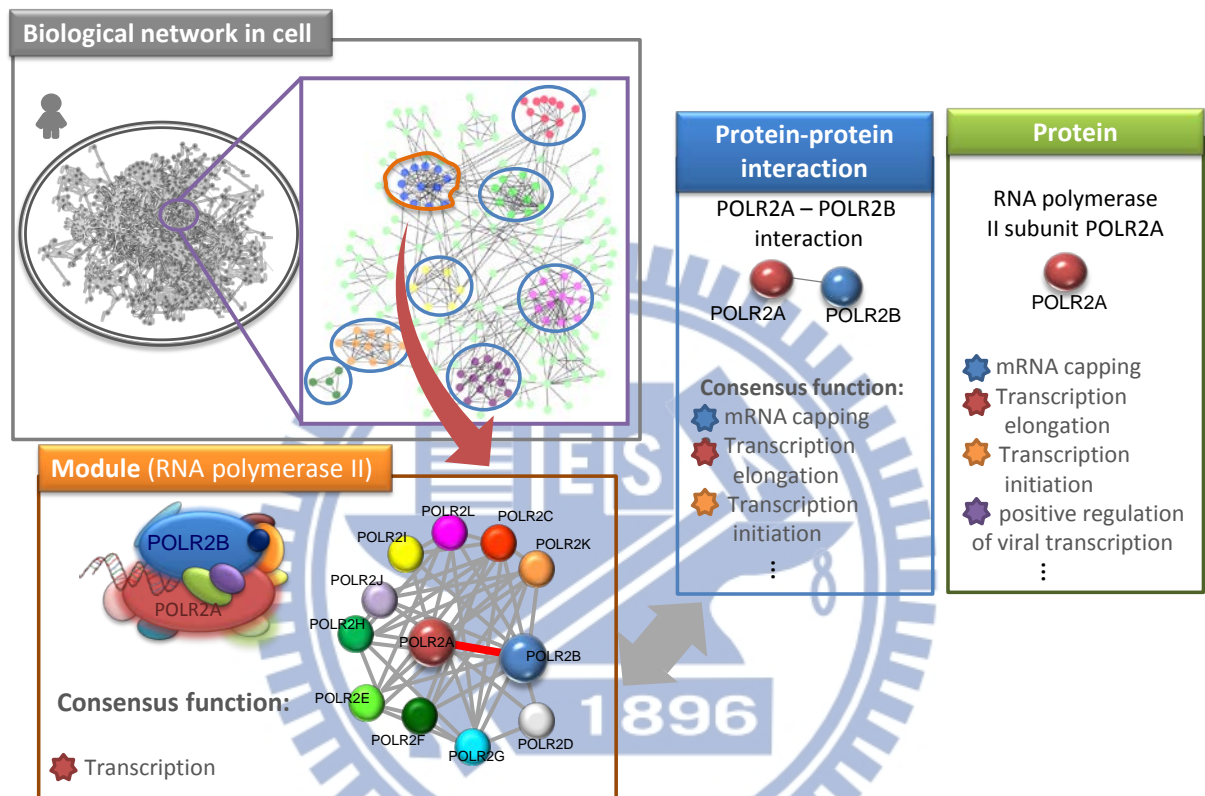
一群蛋白質在生物體中形成模組去執行特定的生化反應途徑(如:變異、發育)時,相較於此物種交互作用網路的其他部分,具有較高的自主性(圖一)[2, 3]。近年來開始有研究提出模組與模組之間的交互作用在細胞網路中構成生化機制反應的主要骨幹[4, 5],因此了解模組和模組交互作用(module-module interaction;MMI)網路,提供一個新系統生物觀點去了解細胞內生化機制(圖二)。

模組辨識已經有很多成形的的方法,如:使用基因表現(gene expression)的實驗資料去定義出模組、使用圖學(graph-based)的方法在蛋白質交互作用網路中定義出模組[2, 6, 7]。然而,在模組與模組交互作用中,相關文獻指出模組與模組之間的蛋白質交互作用必需具有顯著數量且可靠的交互作用,才可被認定兩模組具有交互作用的關係[4, 5]。

新序列基因可透過尋找直向同源(orthologs)關係的已知蛋白質序列來提供一些線索註解此新序列基因的功能。然而,一個蛋白質往往可在蛋白質相互作用網路中(PPI network)透過與其交互作用之蛋白質來提供新的生化功能註解[8],因此同源蛋白質(homologous proteins)與同源蛋白質交互作用(homologous PPI)[或稱蛋白質交互作用家族(PPI family)]可在多個物種的蛋白質交互作用網路中提供新的蛋白質功能的註解[如:演化高度保留的功能區塊(Domain)與區塊間的交互作用]。隨著大量蛋白質交互作用資料的

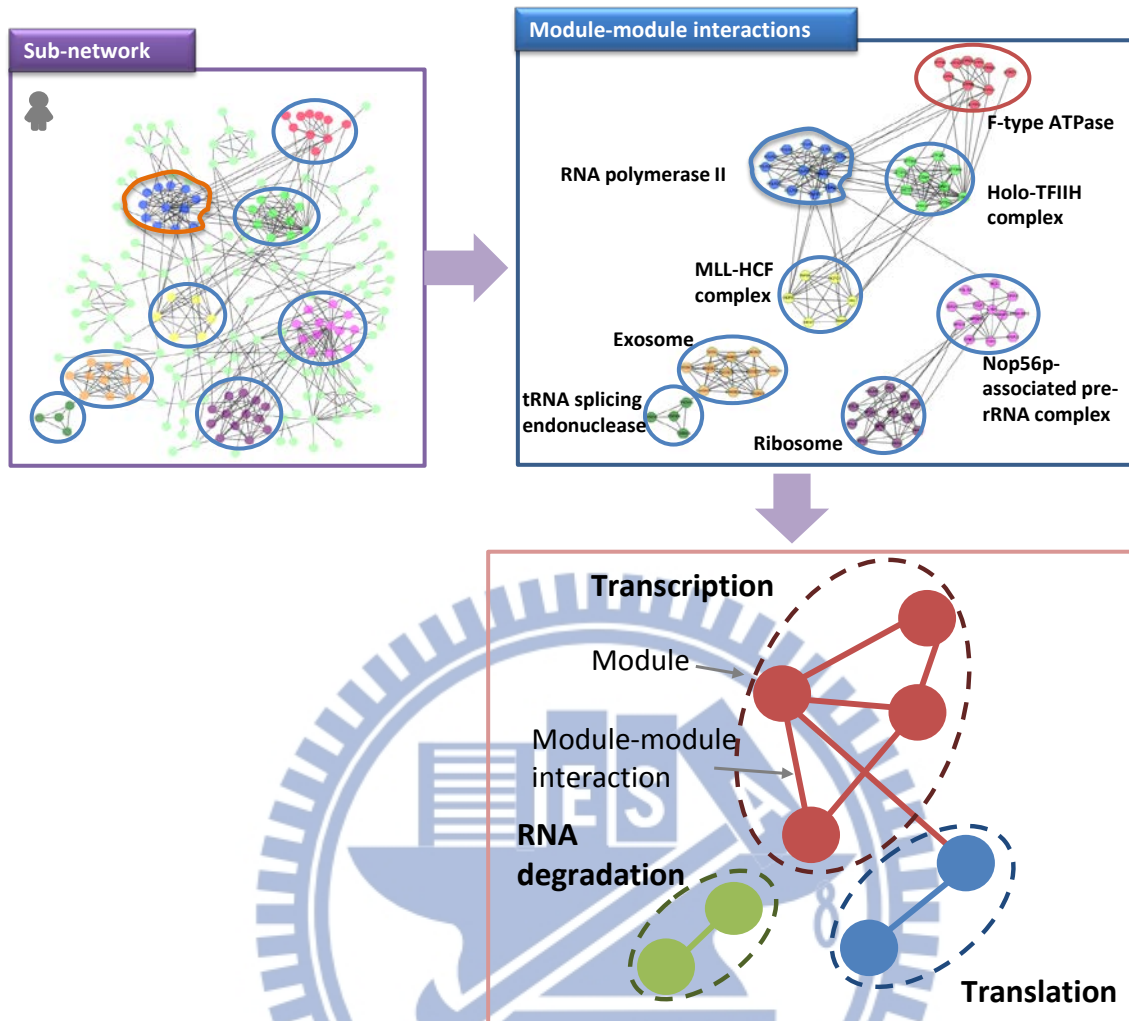


快速增加以及同源蛋白質交互作用概念，不僅有助於建立同源模組及模組交互作用，還可提供比較各物種模組交互作用網路，找出演化保留性的和物種特化之模組及模組交互作用，進而，藉由跨物種比較來探討生物系統中生物功能調控的差異性。



圖一、舉例並描述如何藉由細胞網路中的模組來探討細胞生化機制及交互作用網路之形成

模組在生物體中會群聚在一起執行特定的生化反應途徑，相較於此物種交互作用網路的其他部分，具有較高的自主性，如圖中所示在人體蛋白質交互作用網路中的 RNA polymerase II 模組是執行 DNA 轉錄成 mRNA 之前驅物，而此模組中的蛋白質與蛋白質交互作用(例: POLR2A 和 POLR2B)將有助於了解此模組的生物機制和註解模組中其他未知蛋白質的註解。



圖二、模組與模組之間的交互作用是構成生化機制反應的主要骨幹

模組是具有高度連結並執行特定生物功能的蛋白質所組成，例: RNA polymerase II 和 F-type ATPase，藉由不同模組之間的模組交互作用(module-module interaction, MMI)，可提供新的系統生物角度去探討細胞的生化機制。

## 二、研究動機

過往已經有很多方法可以定義出有意義的模組(如:功能性模組[2, 6]和演化模組[7, 9])且已經有數個資料庫提供跨物種的生化模組[10-12]，然而我們觀察到目前缺乏對於模組之間的關聯性的探討，近年來，新系統生物角度開始重於觀察模組以及模組與模組交互作用，並針對單一特定物種(如：人類、酵母菌)[4, 5, 13]，本研究中我們想要發展建立跨多個物種的模組、模組交互作用、及以模組為單元建構之模組網路(MMI network)，透過這些研究成果將有助於分析細胞內生物功能和生化途徑，並藉由跨物種的比較觀察尋找模組網路中演化保留或特化的部分。過去的相關研究，大都僅針對單一物種去找出模組及建立模組交互作用網路，因此，這些方法在同時建立多個物種的模組交互作用網路上是十分耗時或受到侷限。我們提出模組家族(module family)的概念，並透過大型完整基因資料庫(如:Integr8 資料庫包含 2,274 個物種[14])來建構出同源模組及跨多個物種模組交互作用網路，模組家族的概念相似且延伸自過去研究所提出的蛋白質序列家族[15]、蛋白質複合體家族(protein structure family)[16]以及蛋白質與蛋白質交互作用家族(PPI family)[17]，透過建構出跨物種的同源模組及模組交互作用網路，將有助於生物學家去探討模組演化以及細胞代謝過程等重要議題。

針對這項研究，我們建置出同源網路模組家族(MoNetFamily)網站，此網站可針對使用者所詢問之一個或數個蛋白質，定義其同源模組(module family)；圖像化此模組位模組網路之位置及鄰近之交互作用模組；兩模組間詳細蛋白質交互作用組成；模組家族高度保留的 Gene Ontology(GO)生化功能註解。根據目前我們所知，同源網路模組家族網站是第一個提供相關服務的全球服務網站。

### 三、論文總覽

我們提出模組家族的概念及建立多個脊椎生物之模組與模組交互作用網路，表一展現其建構過程中所使用的資料庫，而其做法概要描述於第二章節中，首先取模板模組建構出同源模組(homologous module)[或稱模組家族 (module family)]，接著透過有實驗紀錄的蛋白質—蛋白質交互作用資料庫和我們實驗室已發表的同源蛋白質—蛋白質交互作用資料庫，結合模組家族建構出跨物種的模組交互作用網路。進一步，基於同源模組的關係代表著演化上它們來自相同祖先，應會執行類似的生物功能，且其鄰近的模組也會執行相似的生化途徑，因此，我們使用透過同源模組與網路鄰近模組之 GO term (生化功能註解)去註解使用者詢問之蛋白質。在第三章節，我們驗證我們所建構出的跨物種模組網路是否符合生物及網路拓樸特性，並舉例說明如何透過我們的模組家族和跨物種之模組交互作用網路提供使用者新的系統生物角度來分析註解所詢問之蛋白質。

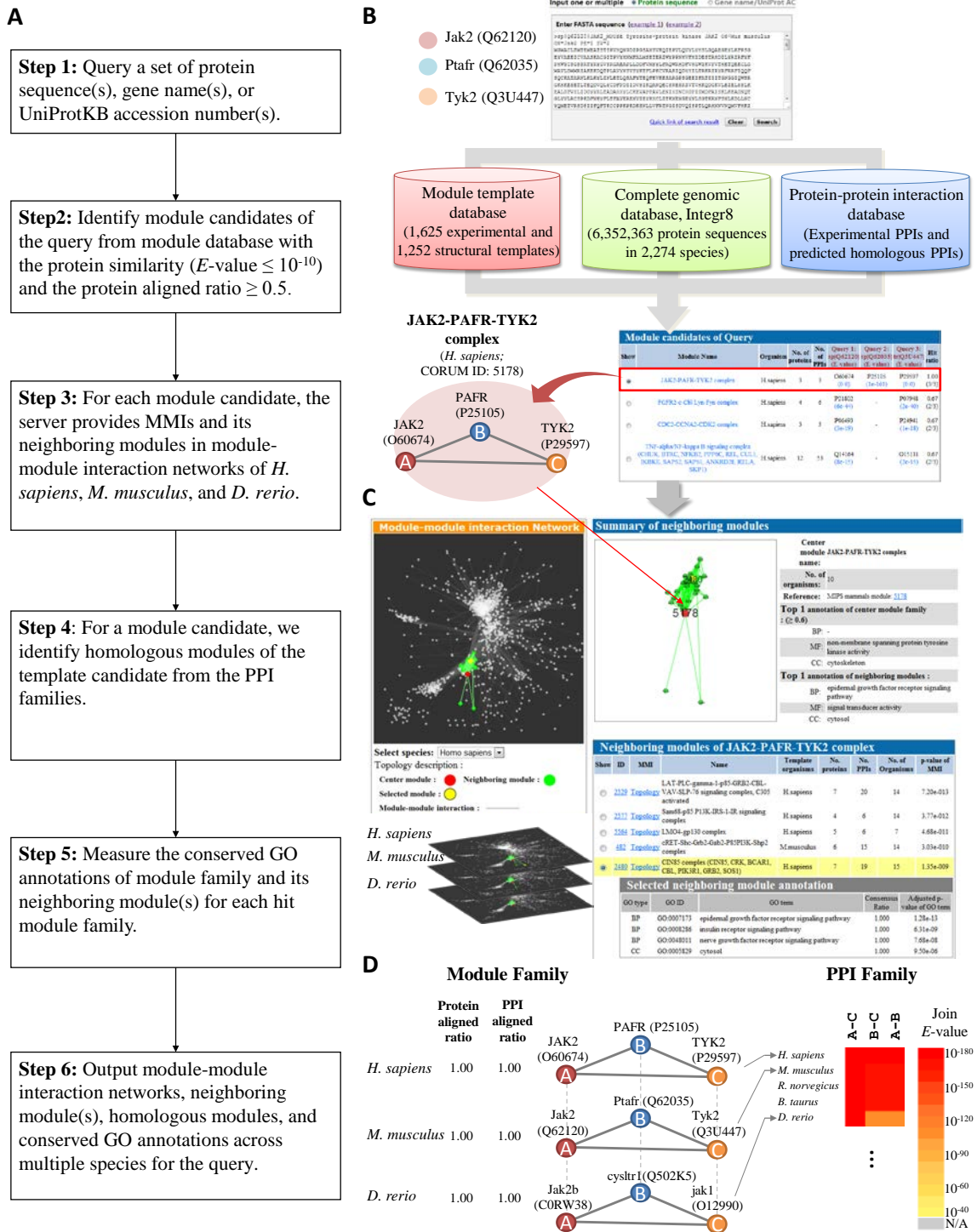


表一、建構和驗證所使用的資料庫列表

Data sets	Comments
<b>MIPS CORUM database [18]</b>	The CORUM database using as module template set provides manually annotated protein complexes, which assemble multiple proteins to perform biological functions, from mammalian organisms.
<b>Annotated PPI database</b>	275,787 experimental PPIs in the annotated PPI database (IntAct [19], BioGRID [20], DIP [21], MIPS [22], and MINT [23])
<b>Predicted homologous PPI set</b>	Our previous sequence-based and structure-based homologous PPIs with joint E-value $\leq 10^{-40}$ [24] and Z-score $\geq 3$ [25], including 290,137 sequence-based PPI families and 86,252 structure-based PPI families
<b>Integr8 database[14]</b>	A complete genomic database (Integr8 version 103, containing 6,352,363 protein sequences in 2,274 species)
<b>KEGG MODULE database [26]</b>	KEGG organism-specific modules is defined as a tight functional unit and complexes in the pathway through a set of orthologs
<b>Gene Ontology (GO) database [27]</b>	We derive GO biological process (BP) to annotate homologous modules and GO molecular function (MF) to annotate core components of module family.
<b>Extended module data set</b>	Extending one-layer PPIs and proteins for each protein in an original module through homologous PPIs
<b>Random data sets</b>	Each module template constructed 50 random modules, which were selected randomly the same protein number from the genome of template's organism, and each random module was the same number of proteins with the module template.

