PO-YI(Brian), LI

Insilico Medicine Summer Internship

EDUCATION

National Yang Ming Chiao Tung University
Biotechnology (Major)
Computer Science (minor)
Sep 2018 – present
Current GPA: 3.7/4.3

Shanghai Jiao Tung University(exchange)
Electronic Engineering
Jul 2019 – Aug 2019
GPA: 4/4

LANGUAGE / SKILLS

Language:
- Mandarin (Native)
- English (Professional)

Skills:
- Python
- C++
- Php
- Bash
- Pymol
- Html
- Mac OS
- Windows
- Microsoft Office

COURSES

Biotechnology:
- Molecular Biology
- Biochemistry
- Cell Biology
- Biostatistics
- Computational biology lab
- Machine Learning in computational biology

Computer Science:
- Data Structure
- Object-oriented Programming
- Algorithms
- Discrete Mathematic

Personal Information

Hi! I' m Brian, a junior student with a major in biotechnology and minor in computer science. I have been doing research in prof. Jinn-Moon Yang' s lab (BioXGEM) since 2020. I picked biotech as my major because I used to have a dream of becoming a biotech engineer. However, changes had happened after I took the course "Introduction to computational biology" in which I used Python programming to solve bundles of biological problems. At the same time, not only did I discover my passion for programming, but I also found my interest in entering the field of bioinformatics.

I trust all the skills I possess increases my ability to make a greater impact using computational biology which is why I am applying for the intern position at Insilico Medicine. Thank you for your consideration, and I look forward to hearing from you.

Projects / Competition

1. **PPISU: A Web Server for comprehensive protein-protein interface analysis (Reviewing, Contributed on NAR)**
Jun 2020 – Jan 2021 / BioxGEM
Description: In this project, I was responsible for data analysis with Python and C++. Besides, I also participated in user interface and user experience of the web design. This process helped me acquire the ability of communicate effectively with colleagues and allowed me to explore more about Protein-protein interaction.

2. **Predict Protein-protein Hot Interacting Regions (HIRs) With Random Forest Model (Reviewing, College student research program)**
Jan 2021 – present / BioXGEM
Description: After one semester of learning background knowledge and some professional skills from the last project, I am now doing research independently. This project has the main focus on protein-protein interaction and machine learning algorithm. With random forest algorithm, I could take a different approach on the conditional problem of finding the main reasons behind Protein-protein Interaction.

3. **SJTU Entrepreneurship competition (4th place)**
Jul 2019 – Aug 2019 / Shanghai
Description: Different from common entrepreneurship competition that have group registrations, contestants were randomly grouped and should come up with a startup idea in a week. In my group, I was the organizer and speaker. I organized different ideas from group members and then presented the concepts clearly to the audiences and judges. It' s also the time that I found my place in a group, except for only being a leader, I can also listen to others and improve the consensus.

Extracurricular Activities

1. **Class leader / Sep 2018 - present**
Organization: Class of 2022, Biotechnology department
2. **Minister of Intra-Association Affair / Sep 2018 - present**
Organization: Student Association of Biotechnology department
3. **Undergraduate Research Fellow / June 2020 - present**
Organization: BioXGEM (Prof. Jinn-Moon Yang' s Lab

brian7149@gmail.com 0902-200-972 Anhe Rd., Daan Dist., Taipei City, Taiwan

科技部大專生研究計畫（審查中）

作者：李柏毅（PO-YI, LI）
校系級：國立陽明交通大學生物科技學系大三
投稿日期：2021/2/12
指導教授：楊進木
題目：以隨機森林模型預測蛋白質交互熱區
Predict Protein-protein Hot Interacting Regions (HIRs) With Random Forest Model

摘要：

蛋白質－蛋白質交互作用（protein-protein interaction，PPI）在生物體執行功能過程中扮演著相當重要的角色，而本實驗室也一直致力於尋找交互作用面上重要的交互區域（interacting regions，IRs），藉以應用於疾病基因突變研究及電腦輔助藥物設計上。本實驗室認為會影響蛋白質結合能力、生物功能及疾病的關鍵並非是由單個殘基（residues）所造成，而是受一個區域環境改變所影響，因此，定義 IR 係由「一群」殘基所組成，並進一步定義蛋白質交互作用面上相對具有影響力的一個區域為交互熱區（hot interacting regions，HIRs）。為了進一步探討蛋白質上的哪些區域與疾病等重要生物功能相關，本計畫將透過機器學習（machine learning）中的隨機森林模式（random forest models），預測蛋白質上的交互熱區。

本實驗室已經從蛋白質資料庫（Protein Data Bank，PDB）收集了 72,961 蛋白質結晶結構，並透過實驗室先前累積的技術，將 144,636 個蛋白質－蛋白質交互作用面（protein-protein interfaces）建立了 2,733,029 組 IRs。並且從 SKEMPI2.0、Uniprot 及 OMIM 等資料庫分別收集了與蛋白質結合能力、生物功能及疾病相關的重要殘基，並定義含有重要殘基（例如：熱點）的 IRs 為 HIRs，其餘為 IRs，HIRs 在系統生物學領域中不僅可能成為影響蛋白質作用面的關鍵，更能在日後成為預測熱點的工具。

在本計畫的初步結果中，已經整理出 IRs 的各種組成，但是對於其中與哪些特徵的關聯依然陌生，這對未來希望做出熱點預測或甚至是藥物設計都是不可或缺的資訊，因此，我將透過隨機森林的方法建立多種特徵（features）的模型進行運算，在各種可能的特徵中，找出哪一項亦或者是哪一群主導了 IRs 的重要性，並在計算出各個特徵的權重後，設計一套具有生物意義的等式進行分析，為此，本計畫將利用本實驗室積累豐富的經驗以及優勢解決上述議題：(1) 我們是全球第一個提出分子介面家族以及交互作用結構單元的團隊；(2) 我們於計算系統生物學和藥物設計領域皆建立許多新穎模型。相信透過此模型，將能找到一些尚未被發現確與疾病密切相關的區域。