BMEG3105 - Data analytics for personalized genomics & precision medicine
Professor: Yu Li
Scribing L1 – Introduction
Wed 04-0-2024

Introduction

1 Why is there data analytics?

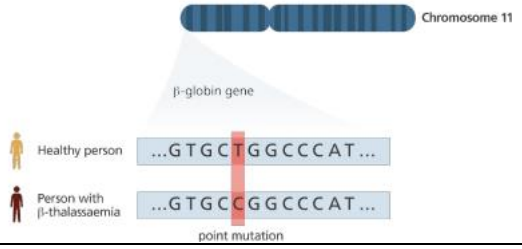
- There's lots of data collected & warehoused in different ways:
  - **Data from the internet**, this is what we produce ourselves every day by surfing on the internet, this is increasing every day as more people are joining social media (e.g. Facebook, Google, Amazon)
  - **Biological data**, from (medical) research (e.g. DNA sequences, protein structures)
  - **Bank transaction data,** coming from all the transactions we make when buying stuff (e.g. Alipay, Paypal)
  - **Mobile data,** coming from the providers (e.g. CSL, China Mobile)
- It's becoming easier and cheaper to store data since computers are more powerful & cheaper than before
- It's useful:
  - Data can be aggregated
  - Hypotheses can be generated
  - Leads to conclusions that can be used to create more personnel advertisements

2 What's the use of data analysis in genomics & precision medicine?

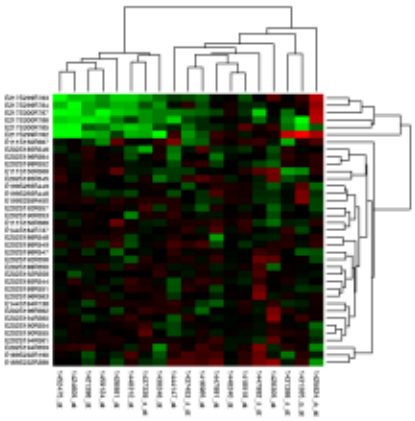
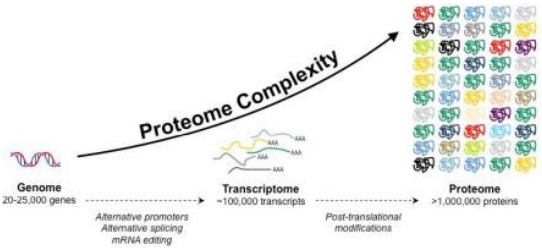
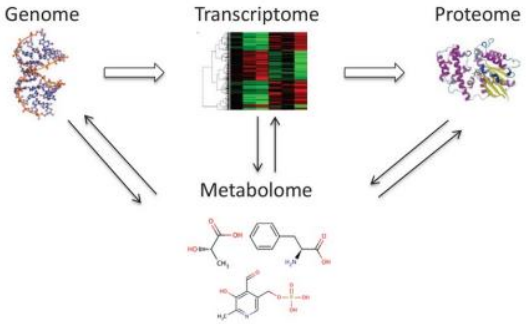
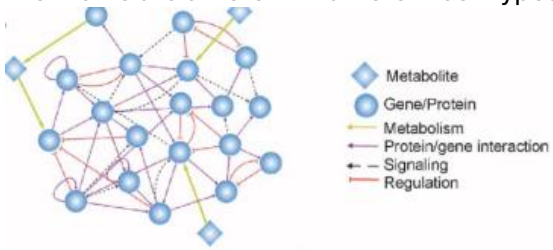
- We already have the data and the analysis is very useful
- The sequencing cost has decreased dramatically ($100M in 2001 vs. $1K in 2015)
- Single cell data accumulates
- Global efforts in building biobank increase

3 Data we have to measure a person

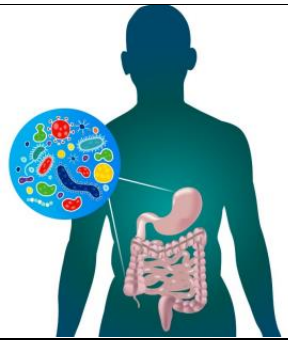
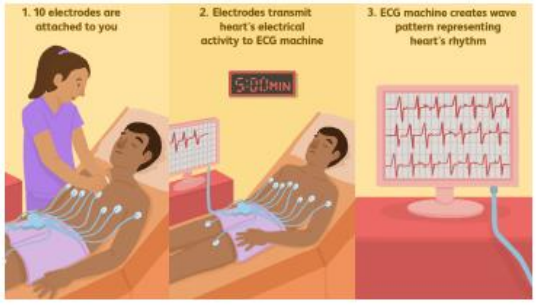
- All kinds of data ordered from smallest to highest level, meaning going from the isolated micro-world of genomics to the interacted macro-world off infectious diseases:

Data	Example
Gene & mutations	Point mutations in the base pairs of DNA (T-C instead of T-A) 

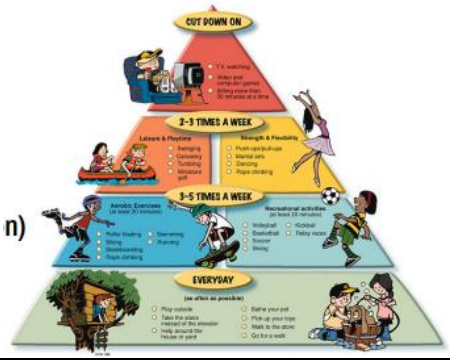
BMEG3105 - Data analytics for personalized genomics & precision medicine
Professor: Yu Li
Scribing L1 – Introduction
Wed 04-0-2024

Gene expression (transcriptome)	Analysis on how much a gene is expressed or how much copies there are 
Proteome	These are the final molecules in our body and can be measured to see if a person has a certain disease or not 
Metabolome	complete set of small-molecule chemicals in a biological sample, influenced by genome & environment 
Molecular network & cellular network	-Within the cell molecules can pass some signals & communicate with each other -networks are different in different cell types 
Microbiome	Oral & gut microbes/viruses that live together with us

BMEG3105 - Data analytics for personalized genomics & precision medicine
Professor: Yu Li
Scribing L1 – Introduction
Wed 04-0-2024

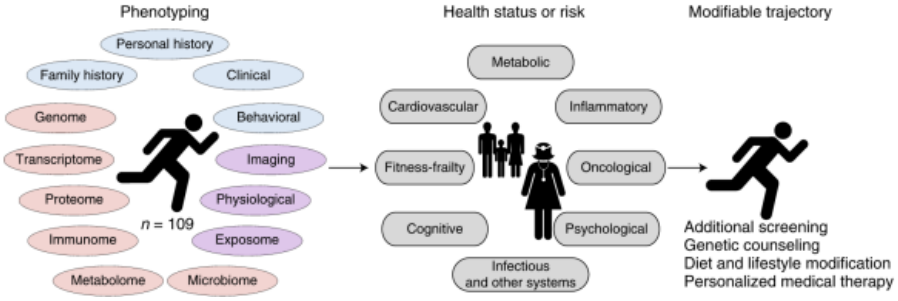
<table>
<tr><td></td><td>  </td></tr>
<tr><td>Organ</td><td>Biomedical imaging is used to measure whether our organs are normal or not</td></tr>
<tr><td>Hospital test</td><td>Blood test to discover diseases or deviations from normal values
<table>
<tr><th>Blood Test</th><th>Result</th><th>Normal Value</th></tr>
<tr><td>WBCs (billion/L)</td><td>8.00</td><td>3.5 to 10.5</td></tr>
<tr><td>Neutrophils (%)</td><td>62</td><td>40 to 70</td></tr>
<tr><td>Lymphocytes (%)</td><td>28</td><td>25 to 45</td></tr>
<tr><td>Monocytes (%)</td><td>10</td><td>2 to 8</td></tr>
<tr><td>Eosinophils (%)</td><td>1</td><td>1 to 5</td></tr>
<tr><td>Basophils (%)</td><td>0</td><td>0 to 1</td></tr>
<tr><td>RBCs (trillion/L)</td><td>3.84</td><td>4.3 to 5.7</td></tr>
<tr><td>Hb (g/dL)</td><td>11.7</td><td>13 to 17</td></tr>
<tr><td>Hematocrit (%)</td><td>37</td><td>37 to 52</td></tr>
<tr><td>Platelets (billion/L)</td><td>262</td><td>150 to 450</td></tr>
</table></td></tr>
<tr><td>Electrocardiography = EGC</td><td>process of producing an electrocardiogram which is a recording of the heart's electrical activity through repeated cardiac cycles