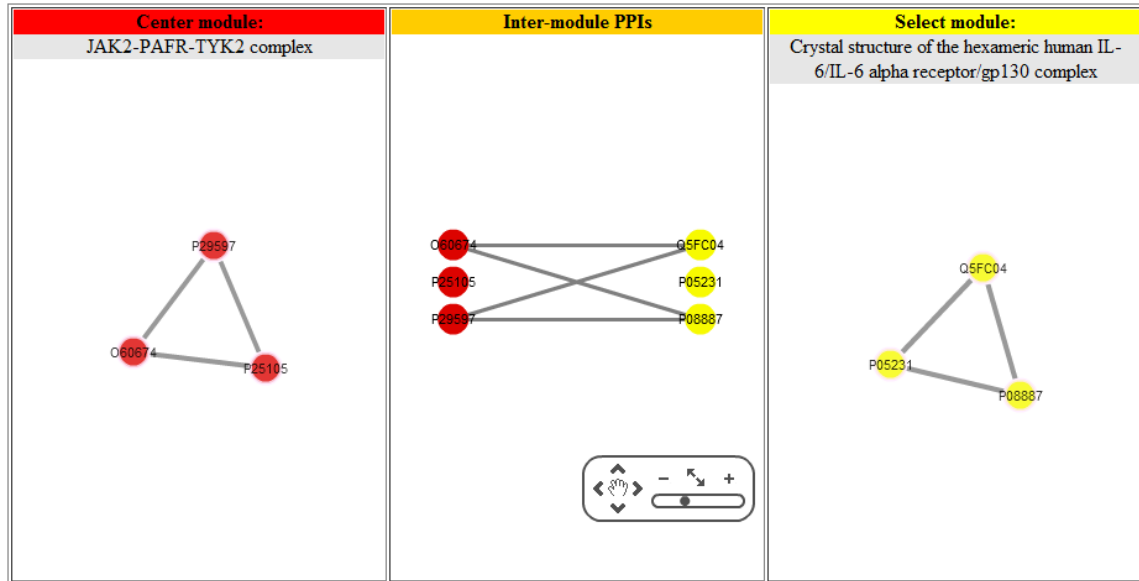
## 貳、研究材料與方法

我們要建立出同源模組及多個脊椎生物之模組交互作用網路並且將它實現於網站上供使用者使用，總流程如(圖三 A)，使用者可以輸入他們有興趣的蛋白質序列、基因名稱、UniprotKB，接著我們使用 BLASTP 為使用者找出有哪些模板模組內蛋白質 ( $E\text{-values} \leq 10^{-10}$ ) 符合使用者所輸入蛋白質(圖三 B)，所搜尋的資料庫是取自哺乳動物蛋白質複合物資料庫(CORUM[ref])共 1,785 個模版模組以及蛋白質資料庫共 1,252 個模版模組(PDB[ref])，將這兩個資料來源整合去除重覆的模組後共保留 1,975 個模組，所有模組總共含有 4,659 個蛋白質，並且藉由蛋白質蛋白質交互作用家族提供使用者此模板模組在其他物種(人類、老鼠、斑馬魚)是否有出現同源的模組[8, 28, 29] (圖三 D)，進一步我們提供多個脊椎生物之模組與模組交互作用網路，也供應在各個物種中此模組與各個鄰近模組中有哪些蛋白質有參與交互作用(圖四)，而在如何判定模組與模組之間是否存在交互作用的計算方式，則是使用超幾何分布方法去計算出。在同源網路模組家族 (MoNetFamily) 網站中，我們主要供應脊椎生物中 3 個物種的模組交互作用網路(人類共找尋到 1,818 個模組、老鼠共找尋到 1,801 個模組、斑馬魚共找尋到 1,257 個模組)，建構出的同源模組以及定義出的鄰近模組，我們使用 gene ontology 的註解(biological processes (BP), cellular components (CC), and molecular functions (MF))來驗證(圖三 C)，驗證的方式是觀察他們是否在這些註解中具有一致性(執行相同的功能)、以及在這些註解中的 p-values 是否有過我們的臨界值。這網站不僅能提供同源模組的查詢，也能視覺化的提供此模組在跨多個脊椎物種的模組交互作用網路中與鄰近的模組彼此交互作用的差異性。



圖三、同源網路模組家族網站架構圖

透過使用者所查詢之 Jak2, Ptafr, and Tyk2 蛋白質作為例子，說明同源網路模組家族網站之運作原理及結果分析。(A) 主要運作流程圖。(B) 針對所查詢之蛋白質透過局部序列比對工具(如 BLASTP)去搜尋模板模組資料庫，並找出相似的候選模組。(C) 提供所選擇之候選模組(如：CORUM ID: 5178)在人類、老鼠、斑馬魚模組網路上的鄰近模組和詳細模組交互作用，並提供候選模組家族之保留性生化功能註解(如：GO 註解)。(D) 提供所選擇之候選模組其建構出的同源模組與其內部蛋白質及蛋白質交互作用之詳細組成。



圖四、模組與模組交互作用之蛋白質交互作用組成分析圖

透過使用者所查詢之 Jak2, Ptafr, and Tyk2 蛋白質所搜尋到的模板模組 (CORUM ID: 5178; JAK2-PAFR-TYK2 模板)，觀察此模組與其鄰近模組(Crystal structure of the hexameric human IL-6/IL-6 alpha receptor/gp130) 內部及兩模組間之詳細蛋白質交互作用組成

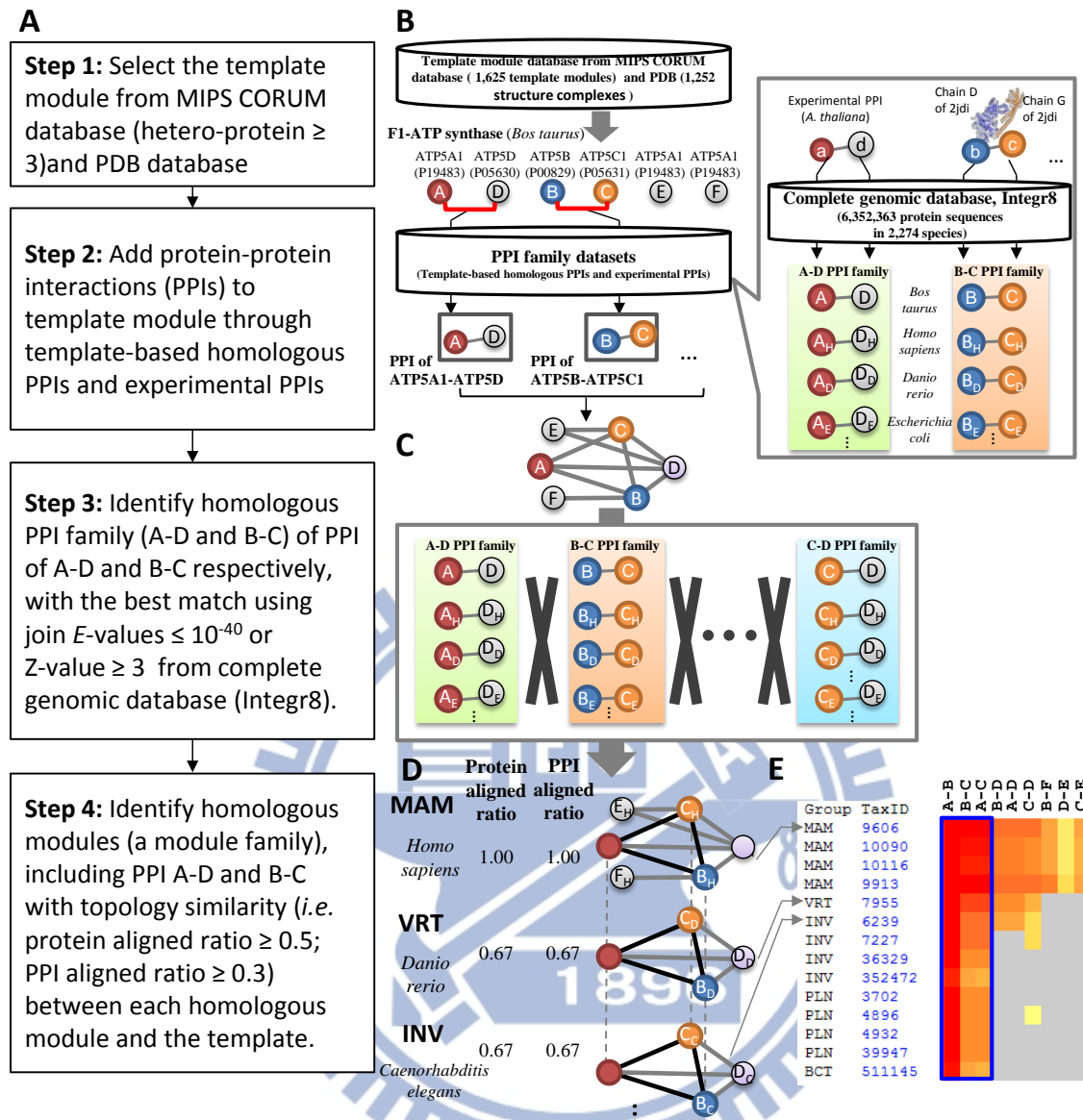


## 一、同源模組 (Homologous module)

要建構出跨物種的模組交互作用網路，首要的步驟就是要先建構出跨物種的同源模組(模組家族)，因此圖四是在展現我們如何以模板為基礎來定義同源模組的方法，流程步驟如(圖五 A)：首先的步驟我們要先選擇模版模組資料庫，我們分別取自哺乳動物蛋白質複合物資料庫(CORUM[30])共 1,785 個模版模組以及蛋白質資料銀行共 1,252 個模版模組(PDB[31])，在 CORUM 資料庫的蛋白質模組是由人工實驗定義出，他們認為的蛋白質模組是由多個蛋白質共同執行特定功能而成[30]，但在 CORUM 資料庫裡的模組並沒有賦予這些模組內部的蛋白質與蛋白質交互作用所以我們為這些模組賦予蛋白質蛋白質交互作用，主要是從有實驗紀錄的蛋白質蛋白質交互作用資料庫共有 461,077 個蛋白質與蛋白質交互作用(i.e. IntAct [19], BioGRID[32], DIP[21], MIPS[33], MINT[23])和預測的同源蛋白質蛋白質交互作用資料庫[21, 34](圖五 B)，在預測同源蛋白質蛋白質交互作用資料庫中我們是使用實驗室先前的研究方法，從完整基因組資料庫中(Integr8 版本 103, 共有 6,352,363 蛋白質序列在 2,274 物種)藉由蛋白質交互作用以及蛋白質複合體家族的概念( $\text{joint } E\text{-value} \leq 10^{-40}$  和  $Z\text{-value} \geq 4$ )找出可能存在的同源蛋白質蛋白質交互作用，接下來定義同源模組也是藉由蛋白質交互作用家族去找尋出蛋白質相似、蛋白質與蛋白質交互作用相似、拓樸相似的同源模組(圖五 C、D)，因此透過模板模組建構出模組家族，這裡舉個範例來說明我們如何定義同源模組，假設我們有個模板模組 T 此模板模組包含 3 個蛋白質(A、B、C)以及 3 條蛋白質與蛋白質交互作用(A-B、A-C、B-C)，定義同源模組主要有 3 步驟: (1)我們將模組 T 內的蛋白質分別去做序列比對(BLASTP  $E\text{-values} \leq 10^{-10}$ )找出最具有顯著相像的同源蛋白質(A'、B'、C');(2) 接下來我們使用共同序列比對( $\text{joint } E\text{-value} \leq 10^{-40}$ )去找出顯著地同源蛋白質與蛋白質交互作用(A'-B'、A'-C'、B'-C');(3) 將同源模組(A'-B'-C')找出後，之後要與模板模組(A-B-C)比對他們之間的拓樸相似度，拓樸比對的方式是使用蛋白質的對齊比與蛋白質與蛋白質交互作用對齊比，對齊比的計算方法： $x/X$ ，x:同源模組的蛋白質數或蛋白質與蛋白質交互作用數 X:模板模

組的蛋白質數或蛋白質與蛋白質交互作用數，我們根據 KEGG MODULE 的 75,706 個同源模組(共有 370 個模板模組)的拓樸相似度統計數據，因此將蛋白質的對齊比定義在 $\geq 0.5$  且蛋白質與蛋白質交互作用對齊比定義在 $\geq 0.3$ (圖六)，要符合這兩項對齊比我們才會認定他是 T 的同源模組。





圖五、透過蛋白質—蛋白質交互作用家族(PPI family)定義同源模組之架構圖

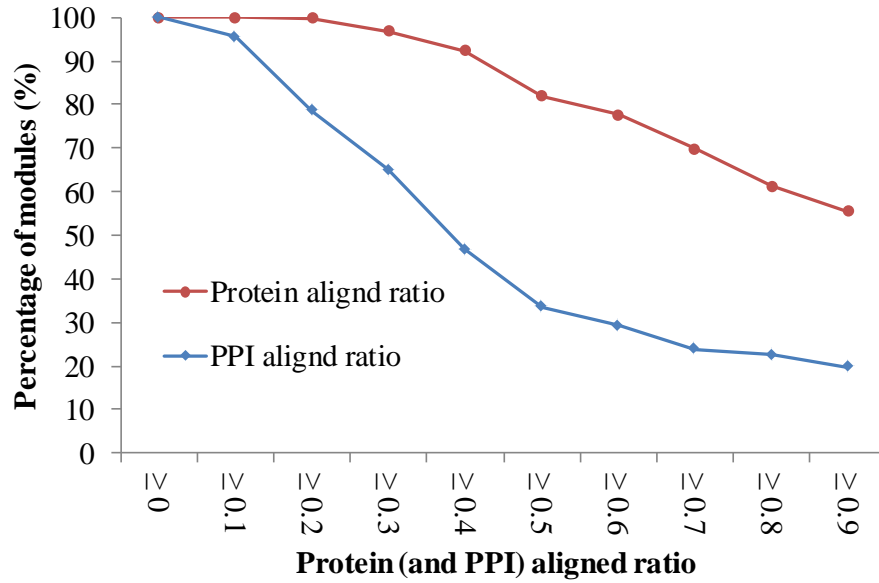
(A) 主要流程圖。

(B) 透過結合實驗紀錄和預測之蛋白質交互作用資料庫為模板模組註解其內部蛋白質交互作用組成。

(C) 藉由多個蛋白質交互作用家族，以實驗或結晶紀錄之模組作為模板建構其模組家族。

(D) 為 F1-ATP synthase 模組所建構之同源模組。

(E) 藉由觀察各物種同源模組之蛋白質交互作用組成，可定義出 F1-ATP synthase-IF1 模組演化高度保留之蛋白質交互作用。



圖六、KEGG MODULE 的拓模相似度統計數據圖

透過分析KEGG MODULE的75,706個同源模組(來自370個參考模組),統計拓模相似度。從圖中可以觀察到在蛋白質對齊比例中有大約82% (62,080)個模組大於0.5以上。然而,在KEGG 原始模組資料庫中並未註解模組內部之蛋白質交互作用,我們採用先前研究以序列為基礎之同源蛋白質交互作用(PPIsearch with joint  $E$ -value  $\leq 10^{-70}$ )和以結晶結構為基礎的同源蛋白質交互作用(PCFamily with  $Z$ -value  $\geq 4$ )來為KEGG 模組註解其蛋白質交互作用,並計算出蛋白質交互作用對齊比例,從圖中觀察到約有65%的模組其對齊比大於0.3以上。因此將蛋白質的對齊比例定義在 $\geq 0.5$ 且蛋白質與蛋白質交互作用對齊比例定義在 $\geq 0.3$ 。

