Scribe: Koo Hoi Ying Kody

---

BMEG 3105: Data analytics for personalized genomics and precision medicine

**Lecture 1: Course Introduction**

Lecturer: Yu LI (CSE) (liyu95.com)

Email address: liyu@cse.cuhk.edu.hk

---

Expected outcome

- Make good use of the resulting huge data sets to understand basic biology and medical conditions
- Apply AI and machine learning in clinical applications in the system level

Pre-course Survey Results:
Response: 47/52 (Mostly BME UG students from U2-U4)

| Drive of Taking the course: | Help may be needed for |
|---|---|
| 1. For the three credits and degree requirement<br>2. For the biological/ genomics/ health applications<br>3. Just exploring a new field<br>4. For the data analytics/ machine learning techniques<br>5. For research project experience | 1. Programming<br>2. Knowledge Application<br>3. Understand concepts in data analytics<br>4. Understand the mathematical background of the data analytical methods<br>5. Additional materials and resources |
| Interested Topics: | What would we provide? |
| 1. Protein-protein / RNA interaction<br>2. Neural networks<br>3. Dynamic programming<br>4. Convolution neural networks<br>5. Cancer genomics | 1. Basic Python tutorials<br>2. Introduce fundamental concepts in data analytics<br>3. Practical usage and project topics<br>4. Explain biology when necessary<br>5. Additional resources and materials |

Course Logistics
Lectures:   Wed 9:30am-11:15am (11:05am), SC L4   Fri 9:30am-10:15am, MMW703
Tutorial:    Fri 10:30-11:15am, MMW703

- ♦ Slides will be available the day before the lecture day
- ♦ Python programming tutorial (Exact schedule determined by the TA)
- ♦ Lecture recording sent on Outlook
- ♦ All materials are available on https://yu979.github.io/BMEG3105-Fall-2024/

Teaching Team:

| Yu Li (Instructor)<br>liyu@cse.cuhk.edu.hk<br>Office hour: 3-5pm, Friday / on request)<br>Location: SHB-106[SEP] | Qinze Yu (TA)<br>qzyu22@cse.cuhk.edu.hk<br>Office hour: 2-4pm, Monday<br>Location: SHB-116 |
|---|---|
| Yimin Fan (TA)<br>fanyimin@link.cuhk.edu.hk<br>Office hour: 2-4pm, Tuesday<br>Location: SHB-904 | Ziqian Lin (TA)<br>linziqian@link.cuhk.edu.hk<br>Office hour: 3-5pm, Friday<br>Location: SHB-904 |

Grading

- Homework (20%): 3 grading homework (5%+5%+5%) and 1 non-grading programming assignment (5%, make sure you learn something from it)
- Scribing (10%): Grading scribing.
  Summarize one of the lectures. Submit it one week after the course. Each student should do at least one lecture. Notice that your note and scribing will be posted online, for others reference. You can choose to remove your name or not. You can sign for at most two, for additional 1%
- In-class quiz (10%): 2 in-class quizzes.
  The questions will be simple. Mainly for checking the participation. The exact dates are on the website: Oct 18, and Nov 27
- Midterm (20%): A grading midterm exam. One bonus question (2%)
- Project (20%): A grading project.
  You can give us your project and seek our help, or we will predefine some projects for you to choose (You should submit a mid-term report (5%), a final report (7%) + presentation (3%) together with the implementation (5%).)
- Final (20%): A grading final

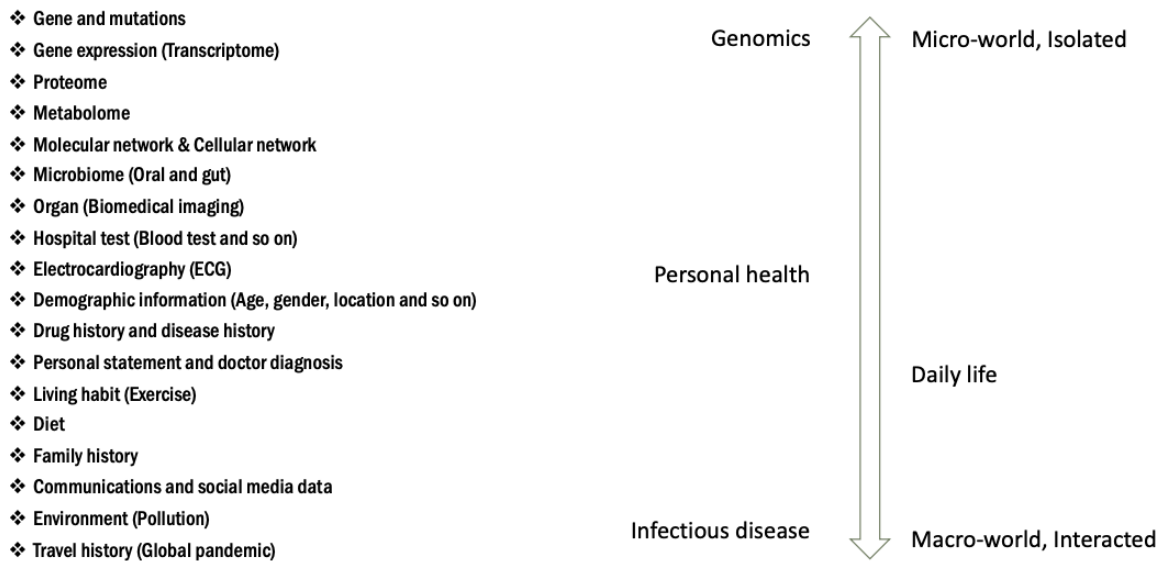## All the exams or quizzes will be open book

Late days policy

- Can be used on A1, A2, A3, PA1, project mid-term report
- Cannot be used on final project report, and scribing report
- 6 late days total, 2 max for any assignment
- Grades will be deducted by 25% for each additional late day
- Please let the TA know when you want to use the late days

## Brief overview of DATA in personalized genomics and precision medicine

Why data analytics?

1. Massive of data is being collected and warehoused

- Web data: Facebook, Google, Amazon, Twitter

- Biological data: DNA sequences, protein structures

- Bank/credit card transaction data: Alipay, PayPal

- Mobile data: China Mobile, CSL

2. **Computers have become cheaper and more powerful**

3. Data analytics are useful

- Aggregate data

- Generate hypothesis

- Support the conclusion

## What data can we have to measure a person?

❖ Gene and mutations
❖ Gene expression (Transcriptome)
❖ Proteome
❖ Metabolome
❖ Molecular network & Cellular network
❖ Microbiome (Oral and gut)
❖ Organ (Biomedical imaging)
❖ Hospital test (Blood test and so on)
❖ Electrocardiography (ECG)
❖ Demographic information (Age, gender, location and so on)
❖ Drug history and disease history
❖ Personal statement and doctor diagnosis
❖ Living habit (Exercise)
❖ Diet
❖ Family history
❖ Communications and social media data
❖ Environment (Pollution)
❖ Travel history (Global pandemic)

Genomics — Micro-world, Isolated

Personal health

Daily life

Infectious disease — Macro-world, Interacted

Scribe: Koo Hoi Ying Kody