</td></tr>
<tr><td>Demographic info</td><td>-Age
-Gender
-Location: air pollution, the higher the pollution in your environment the higher the chance you acquire some sort of lung disease</td></tr>
<tr><td>Drug & disease history</td><td>Overuse may have irreversible effects on your body</td></tr>
</table>

BMEG3105 - Data analytics for personalized genomics & precision medicine
Professor: Yu Li
Scribing L1 – Introduction
Wed 04-0-2024

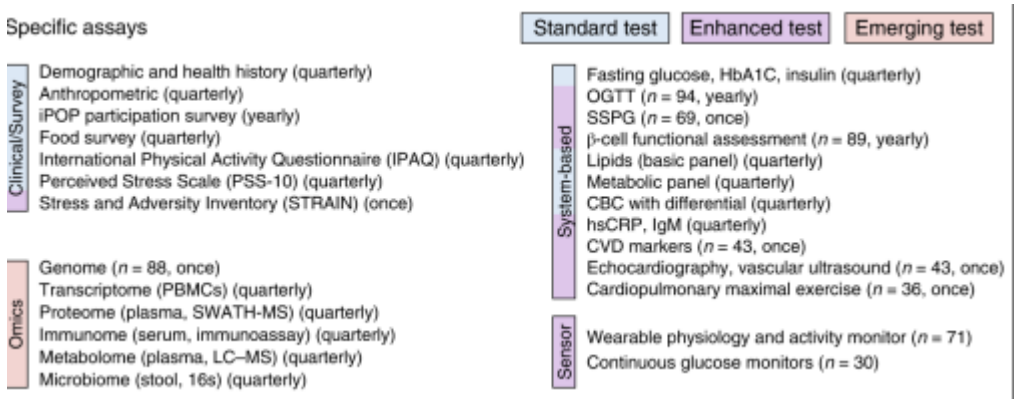
Personal statement & doctors diagnosis	
Living habit	Amount of exercise your perform in your daily life 
Diet	Healthy food, or at least a balanced diet will give you a healthier life
Family history	-Digestion problems may be passed on to offspring -Breast cancer
Communication & social media data	Friends with depression will increase the risk of getting depressed yourself
Travel history	Global pandemic

- Longitudinal big data approach for precision health
    - Process:
        - Phenotyping: test all the data types mentioned above
        - Health status or risk: draw a conclusion from all the data
        - Modifiable trajectory:
            - Additional screening
            - Genetic counselling
            - Diet & lifestyle modification
            - Personalized medical therapy

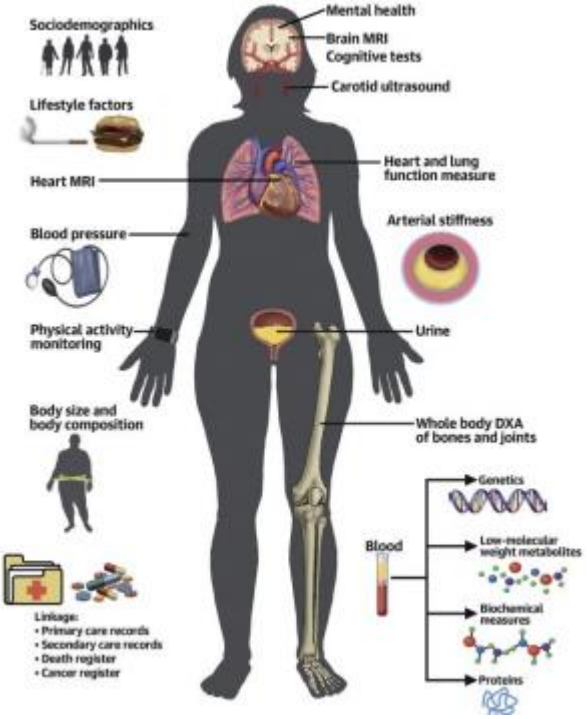
BMEG3105 - Data analytics for personalized genomics & precision medicine
Professor: Yu Li
Scribing L1 – Introduction
Wed 04-0-2024



- Specific assays
  - Clinical survey
  - Omics
  - System-based
  - Sensor tests



- Illustration of a data analysis to measure a person:



BMEG3105 - Data analytics for personalized genomics & precision medicine
Professor: Yu Li
Scribing L1 – Introduction
Wed 04-0-2024

4 Goal of this course

- Learn the fundamental concepts of data analysis
- Know the various data in genomics & mecidine
- Apply data analytic techniques to process the data & resolve problems in biology