## 二、模組交互作用網路 (Module-module interaction network)

判定兩個模組之間是否有交互作用通常都是使用兩個模組之間蛋白質與蛋白質交互作用的質量來判定[3, 27]，因此我們要先賦予在各個兩個模組之間的蛋白質與蛋白質交互作用，我們使用的資料庫是六個公認的蛋白質與蛋白質交互作用資料庫以及預測的同源蛋白質蛋白質交互作用資料庫[17, 35]，篩選的方式是使用超幾何分佈 (hypergeometric distribution)[10]，超幾何分佈是統計學上一種離散的機率分佈，主要是在計算此物件在此群體中出現機率的顯著性， $p$ -value 值越低代表越具有顯著性，公式如

$$\text{下: } p = \sum_{i=x}^n \frac{\binom{M}{i} \binom{N-M}{n-i}}{\binom{N}{n}}$$

$x$  代表兩個模組之間有多少實際的交互作用數(例:在圖七中兩個模組分別是 JAK2-PAFR-TYK2 和 the hexameric human IL-6/IL-6 $\alpha$  receptor/gp130 在此圖例中  $x=4$ )  $n$  代表兩個模組之間所有組合性的交互作用數(例:在圖七中  $n=9$ )

$M$  代表在此模組交互作用網路中所有存在的實際蛋白質與蛋白質交互作用

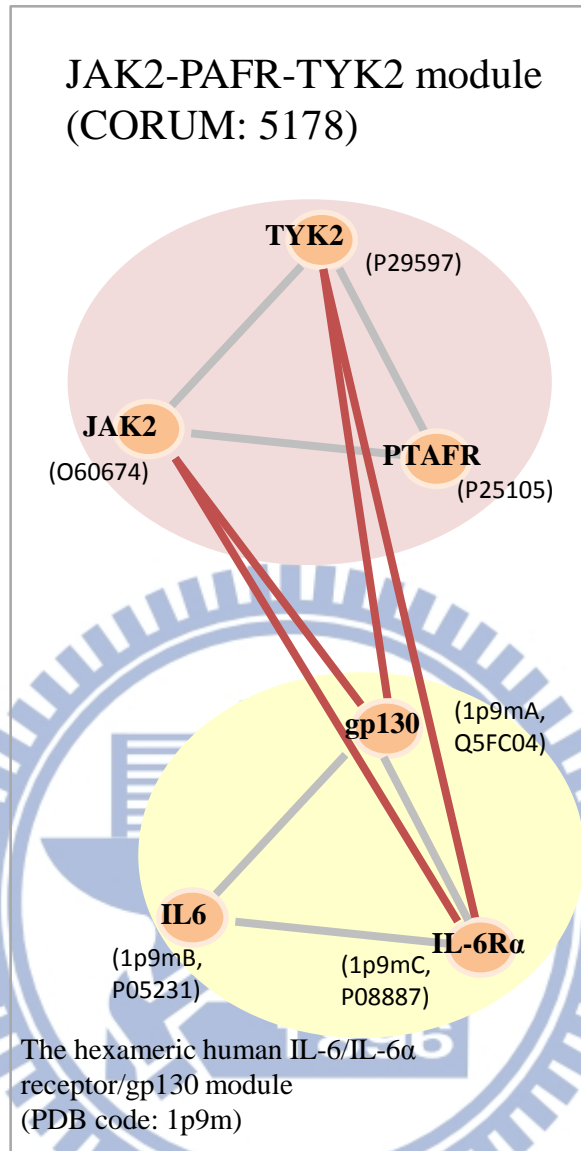
$N$  代表在此模組交互作用網路中蛋白質與蛋白質交互作用所有的可能性組合

在挑選是否有模組模組交互作用的存在我們主要訂定兩個標準

(1)  $p\text{-value} \leq 10^{-4}$

(2) 兩個模組內的各個模組參與模組與模組間的蛋白質與蛋白質交互作用裡面的蛋白質至少要有 2 個蛋白質有參與交互作用

達到以上標準，我們分別在人類、老鼠、斑馬魚 3 個物種中定義出 9,678、8,942、8,722 條模組與模組的交互作用，得以建構成 3 個跨物種的模組交互作用網路。



圖七、JAK2-PAFR-TYK2 模組和 Crystal structure of the hexameric human IL-6/IL-6 alpha receptor/gp130 模組內部及兩模組間之蛋白質交互作用組成

### 三、蛋白質模組註解

我們註解每個蛋白質模組是去觀察此模組是否與同源模組以及鄰近的模組有一致性的 GO terms，假設欲註解的模組有  $K$  個蛋白質我們針對每一個蛋白質的 GO term 去統計此 GO term 共出現在此模組的幾個蛋白質中此數量定義為  $k$ ，假設此 GO term 為  $i$ ，因此我們可以算出此  $i$  在此模組出現的一致性比例(consensus ratio ; $CRM$ )

$$CRM_i = \frac{k}{K}$$

接下來我們使用超幾何分佈去算出每個 GO term 在群體中所出現的顯著性，並且使用邦弗朗尼校正(Bonferroni Correction)進行調整，因此我們藉由統計分析  $CRM > 0.6$  和調整過後的  $p-values \leq 0.05$ [36, 37] 這兩項標準來挑選出模組代表性的 GO terms，在同源網路模組家族網站我們使用模組家族一致性的比例 (consensus ratio of module family; $CRF$ )以及 agreement ratio( $AR$ )來為使用者所查詢的模組和蛋白質加以註解，根據先前的研究[8, 29]， $CRF$  公式定義如下:  $CRF = F_a / F$ ， $F$ :在此模板模組所找尋到所有的同源模組個數， $F_a$ :在這所有的同源模組中出現此 GO term  $a$  的同源模組數量。而  $AR$  的公式如下:

$$AR = \frac{\sum_{i \in Q} A_i(CRF \geq c)}{\sum_{i \in Q} T_i(CRF \geq c)}$$

$Q$ :使用者查詢的模板模組;

$T_i(CRF \geq c)$ :此模組家族(不含模板模組)所出現且符合  $CRF \geq c$  的 GO terms 的所有加總數量。

$A_i(CRF \geq c)$ :此模板家族的模板模組所出現且符合  $CRF \geq c$  的 GO terms 所有加總數量。

## 參、結果與討論

在本研究中，我們建構出模組家族，並藉由模組家族建構出跨物種的模組的交互作用網路，這是一個全新的概念，我們分別取自哺乳動物蛋白質複合物資料庫(CORUM)共 1,785 個模版模組以及蛋白質資料庫共 1,252 個模版模組(PDB)，建構出 1,975 個模組家族，分別在人類找尋到 1,818 個模組、在老鼠找尋到 1,801 個模組、在斑馬魚找尋到 1,586 個模組，因而在模組網路中定義出的模組與模組交互作用，在人類共有 1,440 條、在老鼠共有 1,396 條、在斑馬魚共有 1,257 條。所找到的同源模組其具有高度連通性(connectivity)，且在 GO term 的表現上其同源模組也執行一致的生化功能(biological process ;BP)、相同的細胞位置(cellular component ;CC)，所建構出來的跨物種模組網路，也符合生物體網路的特性(無尺度網路;scale-free 網路)。

### 一、同源模組的合理性

為了瞭解模組家族的功能和特性的合理性，我們從大眾普遍公認的 KEGG MODULE 資料庫中收集出 370 個模板模組分別包含在 1,442 個物種中，共有 75,706 個生物體特定模組(organism-specific modules)，根據這些資料我們計算這些生物體特定模組模板對於 KEGG 的模板模組的蛋白質比例，統計出約有 56%(42,065)的生物特定模組有著 0.9 的蛋白質比例，而約有 82%(62,080)的生物特定模組有著 0.5 的蛋白質比例 (圖六)，而在模組家族中所有的同源模組與模板模組的蛋白質比例都大於 0.5。

針對於模組內蛋白質交互作用的合理性，在 KEGG 資料庫裡並沒有詳述模組內部擁有的蛋白質交互作用，這樣對於我們要驗證拓模相似度的合理性是缺乏的，所以我們為 kegg 資料庫的模組賦予蛋白質交互作用，使用的蛋白質交互作用資料庫主要是下列 3 項：

1)使用實驗紀錄的蛋白質蛋白質交互作用資料庫，共有 461,077 個蛋白質與蛋白質交互



作用(i.e. IntAct [19], BioGRID [20], DIP [21], MIPS [22], MINT [23])。

2)取自蛋白質資料銀行(PDB crystal structures [38])，共有 9,016 條的蛋白質與蛋白質交互作用。

3)我們實驗室先前研究的預測的同源蛋白質蛋白質交互作用資料庫在序列上 joint  $E\text{-value} \leq 10^{-70}$  [17]而在結構預測上  $Z\text{-score} \geq 4$ [35]。

在 75,706 個生物體特定模組中我們為 23,092 個生物體特定模組藉由上述三個資料庫賦予蛋白質交互作用，並且去統計他們的蛋白質交互作用比例，統計結果共有 65%的生物體特定模組的蛋白質交互作用比例大於 0.3(圖六)，基於上述的觀察我們因此個別將預測的同源模組蛋白質比例設在 0.5，將蛋白質交互作用設在 0.3。

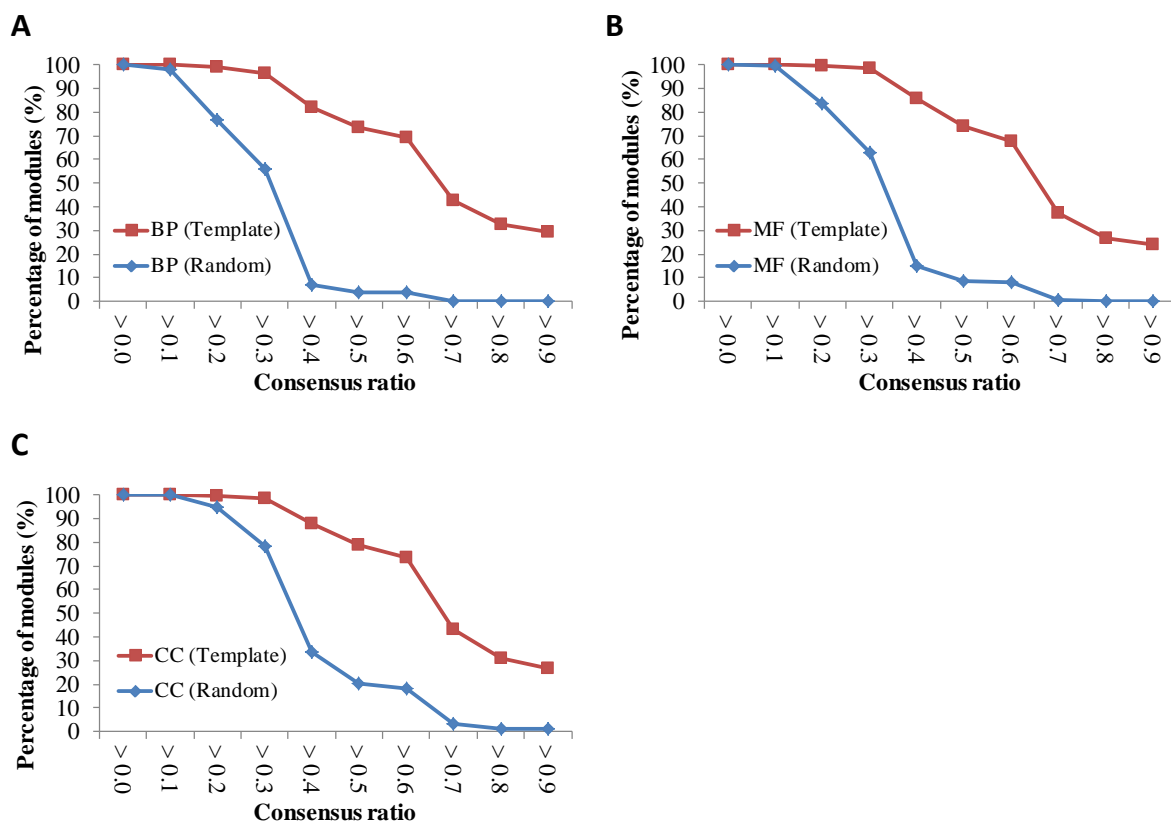
接著要註解模組家族的功能，將從資料庫取出的 1,975 模板模組當作正集合(positive sets)，而負集合(negative sets)的作法是將這 1,975 個模板模組中各個模板模組以隨機亂數的蛋白質產生出 50 個與模板模組相同的蛋白質數模組，所以共產生出 98,750(1,975×50)個亂數模組，接著計算出每個模組裡面每個 GO term(例:生化途徑;BP、細胞位置;CC、分子功能;MF)的一致性比例(consensus ratio ;CRM)，觀察出模板模組的一致性比例是很顯著的優於亂數模組的一致性比例(圖八)，由圖八觀察出模板模組的一致性比例(例:BP、CC、MF)有約 70%(> 1,300)的模板模組大於 0.6，相反的，在亂數模組中大於 0.6 的一致性比例分別只有 3.9%(BP)、8.1%(MF)、18.3%(CC)，且我們也將模板模組中一致性比例大於 0.6 的所有 GO term 進一步去計算超幾何分佈( $p\text{-values}$ )，共有 88.2%(7,776/8,819)的 GO term  $p\text{-values}$  值是小於 0.05 的，意味著模板模組中一致性比例大於 0.6 的 GO term 是具有顯著性的，因此可以用來代表這些模板模組的註解。

接下來為了瞭解模板模組與建構出的模組家族是否執行相似的生物功能，我們使用第二章所闡述的 agreement ratio(AR)以及模組家族一致性的比例(CRF)，分別計算每桶模組家族的 BP、MF、CC 與此桶模板模組的關聯性(圖八)，由此圖可以看出當 CRF 數值高於 0.6 時，AR 數值也是同樣具有很高的一致性，分別在 BP(0.68,綠色)、CC(0.79,藍色)、MF(0.79,紅色)。舉例: TRAP-SMCC mediator 模組家族共有 7 個同源模組其中 CRF

大於 0.6 的 GO terms 分別有 transcription( $CRF = 0.83$  and adjusted  $p$ -value =  $4.59e-08$ )、RNA polymerase II transcription mediator activity ( $CRF = 1.00$  and adjusted  $p$ -value =  $1.41e-11$ )、mediator complex ( $CRF = 1.00$  and adjusted  $p$ -value =  $1.42e-05$ )，這三個具有代表性的 GO terms 可以有效的被用來註解此 TRAP-SMCC mediator 模板模組，這結果顯現同源網路模組家族網站可以為使用者所查詢的模組(例:模板模組)提供高度一致性的 GO terms 以及所查詢模組的同源模組。此外，模板模組與同源模組可以藉由我們的方法推導出他們具有的特定生物功能。

基於這些結果，我們從哺乳動物蛋白質複合物資料庫(CORUM)共 1,785 個模版模組以及蛋白質資料銀行共 1,252 個模版模組(PDB)，將這兩個資料來源整合去除重覆的模組後共保留 1,975 個模組，並且透過實驗室先前所研究以 290,137 條序列為基礎所建構的蛋白質交互作用家族以及以 86,252 個結晶結構為基礎所建構的複合體家族，因此定義出 1,975 個模組家族，此模組家族共在 1,679 個物種中找到 53,529 個模組。





圖八、GO 生化功能註解之一致性比例分析

此圖為 1,975 個模板模組(紅色)與 98,750 個隨機模組(藍色)分別在 GO (A) BP、(B) MF 及(C) CC 註解上之一致性比例(CRM)佔有率分佈圖，可以從圖表看出模板模組的 CRM 都高於隨機模組的 CRM，此外，無論是模板模組的 BP、MF 或 CC，都有約 70%(>1,300) 模組模板具有  $CRM > 0.6$ ，相反地，在隨機模組中， $CRM > 0.6$  僅只有 3.9% (BP)、8.1% (MF)、18.3% (CC)，因此，我們將模組代表性的 GO terms 定義其條件必須符合  $CRM > 0.6$  及統計上顯著的出現比例(adjusted  $p$ -value of GO term  $\leq 0.05$ )。

