**CS2220: Introduction to Computational Biology**

**Group Project**

Due on 15 November 2020

This project contributes 15% to the final course grade

This project contributes 15% to your grade in CS2220. This project should be done in a group of no more than three. Please form your own group. Each group should submit a single report for this assignment. Each group is also expected to make a presentation to fellow students and the module coordinator, as well as submitting a short report to the module coordinator.

**Background**

Predicting the phenotypic effects of mutations is an important application in clinical genetic diagnostics. Many approaches have been proposed for this purpose over the past two decades. The objective of this assignment is, firstly, to provide students an introduction to this topic, an appreciation of some key ideas and the progress made. Secondly, it provides students an opportunity to read, present, and discuss research paper.

Below is a set of representative papers to be read on prediction of mutations with adverse effects:

[bckgrd-hmg2012] Sunyaev, “Inferring causality and functional significance of human coding DNA variants”, *Human Molecular Genetics*, 21:R10-R17, 2012.

[sift-nar2003] Ng & Henikoff, “SIFT: predicting amino acid changes that affect protein function”, *Nucleic Acids Research*, 13(13):3812-3814, 2003.

[snap-nar2007] Bromberg & Rost. “SNAP: predict effect of non-synonymous polymorphisms on function”, *Nucleic Acids Research*, 35(11):3823-3835, 2007.

[polyphen2-natmeth2010] Adzhubei et al., “A method and server for predicting damaging missense mutations”, *Nature Methods*, 7(4):248-249, 2010.

[netdsnp-plos2013] Johansen et al., “Prediction of disease causing non-synonymous SNPs by the artificial neural network predictor NetDiseaseSNP”, *PLoS ONE*, 8(7):e68370, 2013.

[suspect-jmb2014] Yates et al., “SuSPect: Enhanced prediction of single amino acid variant (SAV) phenotype using network features”, *Journal of Molecular Biology*, 426:2692-2701, 2014.

[zhulin-gm2016] Adebali et al., “Establishing the precise evolutionary history of a gene improves prediction of disease-causing missense mutations”, *Genetics in Medicine*, 2016, 18(10):1029-1036, 2016.

[msc-natmeth2016] Itan et al., “The mutation significance cutoff: gene-level thresholds for variant predictions”, *Nature Methods*, 13(2):109-110, 2016.

[weghorn-ng2017] Weghorn & Sunyaev, “Bayesian inference of negative and positive selection in human cancers”, *Nature Genetics*, 49:1785-1788, 2017.

**Tasks and scope**

Students need to form a group of up to 3 persons for this assignment. Each group will be given one of the papers. The main tasks include:

* Review and understand the methodology and findings of assigned paper
* Study supplementary materials to gain background knowledge to interpret the reported findings
* Summarize and present the assigned paper to the class

To accomplish the main tasks effectively, each group is highly encouraged to

* Include additional and related materials to facilitate their understand and interpretation of the GWAS findings
* Maintain efficient team work
* Present information in innovative and effective manners

**Learning objectives**:

Through doing the assignment and participating in the presentation discussion, students learn to:

* Explain typical disease-causing mutation prediction methods
* Elaborate key ideas and advances over the years
* Review and understand academic publications
* Present information effectively
* Work in a team and collaborative setting

**Assessment:**

Team assessment:

* Information included in the final presentation & report
* Effectiveness and clarity of the final presentation & report
* Question and answers after presentation

Individual Assessment:

* Each student will be asked to rate his/her team members (teams are encouraged to discuss the member rating and reach to some consensus)
* Participation in the presentation discussion

**Advice on preparing the presentation & report**

0/ Identify the problem that the paper is solving. Please articulate the motivation, relevance, and technical challenge of the problem.

1/ Describe the method proposed in the paper for solving the problem.

2/ Explain why the propose method is reasonable, as given in the paper and in your own opinion.

3/ Describe how the authors validate the effectiveness of the method.

4/ Analyze and discuss the outcome in 3/.

Please prepare a detailed report on the above. A single report should be submitted for each group. The report should ideally be organized into 6 sections. The first five sections should correspond to the four points above. In the 6th section, please explain clearly the contribution to the report & the presentation of each member of the group; please include and fill the table below:

|  |  |  |  |
| --- | --- | --- | --- |
| Name | Contributions | Report (must add up to 100%) | Present-ation, including preparation of ppt (must add up to 100%) |
| Alpha Wong | He didn’t do anything technical, but he buys coffee for us when the rest of us were reading the papers and writing the report and making the ppt | 10 | 10 |
| Beta Lee | Read the paper, explained the details to the group. Wrote sections 1-3 of the report. Vet and polished the ppt. | 70 | 20 |
| Gamma Chan | Discussed with Beta, wrote section 4 of the report. Prepared the ppt and presented everything. | 20 | 70 |