若手研究者インターナショナル・トレーニング・プログラム(ITP)

バイオインフォマティクスとシステムズバイオロジーの国際連携教育研究プログラム 報告書

Name: Ahmed Mohamed Mohamed
Title: Network-based analysis of cancer pathways
Institute: Bioinformatics Center (Mamitsuka Laboratory), Kyoto University Institute for Chemical Research; Kyoto, Uji, Gokasho 611-0011, Japan
Partner institute of your choice : Bioinformatics Program, Boston University (Prof. Charles DeLisi)
Duration of your choice: January 24th – April 20th , 2013

Report:

Introduction to the host laboratory and general lifestyle in Boston:

During my international training program, I was staying at Professor Charles DeLisi's laboratory in faculty of Engineering and Life Sciences at Boston University (BU). The laboratory focuses on analysis of biological networks, with particular emphasis on applications to cancer. Professor DeLisi was one of the initiators of the human genome project, for which he received the Presidential Citizens medal. In the laboratory, I shared a room with 2 PhD students, as well as 3 postdoctoral researchers, which facilitated constructive discussion in research and life activities.

During my first month of my stay, I lived with a host family on the suburbs of Boston. My host family helped me significantly in getting accustomed to the new lifestyle. After that, I moved closer to BU with an American friend of mine, who is also a graduate student in BU.

Research Objectives:

Cancer develops as a result of pathway dysregulations, that is, in turn, caused by mutations in its participating genes. The successful detection of genetic mutations causing cancer (driver mutations) has been hampered by three main issues. First, complete knowledge of signaling pathways in the cancerous tissues are still lacking. Second, the high proliferative activity of the cancerous tissue cause many additional DNA aberrations (passenger mutations). Third, the small number of patients tested is very small, which prevents identification of gene mutation based on their frequency.

Recent sequencing initiative for different tumor types revealed certain trends in genetic mutations. One of these trends is Mutual exclusivity (ME), where driver mutations in the same biological pathways tend not to occur in the same patient.

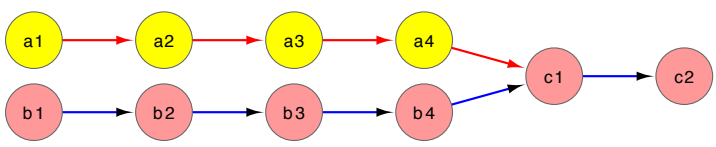

Figure 1 Biological Pathway Illustration

To illustrate the biological interpretation for ME set, we consider the following example: Suppose that C2 gene need to be dysregulated for cancer development. C2 is downstream of C1, which in turn, is regulated by 2 pathways a&b. To successfully dysregulate C1, a & b pathways needs to be non-functioning.

Report (Continued) :

According to the representation in Figure 1, any mutation in genes (a1~a4) is sufficient for pathways "a" de-functioning, and similarly pathway "b". Assuming a perfect tumor evolution process, each tumor sample will have exactly one mutation in each pathway "a" & "b", or a mutated C1 gene. Therefore, mutation network will look like the figure on the right, where blue edges indicate ME, and red edge indicate co-occurrence.

After extensive discussions with researchers in the host laboratory, we have come to the conclusion that protein complexes (multiple genes bound together to form a single functioning unit) can be treated similarly. I started a study aimed to identify ME-mutated protein complexes that highlight commonly perturbed sub-pathways in cancer.

All human protein complexes were fetched from Reactome Database. Data for 27 recent tumor sequencing project data were downloaded from cbioportal.org. To assess the ME of Reactome complexes, the following steps were performed to calculate a P value:

1. For each study, get all mutations from CbioPortal.org
2. Convert Mutations to a Mutation Matrix (Pateint Gene matrix). I used UniProt IDs to index genes because Reactome is curated on Unitpot IDs. Multiple mutations in the same gene were counted as a single mutation.
3. Calculate the hypergeometric probability for all pairwise genes in the mutation matrix.
4. Take Reactome complexes that have at least 2 genes mutated in the current study.
5. For each n-sized complex, calculate Complex "**score**" by summing up all pairwise probabilities.
6. Estimate the empirical distribution of scores for n-sized random genes.
7. Calculate the P value of the complex using ECDF.