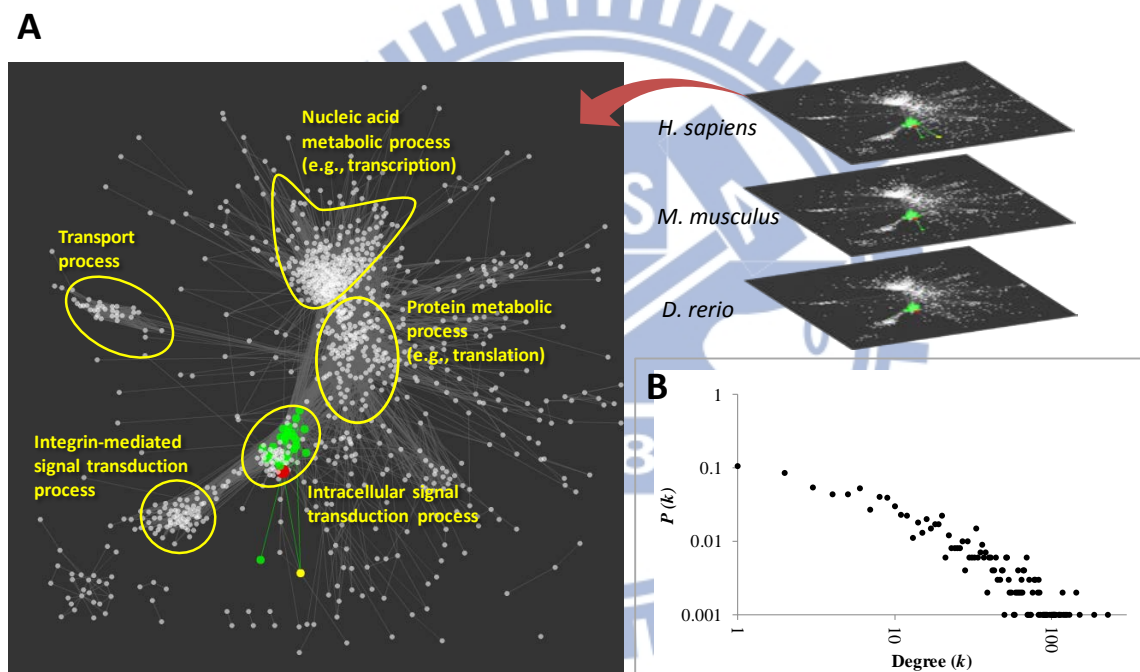
## 二、模組交互作用網路的合理性

對於建構出來的跨物種模組交互作用網路，要去評估此網路的屬性和是否擁有生物意義(圖九~十三)，在評估此網路的屬性中，我們是去計算  $P(k)$ ， $P(k)$ :代表擁有  $k$  條邊的模組占有所有模組的機率，藉以分別導出人類、老鼠、斑馬魚的模組交互作用網路是否屬於 scale-free 的網路特徵，從圖九、圖十 A、圖十 B 中我們可以看到隨著模組的分支數變多，而  $P(k)$  值也隨之的下降，說明了建構出的模組交互作用網路中分支數越高的模組占整體網路中是少數的，往往執行生物途徑中都會有相同的模組參與執行過程(例:負責在細胞內傳遞訊息的模組; Frs2-Grb2-Shp2 module)，這種節點(模組、蛋白質)通常都在生物體中扮演著樞紐的角色(分支度相較於其他的節點來的高)，有文獻指出此種分佈為 scale-free 的網路特徵，而此種特徵最適合解釋細胞網路的架構。文獻進一步指出 scale-free 網路可以藉由分支指數  $\gamma$  值，去判定出是否為 scale-free 網路類型，我們在所建構出三個物種的模組網路中分別算出分支指數(degree exponent) $\gamma$  值，在人類分支指數  $\gamma$  值為 1.183、在老鼠分支指數  $\gamma$  值為 1.143、而在斑馬魚分支指數  $\gamma$  值為 1.218，這些結果與目前一些文獻所研究的生物體細胞網路[39, 40]以及其他文獻所探討酵母菌的模組交互作用網路[41, 42] (圖十 C) 的架構是呈現一致相符的。而在 scale-free 網路中大部分研究將其分支指數  $\gamma$  值定義在  $2 \leq \gamma \leq 3$ ，但也有文獻去指出分支指數  $\gamma < 2$  也是屬於 scale-free 網路[39, 40]，此外，我們也去計算出 3 個模組交互作用網路的模組分支(Degree)中位數值(Median)，分別在人類以及老鼠的中位數值都為 10，而在斑馬魚中位數值為 11，藉由同源網路模組家族網站，使用者可以得知所查詢的模組是否具有高度的分支數，因為高度的分支數通常在生物體網路中是扮演著關鍵作用的角色，譬如在我們所建構的模組交互作用網路中 JAK2-PAFR-TYK2 此模組的分支數為 25，據我們所知此模組在細胞膜與外界接收傳遞訊號中扮演著重要的角色。

我們將所建構的模組交互作用網路分成五個主要的細胞生化反應過程(圖九 A)，這五個分別是 nucleic acid metabolic process(例: 轉錄;transcription)、 protein metabolic process (例:轉譯:translation)、intracellular signal transduction process、integrin-mediated signal transduction process、transport process。藉由模組家族為每個模組的註解，因此可以分析出五個區域的特性，可以觀察出有兩個區域都是在執行基本的細胞代謝過程，而這些過程都是屬於生物體內核心的生化過程(例: central dogma)，譬如執行轉錄的核酸代謝過程以及轉譯的蛋白質代謝過程都有在我們建構出的模組交互作用網路中註解出。而剩餘有被我們註解的區域，分別是在細胞膜上的訊息傳遞以及在細胞質裡扮演傳遞過程的區域，從圖九 A 中很明顯這兩個區域都圍繞在模組交互作用網路的周邊，且都與核心的生化過程有著密切的連接(圖九 A、圖十一)。

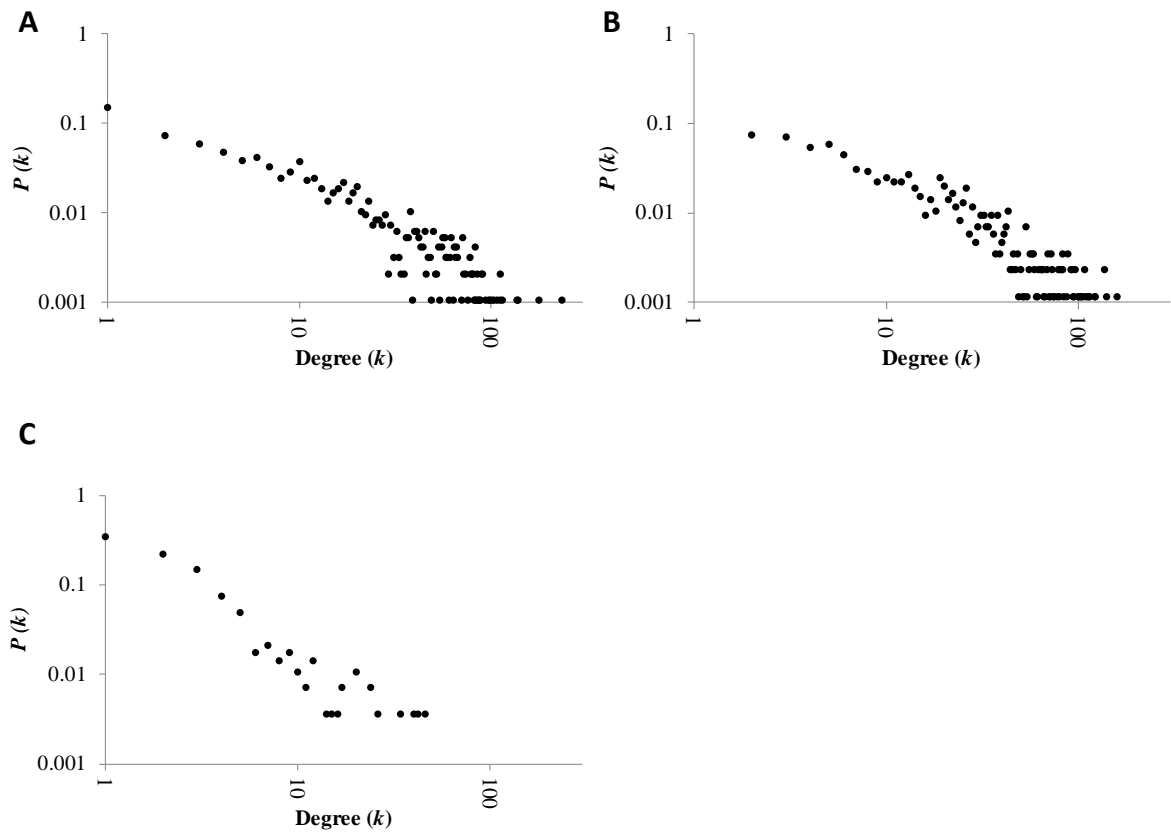
透過模組交互作用網路，我們可以用來分析當細胞皆受到刺激而產生的增值行為，當生物體受傷，體內會啟動修補機制，因此釋放出細胞外基質(Extracellular matrix;ECM)蛋白質和生長因子(growth factors)(圖十二、圖十三)，當細胞增生受到細胞外基質蛋白與整合素受體的共同調節(例: ITGA5-ITGB1-FN1-TGM2 模組負責執行整合素訊號傳遞的過程)和成長因子會與受體結合(例: JAK2-PAFR-TYK2 模組是負責在細胞內接受成長因子的訊息傳遞過程)[43]，這兩個模組接收到啟動因子以後都會去觸發在細胞質內的 Frs2-Grb2-Shp2 模組，進而 Frs2-Grb2-Shp2 模組接收到訊號會去判斷該將訊號送往哪個目的地以及要執行的功能，此模組判定接收到的訊號是要進行細胞增生的動作，因此將訊號傳導至細胞核內的 ALL-1 supercomplex 模組，此模組最主要是執行核酸代謝過程，並且此模組會與鄰近的模組一起執行，在此圖例則列舉出兩個模組分別是 RNA polymerase II 模組以及 TRAP-SMCC mediator 模組一起執行轉錄以及轉譯的功能，進行核酸代謝過程作用後，因此產生出新合成的蛋白質，將此新合成的蛋白質送往內質網讓 kinase maturation module 1 模組(此模組主要執行蛋白質代謝過程)執行後修飾的動作，蛋白質合成完成後，會透過 SNARE 模組(此模組主要執行運輸過程)將蛋白質送往目的地。

從上述的範例，我們可知同源網路模組家族網站可以輔助使用者在網路生物學中探索細胞過程中一個新的視野，同源網路模組家族網站不僅僅能提供單一物種模組網路新視野，在後面的例子分析中我們進而藉由 JAK2-PAFR-TYK2 模組以及 TRAP-SMCC mediator 模組去探討跨物種觀察模組與模組交互作用可以因此為使用者帶來什麼助益。



圖九、模組交互作用網路生化途徑剖析圖

(A)我們將人類、老鼠、斑馬魚的模組交互作用網路大致粗分成五大細胞生化過程。  
 (B)計算於人類模組交互作用網路的節點分支分佈圖，由此圖可以看出人類模組網路是屬於 scale-free 網路特性。

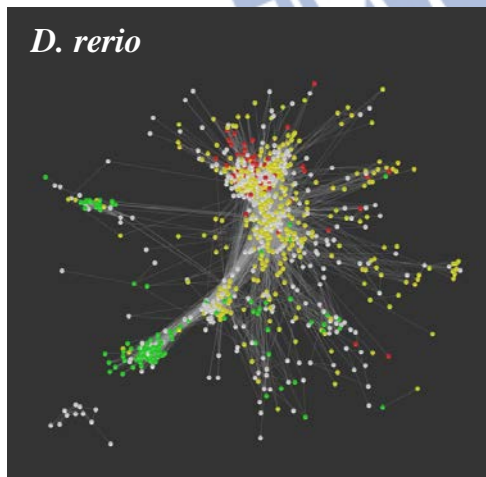
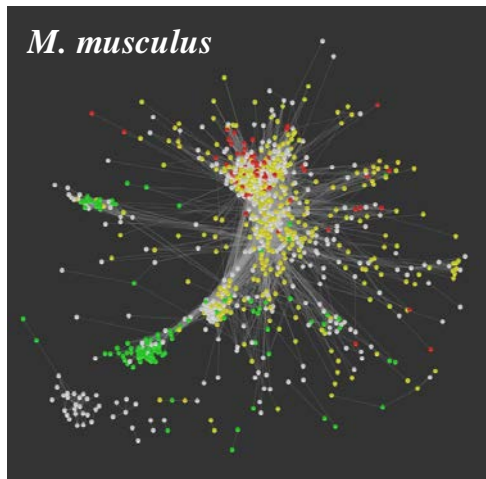
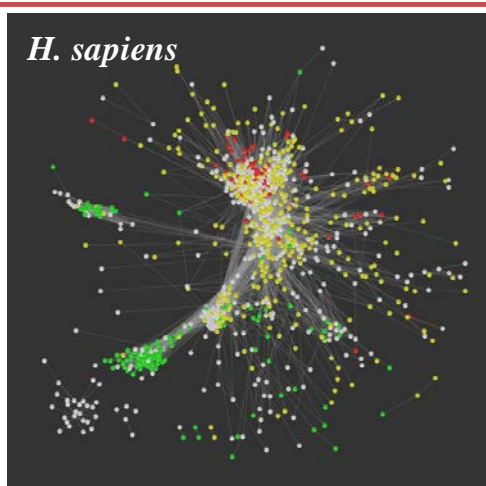


圖十、模組交互作用網路節點分支分佈圖

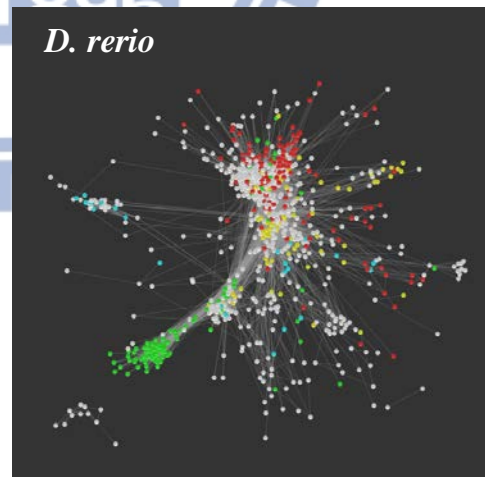
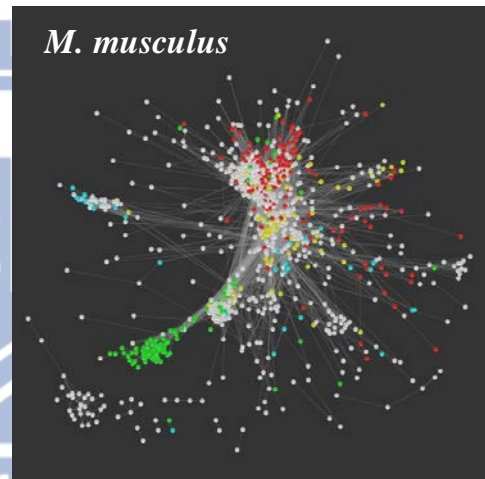
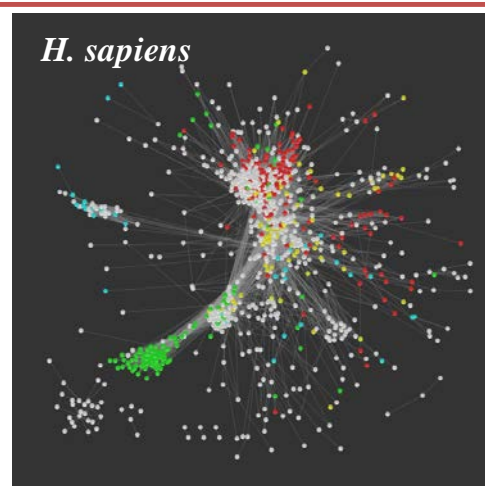
(A)老鼠的模組交互作用網路節點分支分佈圖。

(B)斑馬魚的模組交互作用網路節點分支分佈圖。

(C)酵母菌的模組交互作用網路節點分支分佈圖，此圖文獻來自 Wang's study (2009)[42]。從這些圖與現有的文獻看出我們所建置的模組交互作用網路是屬於 scale-free 網路特性。

**A**

- Nuclear part (GO:0044428)
- Intracellular (GO:0005622)
- Membrane (GO:0016020)
- Others

**B**

- Nucleobase-containing compound metabolic process (GO:0006139)
- Protein metabolic process (GO:0019538)
- Signal transduction (GO:0007165)
- Transport (GO:0006810)
- Others



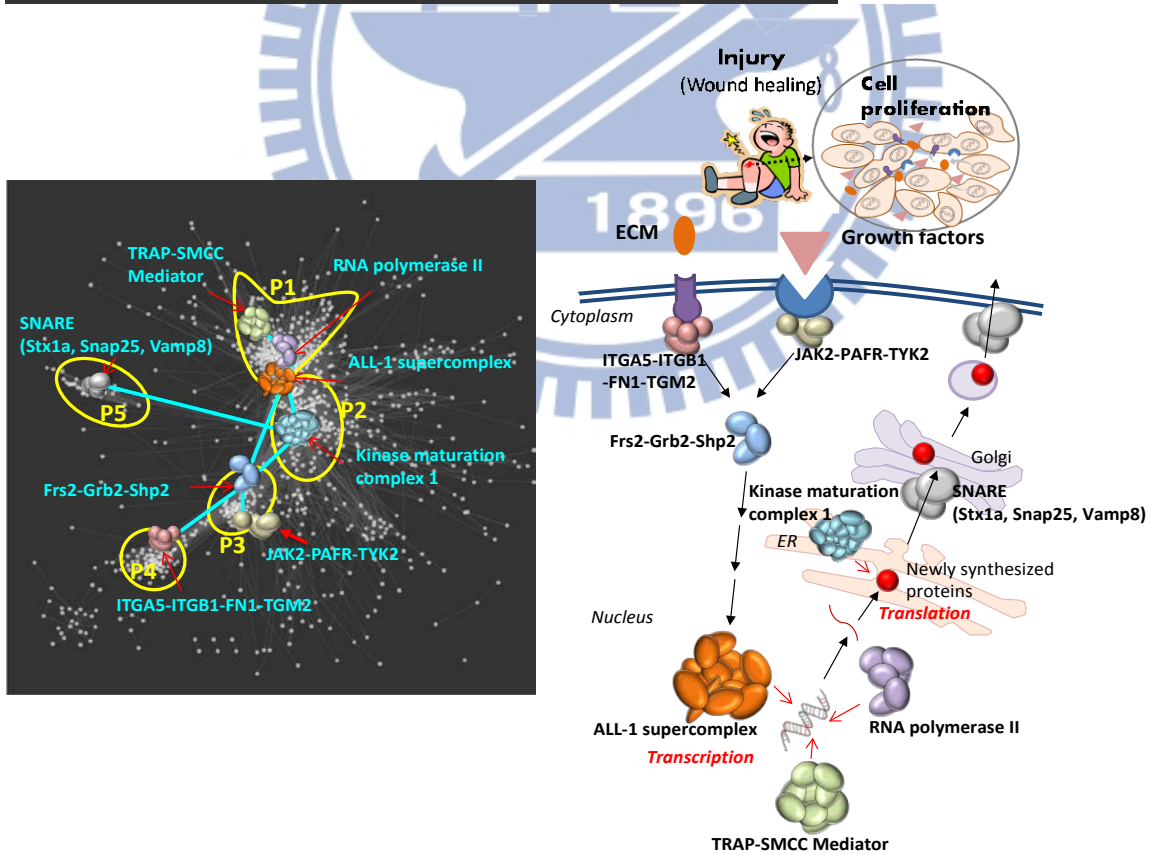
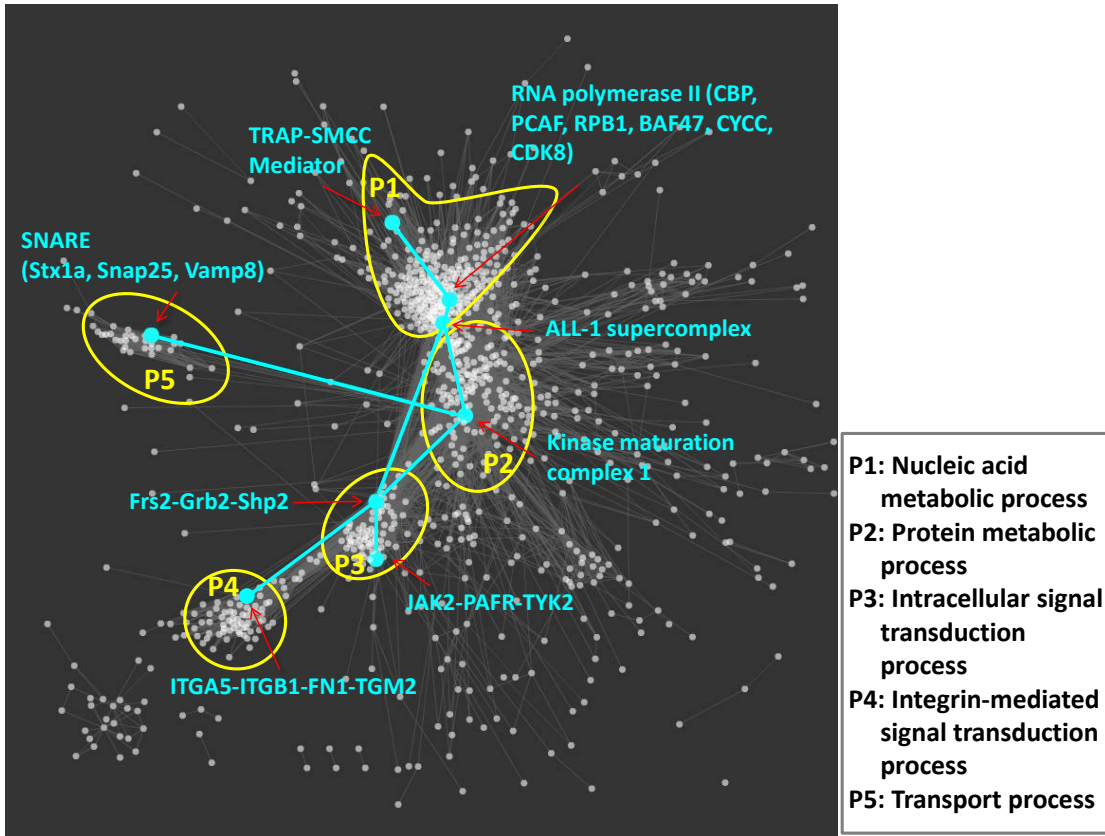
### 圖十一、GO terms(BP、CC)在人類、老鼠、斑馬魚的模組交互作用網路註解分析

使用 GO terms 的註解探討模組交互作用的網路特性[18]。

(A)根據 GO 的細胞位置(cellular component; CC)對各個模組家族進行註解，可以將模組交互作用網路大致區分為四大區塊，分別有 68 個模組家族註解為 nuclear (GO:0044428, 紅色)、402 個模組家族註解為 intracellular (GO: 0005622, 黃色)、141 個模組家族註解為 membrane (GO: 0016020, 綠色)、剩餘註解為 Others (灰色)。

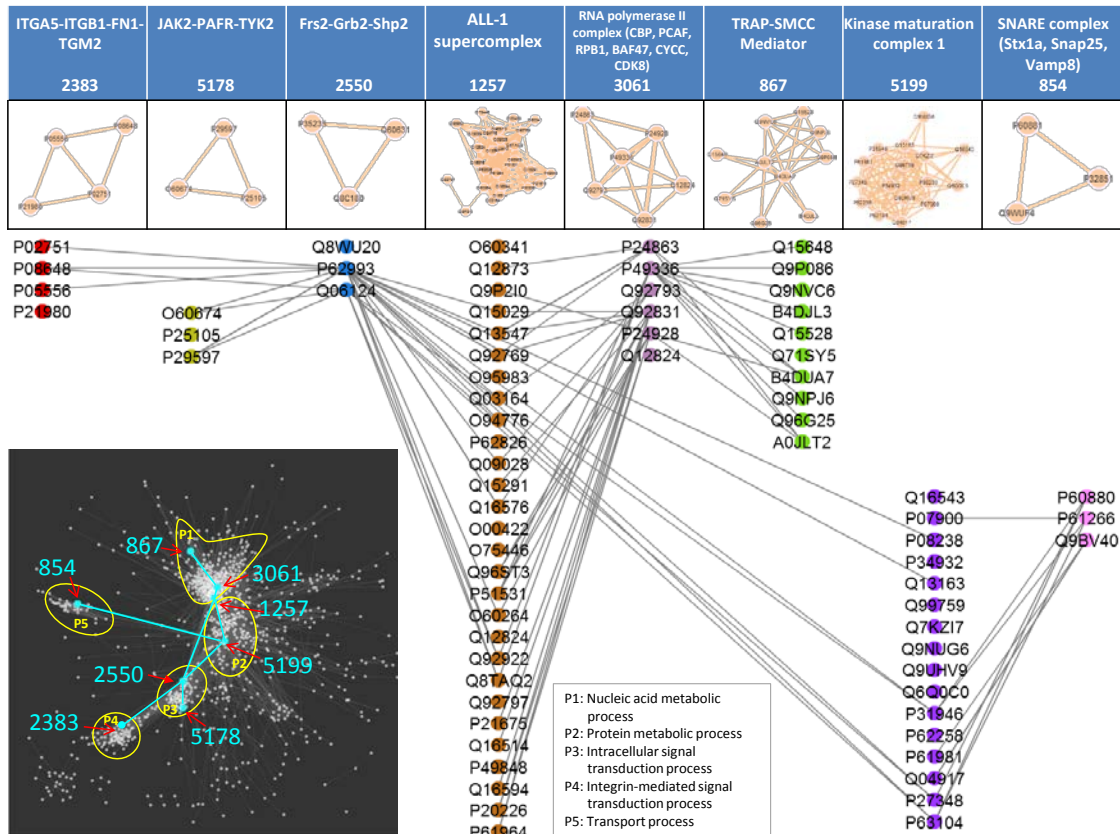
(B) 根據 GO 的生化功能(biological processes; BP)對各個模組家族進行註解，分別有 154 個模組家族註解為 nucleobase-containing compound metabolic process (例: 轉錄作用)(GO: 0006139, 紅色)、83 個模組家族註解為 protein metabolic process (例: 轉譯作用)(GO: 0019538, 黃色)、126 個模組家族註解為 signal transduction (GO: 0007165, 綠色)、37 個模組家族註解為 transport (GO: 0006810, 青色)。

我們使用了 Jaccard coefficient (JC)方法來測量兩個模組之間的相似度，假設有兩個模組蛋白質個數的分別為  $M_1$  和  $M_2$ ，而 JC 定義如下： $JC = \frac{|M_1 \cap M_2|}{|M_1 \cup M_2|}$  [42]，若 JC 值大於 0.6，則將兩模組匯集成同一群集，而每一個群集的代表性模組則選擇蛋白質數量最高的模組，因此，我們分別在人類挑選出 1,278 個代表性模組、老鼠挑選出 1,251 個代表性模組、而在斑馬魚挑選出 1,100 個代表性模組。



圖十二、透過模組交互作用網路描述細胞增殖的行為

細胞的增殖行為可以透過五個主要的細胞生化過程(P1~P5)中的8個模組(青色節點)進行探討,這8個模組分別為 ITGA5-ITGB1-FN1-TGM2、JAK2-PAFR-TYK2、Frs2-Grb2-Shp2、ALL-1 supercomplex、RNA polymerase II complex (CBP, PCAF, RPB1, BAF47, CYCC, CDK8)、TRAP-SMCC Mediator、kinase maturation complex 1、及 SNARE (Stx1a, Snap25, Vamp8)。



圖十三、細胞增殖行為中 8 個模組間的詳細蛋白質交互作用組成

這 8 個模組分別為 ITGA5-ITGB1-FN1-TGM2、JAK2-PAFR-TYK2、Frs2-Grb2-Shp2、ALL-1 supercomplex、RNA polymerase II (CBP, PCAF, RPB1, BAF47, CYCC, CDK8)、TRAP-SMCC Mediator、kinase maturation complex 1、及 SNARE (Stx1a, Snap25, Vamp8), 其位於五個主要的細胞生化過程區塊中(黃色線框區域)。詳細紀錄其模組內部與模組之間的蛋白質交互作用組成。

### 三、範例分析

#### (一)、JAK2-PAFR-TYK2 模組

##### 1.JAK2-PAFR-TYK2 模組家族

圖三是展現出同源網路模組家族網站使用流程，當使用者在同源網路模組家族網站查詢欄位中輸入 Janus kinase 2 (Jak2, UniProt 編號: Q62120)、platelet-activating factor receptor (Pafr 或 Ptafr, UniProt 編號: Q62035)、tyrosine kinase 2 (Tyk2, UniProt 編號:Q3U447) (圖三 B)，同源網路模組家族網站為這三個蛋白質找尋到多個模板模組，其中 JAK2-PAFR-TYK2 模板模組符合所查詢的蛋白質比例為最高(圖十三 B)，我們為 JAK2-PAFR-TYK2 模組(CORUM ID: 5178)找尋到十個同源模組，其中的物種包含有人類、老鼠、斑馬魚、果蠅(圖三 D)，已經有文獻指出 JAK2-PAFR-TYK2 模組負責調控著不同的生理和病理，在人類皮膚中對於執行先天性免疫反應扮演著重要角色[44, 45]，在此模組裡 JAK2 和 TYK2 兩個蛋白質是非受體酪氨酸 JAK 家族，且對於哺乳動物的演化以及免疫系統疾病有著息息相關的密切關係[44, 46]，Monetfamily 建構出 JAK2-PAFR-TYK2 模組家族，並也為此家族註解 GO terms(例:分子功能(MF) - non-membrane spanning protein tyrosine kinase activity 和細胞位置(CC) - cytoskeleton)，此外，Monetfamily 也會提供此模組家族的鄰近模組 GO terms 的註解，有助於使用者去觀察了解此模組還跟那些模組有交互作用，互相調控那些生化功能和共同執行了那些生化過程，以此 JAK2-PAFR-TYK2 模組家族為例，同源網路模組家族網站為此模組的鄰近模組註解出代表性的 GO terms 分別有:生物途徑(BP) - epidermal growth factor receptor (EGFR) 訊息途徑、分子功能(MF) -信號傳導活動、細胞位置(CC)-細胞質(圖三 C)。

## 2. 模組與模組交互作用和鄰近的模組

在人類模組交互作用網路中，JAK2-PAFR-TYK2 模組(紅色節點)，擁有 25 個鄰近模組(綠色節點)，這個子網路共包含 26 個模組以及 131 條模組與模組交互作用(圖十四 C 和表二)，根據 GO term 的註解和 MIPS FunCat [47] 的分析，這些模組可以被大致分為三個群體，分別是 cell surface receptor linked signaling pathway (橘色)、cellular protein metabolic pathway (紫色)、interleukine receptor signaling pathway (藍色)。JAK2-PAFR-TYK2 模組是一個具有高度連通性的樞紐，負責在細胞與細胞間以及細胞內的訊息傳遞(圖十四 A)，我們可以藉由圖十四 A 觀察 JAK2-PAFR-TYK2 模組與其鄰近模組在三個物種上的差別，橘色括號和紫色括號在三個物種中都具有高度的保留，這兩個群體所執行的生物功能分別是蛋白質代謝和訊息傳遞，可看出都是生物體所必需的生物功能，但在藍色括號中就可明顯觀察出各個物種不同的地方，例：IL4-IL4R-IL2RG (CORUM ID: 1515) 模組和 RIN1-STAM2-EGFR (CORUM ID: 3678) 模組以及 IL-6/IL-6R $\alpha$ /gp130 (PDB code: 1p9m) 在斑馬魚中無法找到這些模組。藍色括號主要功能是在執行先天性免疫反應的生化過程，不同的物種對於不同的疾病會有不同的反應，例：愛滋病會感染人類，而斑馬魚並不會被愛滋病所感染。所以有可能導致各個物種在藍色括號中有差異的原因之一。