Results:

To find common complexes between studies, I plotted complexes ($P<0.001$), the colored them by coverage, to see if there are common, and high-coverage complexes. Out of the 27 studies, 24 showed significant ME mutation signature in PI3K-inhibitor complex. The complex is implicated in cancer pathways and the impairment of PI3K signaling pathway has been tightly linked with cancer development. Other complexes showing frequent ME signature belonged to immune-regulation, mRNA splicing and cell cycle pathways, all of which are linked to cancer. This assured us of the potential applicability of our method is detecting driver mutations.

Plan (Continued)

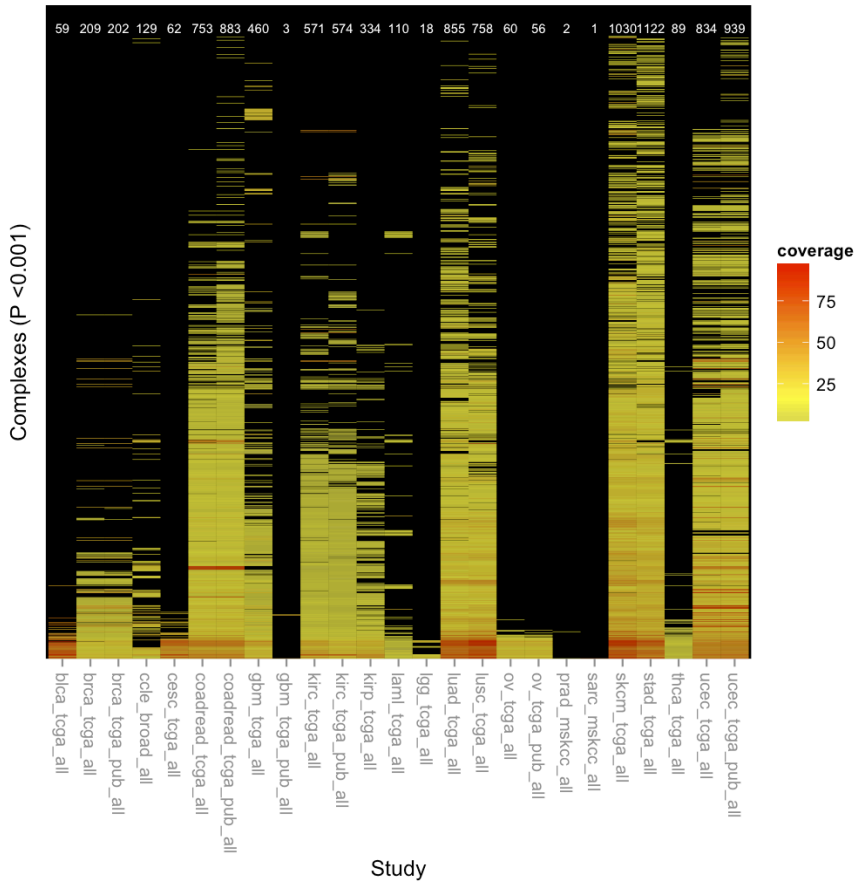

Figure 2 Study-Complex Plot. Complexes are colored according to their coverage. The numbers above the plot indicate complexes passing the filter in each study.

Other Activities:

Aside from wandering in Boston, I exploited my stay in the USA to visit New York City and Philadelphia. In NYC, I visited the landmarks of Times Square, Central Park, and the Empire State Building, while enjoying the magnificent street shows there. In Philadelphia, I visited the Independence Hall, a historic landmark were the American constitution was written.

Boston covered in snow

Parades in NYC

In front of Independence Hall

Acknowledgements:
I would like to express my sincere gratitude to Professor DeLisi and Dr. Zhenjun Hu for their relentless help during my stay. Similar thanks are also extended to the lab members for their hospitality and help in both life and research.

若手研究者インターナショナル・トレーニング・プログラム(ITP)

バイオインフォマティクスとシステムズバイオロジーの国際連携教育研究プログラム 報告書