我們進而探討在其他物種中為什麼我們會沒有找到此模組，挑選這 3 個物種在人類中有出現的 IL-6/IL-6R $\alpha$ /gp130 (PDB code: 1p9m) 模板模組來做進一步探討，1p9m 此模組裡共有 3 個蛋白質分別是 IL-6、IL-6R $\alpha$ 、gp130，而此模組是參與免疫調控機制反應，根據我們蛋白質結晶結構家族先前的研究，觀察出此模組在人類中的接合模型分別是 IL-6/IL-6R $\alpha$ /gp130 與老鼠和斑馬魚有很顯著的不同，因而導致在這兩個物種中無法找到此模組，最主要的原因有二項：

1) 從 gp130 和 IL-6 兩個蛋白質中分別觀察老鼠和人類的介面(interface)的 contact-residue (有顏色的部分)中，在老鼠對應於 IL-6 蛋白質最相像的序列中只占 7.7%，

而在斑馬魚占 0%，由此觀察他們的比例是非常低的，因此導致沒有辦法在老鼠和斑馬魚中找尋到 gp130 和 IL-6 的 binding model(圖十五)。

2) 我們使用模板為基礎的計分方式來計算 1p9m 這個模組內 gp130 和 IL-6 這兩個蛋白質對的介面相似度，在老鼠得到的 Z-values 為 0.923、在斑馬魚得到的 Z-values 為 -1.638，這分數沒有通過臨界值導致這對蛋白質的交互作用在老鼠和斑馬魚中我們認為不存在

### 3.治療牛皮癬的潛在藥物靶標

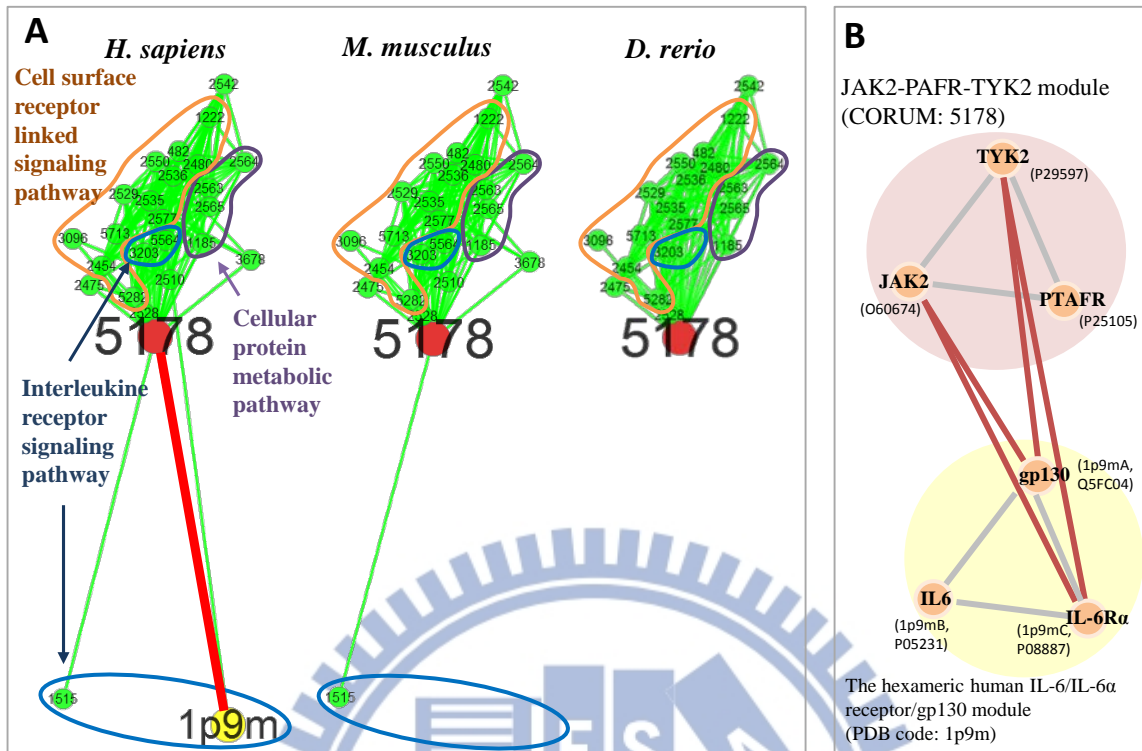
牛皮癬(psoriasis)是一種自身免疫性疾病，在現今是人類最常見的皮膚疾病之一[48]，而 JAK2-PAFR-TYK2 此模組已經有研究指出它與牛皮癬的發生有著密切的關係，所以目前在此模組裡面的 JAK2 和 TYK2 已經被提議作為設計牛皮癬藥物的潛在目標，如：ruxolitinib[49]和 tasocitinib[50]這兩種藥物分別用來對 JAK2 和 TYK2 進行抑制，進而減緩牛皮癬的症狀。

有趣的是我們可以透過同源網路模組家族網站去觀察 JAK2-PAFR-TYK2 模組與鄰近模組的關聯，因而提供對於牛皮癬症狀發生的新觀點及新線索，在 JAK2-PAFR-TYK2 模組鄰近的 25 個模組中，共有 12 個模組在 3 個物種都具有高度保留，且這 12 個模組都被註解為細胞表面受體相關的訊息傳遞途徑，例如：SH3P2/OSTF1-CBL-SRC module、SLP-76-Cbl-Grb2-Shc module、Fc receptor gamma-R1 stimulated 以及 CAS-SRC-FAK module 的模組家族都被註解為表皮生長因子受體訊息傳遞途徑 (Epidermal growth factor receptor signaling pathway;EGFR)。牛皮癬發生的原因比較複雜，藉由同源網路模組家族網站提供這些資訊我們因此可以去推斷可能跟免疫系統上訊息傳遞功能紊亂有關聯。

JAK2-PAFR-TYK2 模組(JAK2/PAFR/TYK2)和 1p9m 模組(IL-6/IL6R $\alpha$ /gp130)之間擁有的交互作用分別是 JAK2-gp130、JAK2-IL-6R $\alpha$ 、TYK2-gp130、TYK2-IL-6R $\alpha$ (圖十四 B)。基於超幾何分布的計算結果，這兩個模組之間交互作用的 *p*-value 藉此算出是 1.06e-5，明顯的通過我們的定義。因為我們判定兩個模組有交互作用的關係，因此在 1p9m 模組內 IL-6 蛋白質也被註解為 JAK-STAT 訊息傳遞途徑，目前有文獻研究出當人

體感染了愛滋病(HIV)後，體內的 IL-6 蛋白質的反應也會隨之遞增，而 IL-6 蛋白質是包含在 1p9m 模組裡，因而也會造成 1p9m 模組大量反應，在我們的研究中我們認為 1p9m 模組與 JAK2-PAFR-TYK2 模組有交互作用的產生，所以 JAK2-PAFR-TYK2 模組也會因而有大量的表現。此外，據我們所知 JAK2-PAFR-TYK2 模組同負責調節表皮生長因子受體訊息傳遞途徑(EGFR)和白細胞介素受體訊息傳遞途徑，因為 JAK2-PAFR-TYK2 模組大量的表現，因而導致功能的不正常，而造成牛皮癬的症狀出現可能之一，而人體感染愛滋病後通常會有牛皮癬症狀的發生，因此透過我們的研究觀察到 JAK2-IL6 是否有可能為控制愛滋病與牛皮癬的潛在目標之一，上述的觀察結果藉由 JAK2-PAFR-TYK2 模組和其鄰近模組的交互作用以及模組交互作用網路，提供了 JAK-引發了炎症性相關疾病 (例:牛皮癬和類風溼關節炎)獨特的寶貴見解。





圖十四、在人類、老鼠、斑馬魚的 JAK2-PAFR-TYK2 模組與其鄰近模組

同源網路模組家族網站可提供使用者觀察目標模組與其鄰近模組在人類、老鼠、斑馬魚中的差異。(A)JAK2-PAFR-TYK2 模組(CORUM ID: 5178)與 25 個鄰近模組所構成之鄰近模組子網路，藉由跨物種的比較觀察，可將其大致區分為 3 個區塊，分別是 cell surface receptor linked signaling pathway (橙色)、cellular protein metabolic pathway (紫色)、以及 interleukine receptor signaling pathway (藍色)。其鄰近模組分別對 signal transducer activity 和 cytosol 這 2 個 GO terms 具有高度的一致性。(B)呈現 JAK2-PAFR-TYK2 模組與 the hexameric human IL-6/IL-6 $\alpha$  receptor/gp130 (PDB code: 1p9m)模組之間詳細的交互作用關係，紅色線條表示兩模組之間實際的蛋白質交互作用註解，灰色線條表示模組內部蛋白質之間的交互作用註解。



## A

Query : Chain B of 1p9m (IL6 of *H. sapiens*)  
Sbjct : IL6 of *M. musculus* (P08505)  
Length = 211

Contact-residue identity : 7.7% (1/13)  
Sequence identity : 40.4% (74/183)

Score = 151 bits (382), Expect = 2e-35, Method: Compositional matrix adjust.  
Identities = 74/183 (40%), Positives = 122/183 (66%), Gaps = 3/183 (1%)

Query: 4 VPPGEDSKDVAAPHRQPLTSSERIDKQIRYILGISALRKETCNKSNMCESSKEALAENN 63  
V G+ ++D P+R P+ ++ ++ I ++L I +RKE CN ++ C ++ +ALAENN  
Sbjct: 30 VRRGDFTEd-TTPNR-PVYTTSQVGGGLITHVLWEIVEMRKELCNGNSDCMNNDDALAENN 87

Query: 64 LNLPKMAEKDGCQSGFNEETCLVKIITGLLEFEVYLEYLQNRf-ESSEEQARAVQMSTK 122  
L LP++ DGC+Q+G+N+E CL+KI +GLLE+ YLEY++N ++ +++AR +Q T+  
Sbjct: 88 LKLPEIQRNDGQYQTGYNQEI CLLKISSGLLEYHSYLEYMKNNLKDKNKKDKARVLQRDTE 147

Query: 123 VLIQFLQKKAKNLDAITTPDPTNASLLTKLQAQNQWLQDMTTHLILRSFKEFLQSSLRA 182  
LI ++ K+L I P P +NA L KL++Q +WL+ T IL+S +EFL+ +LR+  
Sbjct: 148 TLIHIFNQEVKDLHKIVLPTPISNALLTDKLESQKEWLRRTKTIQFILKSLEEFKVTLSR 207

Query: 183 LRQ 185  
RQ  
Sbjct: 208 TRQ 210

### Contact residues

Hydrogen-bond residues: green

Others : gray

## B

Query : Chain B of 1p9m (IL6 of *H. sapiens*)  
Sbjct : IL6 of *D. rerio* (GenBank: AFC76325.1)  
Length = 231

Contact-residue identity : 0% (0/13)  
Sequence identity : 23.6% (29/123)

Score = 36.2 bits (82), Expect = 1.1, Method: Compositional matrix adjust.  
Identities = 29/123 (24%), Positives = 56/123 (46%), Gaps = 7/123 (6%)

Query 64 LNLPKMAEKDGCQSGFNEETCLVKIITGLLEFEVYLEYLQNRfESSEEQARAVQMSTKV 123  
++ P + D C F+ E CL +I + L ++ Y++ +S ++ STK  
Sbjct 95 ISTPLLKPSDRCLSKNFSTERCLTRIYSVLTWYKDNWNYIEKENLTS-VLVNDIKHSTKR 153

Query 124 LIQFLQKKAKNLDAITTPDPTNASLLTKLQAQNQWLQDMTTHLILRSFKEFLQSSLRAL 183  
L++ + + + D D T++ T L ++ W + T H IL +F + + RA+  
Sbjct 154 LLEAINSQQLQVRDG--EMDQTSS----TSLSFKSAWTRKTTVHSILFNFSVMIDACRAI 207

Query 184 RQM 186  
M  
Sbjct 208 NYM 210

### Contact residues

Hydrogen-bond residues: green

Others : gray

圖十五、人類 IL-6 蛋白質(chain B of 1p9m)與老鼠、斑馬魚之同源蛋白質的序列比對圖

以 BLASTP 進行人類 IL-6 蛋白質與(A)老鼠或(B)斑馬魚之同源蛋白質中的序列比對，並標示其結合界面(interface)上的 contact residues。

## (二)、TRAP-SMCC mediator 模組

TRAP-SMCC mediator 模組是轉錄機制中的中樞調控子[51, 52]，Mediator of RNA polymerase II transcription subunit 19 (Med19)，此蛋白質是 TRAP-SMCC mediator 模組內的成員之一，對於肺腫瘤[52]以及乳癌[53]都有所影響。我們在同源網路模組家族網站輸入此蛋白質，可以因此找尋到 TRAP-SMCC mediator 模組，且此模組找尋到 7 個同源模組，TRAP-SMCC mediator 模組註解了 3 個 GO terms，分別是 transcription ( $CRF=0.83$ )、RNA polymerase II transcription mediator activity ( $CRF=1.00$ )、和 mediator complex ( $CRF=1.00$ ) (圖十六 B)，有趣的是，此模組的 15 個鄰近模組(綠色節點)所註解的 GO term，也與 TRAP-SMCC mediator 模組家族具有高度的一致性，這結果顯示我們的網站可以利用模組家族和其相鄰的模組賦予的 GO terms 的註解為使用者所查詢的蛋白質預測此蛋白質的細胞功能。

在人類與老鼠的模組交互作用網路中，TRAP-SMCC mediator 模組(紅色節點)共有 15 個鄰近模組，而這 16 個模組中共有 57 條模組交互作用(圖十六 B)，根據 GO terms 和 MIPS FunCat 的分析，這些模組互相動態的調節執行轉錄作用，大致可以分成 3 個區塊，這 3 個區塊分別是 transcription activation、transcription repression、DNA conformation modification (例:chromatin structure modification) (圖十七)。從我們的模組交互作用網路中觀察到在 TRAP-SMCC mediator 模組中的 MED19 蛋白質和 mediator of RNA polymerase II transcription subunit 29 (IXL or MED29, B4DUA7)蛋白質在哺乳動物中與鄰近的模組都有著高度的交互作用，這結果意味著 MED19 和 MED29 這兩個蛋白質在腫瘤(tumorigenesis)、乳腺癌(breast)、肺癌(lung)、胰臟癌(pancreatic cancers)扮演著關鍵性的角色。

**A**

Input one or multiple  Protein sequence  Gene name/UniProt AC

Enter FASTA sequence (example 1) (example 2)

```
>sp|Q0C1S0|MED19_MOUSE Mediator of RNA polymerase II transcription
#Uniq: 19
MGVPMAGDQPPPPSSALGPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP
MERFALPGAGTDFPPPPSALGPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP
QSQFFPLMELFNGSTELTGFTHLITHEHQATWFCQKRVKELSGFGLFPMHIDISG
SDEKSLKGLLEKPPFLLGGPPFPITVMDKGGPGLATVPLPQGLDMLGQPFKEKRRK
QSPFQDPVFPFPPDSDSXXXXXXXXXXXXPPEPKKXXXXXXXXXXXXXXXXXSSR
SSLR
```

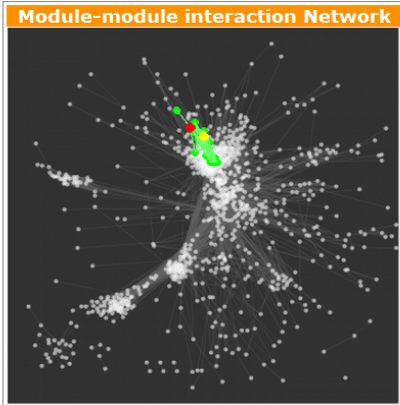
[Quick link of search result](#)

Input one or multiple  Protein sequence  Gene name/UniProt AC

Enter gene name or UniProtKB accession number (example 1) (example 2)

MED19

[Quick link of search result](#)

**B****NetMoFamily**

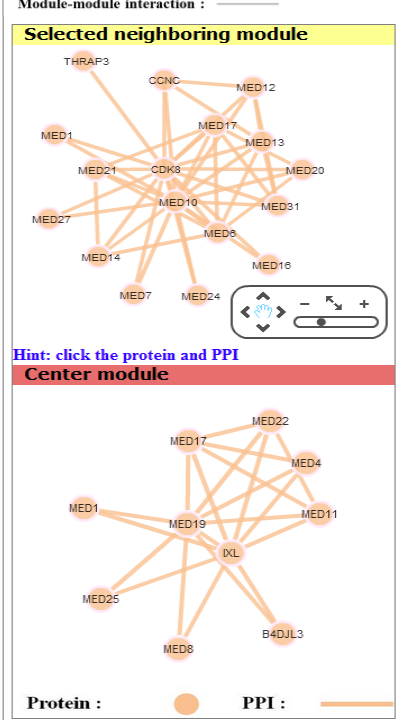
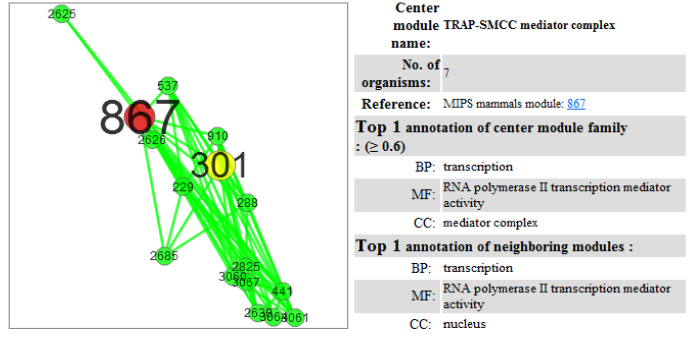
Select species:

Topology description :

Center module : ● Neighboring module : ●

Selected module : ●

Module-module interaction : \_\_\_\_\_

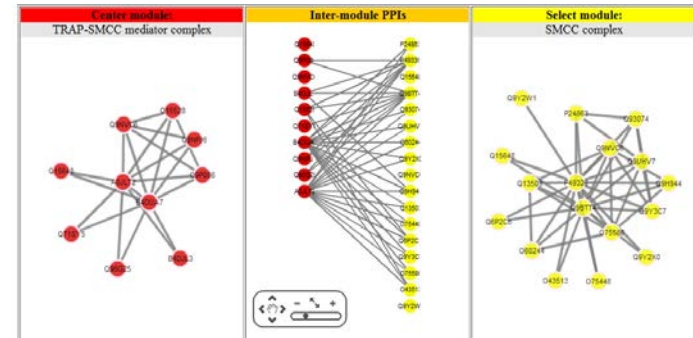
**Summary of neighboring modules****Neighboring modules of TRAP-SMCC mediator complex**

Show	ID	MMI	Name	Template organisms	No. proteins	No. PPIs	No. of Organisms	p-value of MMI
<input checked="" type="radio"/>	301	Topology	SMCC complex	H.sapiens	17	44	12	1.16e-042

**Selected neighboring module annotation**

GO type	GO ID	GO term	Consensus Ratio	Adjusted p-value of GO term
BP	GO:0006350	transcription	0.636	7.32e-07
MF	GO:0016455	RNA polymerase II transcription mediator activity	1.000	1.78e-23
CC	GO:0016592	mediator complex	1.000	2.78e-20

- 288 Topology ARC complex H.sapiens 15 36 12 5.12e-039
- 229 Topology NAT complex H.sapiens 7 14 13 9.16e-036
- 537 Topology Mediator complex M.musculus 8 9 10 1.40e-021
- 2626 Topology CCNC-CDK8-MED1-MED6-MED7 complex H.sapiens 5 4 35 3.45e-019
- 910 Topology CRSP-Mediator 2 complex H.sapiens 10 5 11 3.42e-015
- 2825 Topology BRCA1-RNA polymerase II complex H.sapiens 26 106 53 2.91e-014
- 2625 Topology CDK8-MED6-PARP1 complex H.sapiens 3 2 17 4.72e-014
- 1060 Topology RNA polymerase II complex (RPB1, RAP74, CDK8, CYCC, SRB7, BAF190, BAF47), chromatin structure modifying H.sapiens 7 14 50 6.70e-014
- 2685 Topology RNA polymerase II (RNAPII) H.sapiens 17 56 56 3.80e-013
- 3061 Topology RNA polymerase II complex (CBP, PCAF, RPB1, BAF47, CYCC, CDK8), chromatin structure modifying H.sapiens 6 13 63 3.05e-011

**C****D**

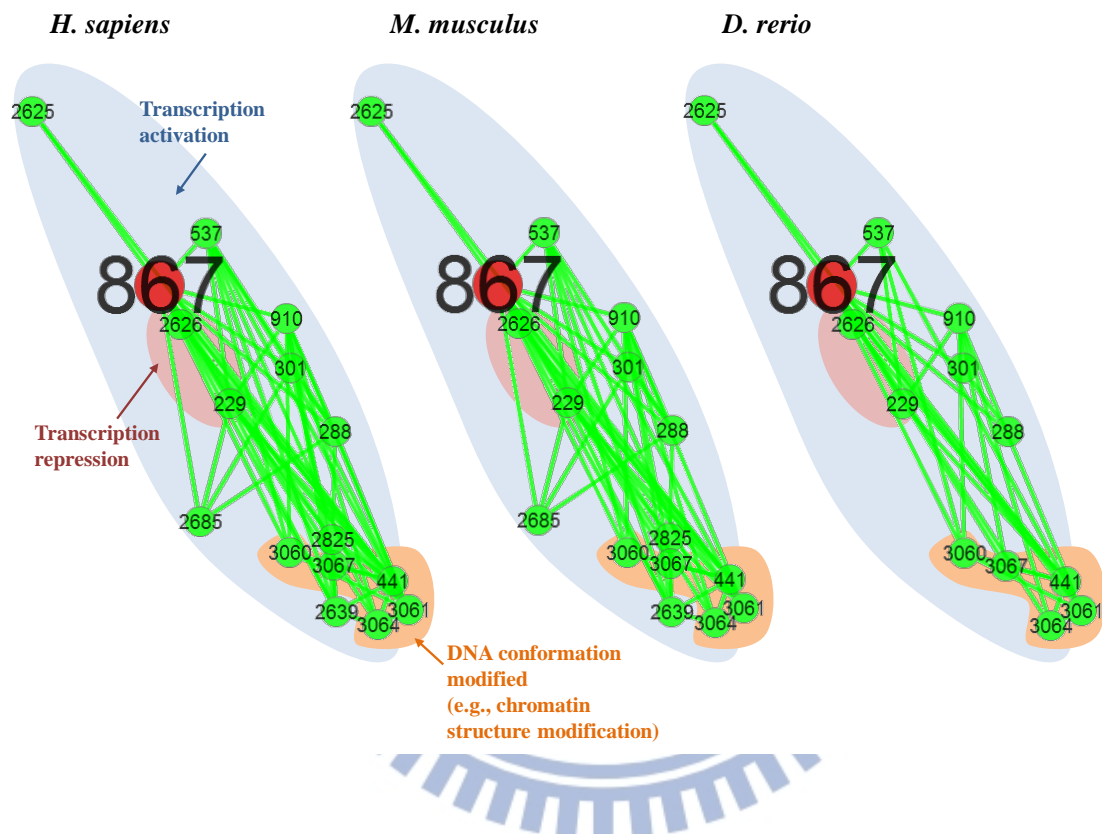
**Homologous modules of TRAP SMCC mediator complex**

Species	Division group	Organism	MED19 (E-value)	MED1 (E-value)	MED21 (E-value)	MED27 (E-value)	DKL (E-value)	MED4 (E-value)	MED11 (E-value)	Hit ratio
●	MAM	Mus musculus	Q0C1S0	Q8Q9G2	Q8Q9G4	Q8Q9G3	Q8Q9G5	Q8Q9G6	Q8Q9G7	1.00 (1-1)
●	MAM	Homo sapiens	A8L7J2	Q89P88	Q89P90	R8D3L3	Q11593	Q11592	R8D3L7	1.00 (1-1)
●	MAM	Bos taurus	A1YVE3	Q2U3L3	Q2U3L2	Q2U3L1	Q2U3L4	A2A3G8	Q2U3L5	1.00 (1-1)
●	VRT	Danio rerio	R6A383	R6A384	A5N2K7	Q893Y1	Q7I3H7	A7V3L3	Q892F8	1.00 (1-1)
●	VRT	Tetraodon nigroviridis	Q585R4	Q585R3	( )	Q585R2	Q585R1	Q585R5	Q585R6	1.00 (1-1)
●	INV	Anopheles gambiae	Q3F3V3	Q3F3V4	( )	Q3F3V2	Q3F3V1	Q3F3V5	Q3F3V6	1.00 (1-1)
●	INV	Aedes aegypti	Q1T9J3	Q1T9J2	Q1T9J1	Q1T9J4	Q1T9J5	Q1T9J6	Q1T9J7	1.00 (1-1)

**PPI Profiles of TRAP-SMCC mediator complex**

### 圖十六、同源網路模組家族網站搜尋 TRAP-SMCC mediator 模組之網頁介紹

使用者在同源網路模組家族網站查詢 mediator of RNA polymerase II transcription subunit 19 (Med19)。(A) 使用者可以輸入蛋白質序列、gene name、UniProtKB。(B) 提供圖像化 TRAP-SMCC mediator 模組、鄰近的模組、鄰近模組交互作用網路、此模組家族及鄰近模組之演化保留性 GO 註解。(C) 提供 TRAP-SMCC mediator 模組與所選擇之鄰近模組(例: SMCC 模組)間詳細蛋白質交互作用組成。(D) TRAP-SMCC mediator 模組的同源模組之蛋白質或蛋白質組成。



### 圖十七、人類、老鼠、斑馬魚之 TRAP-SMCC mediator 模組與鄰近模組

TRAP-SMCC mediator 模組(CORUM ID: 867)與 15 個鄰近模組所構成的鄰近模組子網路，藉由跨物種的觀察比較，可將其大致分成 3 個區塊，分別為 transcription activation (藍色)、transcription repression (紅色)、DNA conformation modification (橙色)。其鄰近模組在 including transcription、RNA polymerase II transcription mediator activity、mediator complex 的 3 個 GO terms 註解上具有相當高的一致性。

## 肆、結語

### 一、總結

我們這研究提出了模組家族以及跨物種模組交互作用網路的新概念，並將此概念藉由網站的呈現讓使用者便於查詢及使用，同源網路模組家族網站透過同源模組去定義出模組之間的交互作用和模組交互作用網路，且此網站服務是第一個提供鄰近模組和藉由跨物種的模組交互作用網路去觀察各個模組交互作用之間的差異性，以及使用者所查詢的模組家族中所有物種的蛋白質與蛋白質蛋白質交互作用的所有資訊，並且會使用 GO terms 為所查詢的模組家族和鄰近的模組作註解。

從我們第三節的結果與討論中，我們證明了這網站服務，讓使用者輸入欲查詢的蛋白質序列或者是基因名稱，此網站服務能提供此蛋白質有出現在哪個模組中，且提供可視化的跨多個脊椎動物的模組交互作用網路，並且使用模組家族和鄰近的模組來為使用者所查詢的蛋白質加以註解。

我們相信同源網路模組家族網站可以提供一個快速搜尋同源模組和模組交互作用的服務，並且能夠為模組演化和細胞生化過程提供有價值的新見解。

表二、JAK2-PAFR-TYK2 模組的鄰近模組功能註解

Functional groups	Module ID	Name	<i>p</i> -value of MMI	Module family annotation	Module family annotation at level 5	FunCat annotations [54]
cellular protein metabolic pathway	2563	FGFR2-c-Cbl-Lyn-Fyn complex	2.28E-08	protein phosphorylation	cellular protein metabolic process (GO: 0044267)	proteasomal degradation (ubiquitin/proteasomal pathway); tyrosine kinase; FGF-receptor signalling pathway; structural cell of tissue (fibroblast, osteoblast, etc.)
cellular protein metabolic pathway	2564	p21(ras)GAP-Fyn-Lyn-Yes complex, thrombin stimulated	2.28E-08	protein phosphorylation	cellular protein metabolic process (GO: 0044267)	enzyme mediated signal transduction; blood coagulation; platelet
cellular protein metabolic pathway	2565	CD20-LCK-LYN-FYN-p75/80 complex, (Raji human B cell line)	2.28E-08	protein phosphorylation	cellular protein metabolic process (GO: 0044267)	tyrosine kinase; B-cell
cellular protein metabolic pathway	1185	EGFR-containing signaling complex	3.99E-05	protein phosphorylation; cell communication; phosphoinositide-mediated signaling	cellular protein metabolic process (GO: 0044267)	EGF-receptor signalling pathway
cell surface receptor linked signaling pathway	5713	SH3P2/OSTF1-CBL-SRC complex	1.06E-05	epidermal growth factor receptor signaling pathway	cell surface receptor linked signaling pathway (GO: 0007166)	cell adhesion; actin cytoskeleton

表二、 JAK2-PAFR-TYK2 模組的鄰近模組功能註解(2)

Functional groups	Module ID	Name	<i>p</i> -value of MMI	Module family annotation	Module family annotation at level 5	FunCat annotations [54]
cell surface receptor linked signaling pathway	1222	Irs1-Grb2-Ptpn1 complex	2.17E-09	insulin receptor signaling pathway	cell surface receptor linked signaling pathway (GO: 0007166)	modification by phosphorylation, dephosphorylation, autophosphorylation; insulin receptor signalling pathway
cell surface receptor linked signaling pathway	2577	Sam68-p85 P13K-IRS-1-IR signaling complex	3.77E-12	insulin receptor signaling pathway	cell surface receptor linked signaling pathway (GO: 0007166)	insulin receptor signalling pathway; cytoplasm
cell surface receptor linked signaling pathway	2535	SLP-76-Cbl-Grb2-Shc complex, Fc receptor gamma-R1 stimulated	2.28E-08	epidermal growth factor receptor signaling pathway	cell surface receptor linked signaling pathway (GO: 0007166)	transmembrane receptor protein tyrosine kinase signalling pathways; immune response
cell surface receptor linked signaling pathway	2550	Frs2-Grb2-Shp2 complex, FGF stimulated	1.06E-05	fibroblast growth factor receptor signaling pathway	cell surface receptor linked signaling pathway (GO: 0007166)	FGF-receptor signalling pathway
cell surface receptor linked signaling pathway	5282	CAS-SRC-FAK complex	1.06E-05	epidermal growth factor receptor signaling pathway; nerve growth factor receptor signaling pathway; integrin-mediated signaling pathway	cell surface receptor linked signaling pathway (GO: 0007166)	transmembrane receptor protein tyrosine kinase signalling pathways

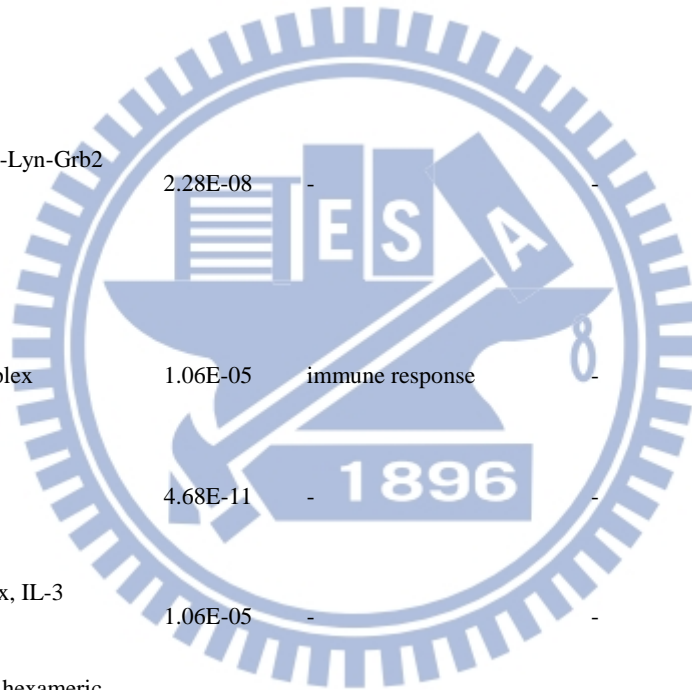
表二、JAK2-PAFR-TYK2 模組的鄰近模組功能註解(3)

Functional groups	Module ID	Name	p-value of MMI	Module family annotation	Module family annotation at level 5	FunCat annotations [54]
cell surface receptor linked signaling pathway	3096	ITGA6-ITGB4-SHC1-GRB2 complex	3.99E-05	integrin-mediated signaling pathway	cell surface receptor linked signaling pathway (GO: 0007166)	non-enzymatic receptor mediated signalling; cytoskeleton/structural proteins; epithelium; skin
cell surface receptor linked signaling pathway	2529	LAT-PLC-gamma-1-p85-GRB2-CBL-VAV-SLP-76 signaling complex, C305 activated	7.20E-13	T cell receptor signaling pathway; platelet activation; nerve growth factor receptor signaling pathway; T cell costimulation	cell surface receptor linked signaling pathway (GO: 0007166)	polyphosphoinositol mediated signal transduction; transmembrane receptor protein tyrosine kinase signalling pathways; adaptive cell mediated response (T-cells)
cell surface receptor linked signaling pathway	2454	CIN85-CBL-SH3GL2-EGFR complex, EGF stimulated	1.06E-05	-	-	vesicle formation; endocytosis EGF-receptor signalling pathway
cell surface receptor linked signaling pathway	482	cRET-Shc-Grb2-Gab2-P85PI3K-Shp2 complex	3.03E-10	-	-	second messenger mediated signal transduction; transmembrane receptor protein tyrosine kinase signaling pathway



表二、 JAK2-PAFR-TYK2 模組的鄰近模組功能註解(4)

Functional groups	Module ID	Name	<i>p</i> -value of MMI	Module family annotation	Module family annotation at level 5	FunCat annotations [54]
cell surface receptor linked signaling pathway	2480	CIN85 complex (CIN85, CRK, BCAR1, CBL, PIK3R1, GRB2, SOS1)	1.35E-09	-	-	transmembrane receptor protein tyrosine kinase signalling pathways
cell surface receptor linked signaling pathway	2536	PLC-gamma-2-SLP-76-Lyn-Grb2 complex	2.28E-08	-	-	polyphosphoinositol mediated signal transduction;transmembrane receptor protein tyrosine kinase signalling pathways;
interleukine receptor signalling pathway	1515	IL4-IL4R-IL2RG complex	1.06E-05	immune response	-	blood coagulation; platelet interleukine receptor signalling pathway
interleukine receptor signalling pathway	5564	LMO4-gp130 complex	4.68E-11	-	-	transcription activation; interleukine receptor signalling pathway
interleukine receptor signalling pathway	3203	Shc-Ship-Grb2 complex, IL-3 stimulated	1.06E-05	-	-	interleukine receptor signalling pathway
interleukine receptor signalling pathway	1p9m	Crystal structure of the hexameric human IL-6/IL-6 alpha receptor/gp130 complex	1.06E-05	-	-	-



表二、JAK2-PAFR-TYK2 模組的鄰近模組功能註解(5)

Functional groups	Module ID	Name	<i>p</i> -value of MMI	Module family annotation	Module family annotation at level 5	FunCat annotations [54]
-	2528	ERBB2-MEMO-SHC complex	1.06E-05	-	-	cell motility; microtubule cytoskeleton protein targeting, sorting and translocation;
-	3678	RIN1-STAM2-EGFR complex, EGF stimulated	1.06E-05	-	-	endocytosis; EGF-receptor signalling pathway; endosome
-	2542	EGFR-CBL-GRB2 complex	2.17E-09	-	-	cellular signalling; endosome
-	2510	ZAP70-CRKL-WIPF1-WAS complex	3.99E-05	-	-	cellular signalling
-	2475	Cbl-Crkl-Rapgef1 complex	1.06E-05	blood vessel development	vasculature development (GO: 0001944)	integrin receptor signalling pathway; cell migration

## References

1. Gavin AC, Aloy P, Grandi P, Krause R, Boesche M, Marzioch M, Rau C, Jensen LJ, Bastuck S, Dumpelfeld B *et al*: **Proteome survey reveals modularity of the yeast cell machinery.** *Nature* 2006, **440**(7084):631-636.
2. Segal E, Shapira M, Regev A, Pe'er D, Botstein D, Koller D, Friedman N: **Module networks: identifying regulatory modules and their condition-specific regulators from gene expression data.** *Nature Genetics* 2003, **34**(2):166-176.
3. Wagner GP, Pavlicev M, Cheverud JM: **The road to modularity.** *Nat Genet* 2007, **8**(12):921-931.
4. Malovannaya A, Lanz RB, Jung SY, Bulynko Y, Le NT, Chan DW, Ding C, Shi Y, Yucer N, Krenciute G *et al*: **Analysis of the Human Endogenous Coregulator Complexome.** *Cell* 2011, **145**(5):787-799.
5. Wang H, Kakaradov B, Collins SR, Karotki L, Fiedler D, Shales M, Shokat KM, Walther TC, Krogan NJ, Koller D: **A complex-based reconstruction of the *Saccharomyces cerevisiae* interactome.** *Molecular & cellular proteomics : MCP* 2009, **8**(6):1361-1381.
6. Ideker T, Ozier O, Schwikowski B, Siegel AF: **Discovering regulatory and signalling circuits in molecular interaction networks.** *Bioinformatics* 2002, **18 Suppl 1**:S233-240.
7. Bader GD, Hogue CW: **An automated method for finding molecular complexes in large protein interaction networks.** *Bioinformatics* 2003, **4**:2.
8. Stelzl U, Worm U, Lalowski M, Haenig C, Brembeck FH, Goehler H, Stroedicke M, Zenkner M, Schoenherr A, Koeppen S *et al*: **A human protein-protein interaction network: a resource for annotating the proteome.** *Cell* 2005, **122**(6):957-968.
9. Espinosa-Soto C, Wagner A: **Specialization Can Drive the Evolution of Modularity.** *Plos Comput Biol* 2010, **6**(3).
10. Kirschner M, Gerhart J: **Evolvability.** *P Natl Acad Sci USA* 1998, **95**(15):8420-8427.
11. Lenski RE, Barrick JE, Ofria C: **Balancing robustness and evolvability.** *Plos Biol* 2006, **4**(12):2190-2192.
12. Wagner A: **Robustness, evolvability, and neutrality.** *Febs Lett* 2005, **579**(8):1772-1778.
13. Bandyopadhyay S, Mehta M, Kuo D, Sung MK, Chuang R, Jaehnig EJ, Bodenmiller B, Licon K, Copeland W, Shales M *et al*: **Rewiring of Genetic Networks in Response to DNA Damage.** *Science* 2010, **330**(6009):1385-1389.
14. Kersey P, Bower L, Morris L, Horne A, Petryszak R, Kanz C, Kanapin A, Das U, Michoud K, Phan I *et al*: **Integr8 and Genome Reviews: integrated views of complete genomes and proteomes.** *Nucleic Acids Research* 2005, **33**:D297-D302.

15. Punta M, Coggill PC, Eberhardt RY, Mistry J, Tate J, Boursnell C, Pang N, Forslund K, Ceric G, Clements J *et al*: **The Pfam protein families database**. *Nucleic Acids Research* 2012, **40**(Database issue):D290-301.
16. Andreeva A, Howorth D, Brenner SE, Hubbard TJP, Chothia C, Murzin AG: **SCOP database in 2004: refinements integrate structure and sequence family data**. *Nucleic Acids Research* 2004, **32**:D226-D229.
17. Chen CC, Lin CY, Lo YS, Yang JM: **PPISearch: a web server for searching homologous protein-protein interactions across multiple species**. *Nucleic Acids Research* 2009, **37**(Web Server issue):W369-375.
18. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT *et al*: **Gene ontology: tool for the unification of biology. The Gene Ontology Consortium**. *Nat Genet* 2000, **25**(1):25-29.
19. Aranda B, Achuthan P, Alam-Faruque Y, Armean I, Bridge A, Derow C, Feuermann M, Ghanbarian AT, Kerrien S, Khadake J *et al*: **The IntAct molecular interaction database in 2010**. *Nucleic Acids Research* 2010, **38**:D525-D531.
20. Stark C, Breitkreutz B-J, Chatr-aryamontri A, Boucher L, Oughtred R, Livstone MS, Nixon J, Van Auken K, Wang X, Shi X *et al*: **The BioGRID Interaction Database: 2011 update**. *Nucleic Acids Res* 2011, **39**:D698-D704.
21. Xenarios I, Salwinski L, Duan XQJ, Higney P, Kim SM, Eisenberg D: **DIP, the Database of Interacting Proteins: a research tool for studying cellular networks of protein interactions**. *Nucleic Acids Research* 2002, **30**(1):303-305.
22. Mewes HW, Dietmann S, Frishman D, Gregory R, Mannhaupt G, Mayer KFX, Muensterkoetter M, Ruepp A, Spannagl M, Stuempflen V *et al*: **MIPS: analysis and annotation of genome information in 2007**. *Nucleic Acids Res* 2008, **36**:D196-D201.
23. Ceol A, Aryamontri AC, Licata L, Peluso D, Briganti L, Perfetto L, Castagnoli L, Cesareni G: **MINT, the molecular interaction database: 2009 update**. *Nucleic Acids Research* 2010, **38**:D532-D539.
24. Chen C-C, Lin C-Y, Lo Y-S, Yang J-M: **PPISearch: a web server for searching homologous protein-protein interactions across multiple species**. *Nucleic Acids Res* 2009, **37**:W369-W375.
25. Chen Y-C, Lo Y-S, Hsu W-C, Yang J-M: **3D-partner: a web server to infer interacting partners and binding models**. *Nucleic Acids Res* 2007, **35**:W561-W567.
26. Bork P, Jensen LJ, Kuhn M, Stark M, Chaffron S, Creevey C, Muller J, Doerks T, Julien P, Roth A *et al*: **STRING 8-a global view on proteins and their functional interactions in 630 organisms**. *Nucleic Acids Res* 2009, **37**:D412-D416.
27. Kanehisa M, Araki M, Goto S, Hattori M, Hirakawa M, Itoh M, Katayama T, Kawashima S, Okuda S, Tokimatsu T *et al*: **KEGG for linking genomes to life and the environment**. *Nucleic Acids Research* 2008, **36**:D480-D484.

28. Matthews LR, Vaglio P, Reboul J, Ge H, Davis BP, Garrels J, Vincent S, Vidal M: **Identification of potential interaction networks using sequence-based searches for conserved protein-protein interactions or "interologs"**. *Genome Res* 2001, **11**(12):2120-2126.
29. Yu HY, Luscombe NM, Lu HX, Zhu XW, Xia Y, Han JDJ, Bertin N, Chung S, Vidal M, Gerstein M: **Annotation transfer between genomes: Protein-protein interologs and protein-DNA regulogs**. *Genome Res* 2004, **14**(6):1107-1118.
30. Ruepp A, Brauner B, Dunger-Kaltenbach I, Frishman G, Montrone C, Stransky M, Waegle B, Schmidt T, Doudieu ON, Mpfen VS *et al*: **CORUM: the comprehensive resource of mammalian protein complexes**. *Nucleic Acids Research* 2008, **36**:D646-D650.
31. Meyer PA, Ye P, Suh MH, Zhang M, Fu J: **Structure of the 12-Subunit RNA Polymerase II Refined with the Aid of Anomalous Diffraction Data**. *J Biol Chem* 2009, **284**(19):12933-12939.
32. Stark C, Breitkreutz BJ, Chatr-aryamontri A, Boucher L, Oughtred R, Livstone MS, Nixon J, Van Auken K, Wang XD, Shi XQ *et al*: **The BioGRID Interaction Database: 2011 update**. *Nucleic Acids Research* 2011, **39**:D698-D704.
33. Mewes HW, Dietmann S, Frishman D, Gregory R, Mannhaupt G, Mayer KFX, Munsterkotter M, Ruepp A, Spannagl M, Stuempflen V *et al*: **MIPS: analysis and annotation of genome information in 2007**. *Nucleic Acids Research* 2008, **36**:D196-D201.
34. von Ballmoos C, Cook GM, Dimroth P: **Unique rotary ATP synthase and its biological diversity**. *Annu Rev Biophys* 2008, **37**:43-64.
35. Lo YS, Lin CY, Yang JM: **PCFamily: a web server for searching homologous protein complexes**. *Nucleic Acids Research* 2010, **38**:W516-W522.
36. Medina I, Carbonell J, Pulido L, Madeira SC, Goetz S, Conesa A, Tarraga J, Pascual-Montano A, Nogales-Cadenas R, Santoyo J *et al*: **Babelomics: an integrative platform for the analysis of transcriptomics, proteomics and genomic data with advanced functional profiling**. *Nucleic Acids Research* 2010, **38**:W210-W213.
37. Boyle EI, Weng SA, Gollub J, Jin H, Botstein D, Cherry JM, Sherlock G: **GO::TermFinder - open source software for accessing Gene Ontology information and finding significantly enriched Gene Ontology terms associated with a list of genes**. *Bioinformatics* 2004, **20**(18):3710-3715.
38. Lo Y-S, Chen Y-C, Yang J-M: **3D-interologs: an evolution database of physical protein-protein interactions across multiple genomes**. *Bmc Genomics* 2010, **11**.
39. Barabasi AL, Oltvai ZN: **Network biology: Understanding the cell's functional organization**. *Nature Reviews Genetics* 2004, **5**(2):101-U115.
40. Seyed-allaei H, Bianconi G, Marsili M: **Scale-free networks with an exponent less**

- than two. *Phys Rev E* 2006, **73**(4).
41. Li SS, Xu K, Wilkins MR: **Visualization and Analysis of the Complexome Network of *Saccharomyces cerevisiae***. *J Proteome Res* 2011, **10**(10):4744-4756.
  42. Wang HD, Kakaradov B, Collins SR, Karotki L, Fiedler D, Shales M, Shokat KM, Walther TC, Krogan NJ, Koller D: **A Complex-based Reconstruction of the *Saccharomyces cerevisiae* Interactome**. *Molecular & Cellular Proteomics* 2009, **8**(6):1361-1381.
  43. Schwartz MA, Assoian RK: **Integrins and cell proliferation: regulation of cyclin-dependent kinases via cytoplasmic signaling pathways**. *J Cell Sci* 2001, **114**(14):2553-2560.
  44. Lukashova V, Chen ZG, Duhe RJ, Rola-Pleszczynski M, Stankova J: **Janus kinase 2 activation by the platelet-activating factor receptor (PAFR): Roles of Tyk2 and PAFR C terminus**. *J Immunol* 2003, **171**(7):3794-3800.
  45. Fridman JS, Scherle PA, Collins R, Burn T, Neilan CL, Hertel D, Contel N, Haley P, Thomas B, Shi J *et al*: **Preclinical Evaluation of Local JAK1 and JAK2 Inhibition in Cutaneous Inflammation**. *J Invest Dermatol* 2011, **131**(9):1838-1844.
  46. Gu JY, Wang YF, Gu X: **Evolutionary analysis for functional divergence of Jak protein kinase domains and tissue-specific genes**. *J Mol Evol* 2002, **54**(6):725-733.
  47. Ruepp A, Zollner A, Maier D, Albermann K, Hani J, Mokrejs M, Tetko I, Guldener U, Mannhaupt G, Munsterkotter M *et al*: **The FunCat, a functional annotation scheme for systematic classification of proteins from whole genomes**. *Nucleic Acids Research* 2004, **32**(18):5539-5545.
  48. Lowes MA, Bowcock AM, Krueger JG: **Pathogenesis and therapy of psoriasis**. *Nature* 2007, **445**(7130):866-873.
  49. Mesa RA: **Ruxolitinib, a selective JAK1 and JAK2 inhibitor for the treatment of myeloproliferative neoplasms and psoriasis**. *Idrugs* 2010, **13**(6):394-403.
  50. Chrencik JE, Patny A, Leung IK, Korniski B, Emmons TL, Hall T, Weinberg RA, Gormley JA, Williams JM, Day JE *et al*: **Structural and Thermodynamic Characterization of the TYK2 and JAK3 Kinase Domains in Complex with CP-690550 and CMP-6**. *J Mol Biol* 2010, **400**(3):413-433.
  51. Malik S, Roeder RG: **Dynamic regulation of pol II transcription by the mammalian Mediator complex**. *Trends Biochem Sci* 2005, **30**(5):256-263.
  52. Sun M, Jiang R, Li JD, Luo SL, Gao HW, Jin CY, Shi DL, Wang CG, Wang B, Zhang XY: **MED19 promotes proliferation and tumorigenesis of lung cancer**. *Mol Cell Biochem* 2011, **355**(1-2):27-33.
  53. Kuuselo R, Savinainen K, Sandstrom S, Autio R, Kallioniemi A: **MED29, a component of the mediator complex, possesses both oncogenic and tumor suppressive characteristics in pancreatic cancer**. *Int J Cancer* 2011, **129**(11):2553-2565.
  54. Ruepp A, Zollner A, Maier D, Albermann K, Hani J, Mokrejs M, Tetko I, Guldener U,

Mannhaupt G, Munsterkotter M *et al*: **The FunCat, a functional annotation scheme for systematic classification of proteins from whole genomes.** *Nucleic acids research* 2004, **32**(18):5539-5545.

