Problem Based Learning in Bioinformatics (2020)
Gene-Environment Interactions
Facilitator Denise Daley, PhD contact: denise.daley@hli.ubc.ca
March 27 and 30th

September 2019, you just relocated to BC and started in your graduate program in Bioinformatics. You are very excited to start this new phase of your life, but you are also a bit intimidated, as you are working in a lab that focuses on gene-environment interactions, and you have no idea what a gene-environment interaction is!

You ask you supervisor to give you an example of a gene-environment interaction. These may seem like simple and straight forward questions, but they are not, in fact they are ever evolving field of study that you are about to learn a whole lot more about! Sit back, relax and let's start learning some of these concepts together.

Case Objectives:

1) Understand the concept of necessary and sufficient
2) Understand the concept of necessary but insufficient
3) Understand the concept of gene-environment interaction
4) Understand the concept of effect modification
5) Understand the concept of genetic selection and how effect modification and gene-environment interactions contribute to the process
6) Brief Introduction to the concept of genetic selection

At the end of 1st day class, Ali will split you into groups of 5-6 students per group. Each group will select a presenter who will present the groups findings for the following objectives.

Group Objectives:

1) Identify current sources of information on COVID-19 and case distribution and if possible raw data downloads.
2) Identify if there is any evidence for effect modification of COVID-19 infection.
3) Develop a case for or against genetic effect modification of COVID-19 infection.
4) Group presenter will present group findings and summary (no more than 10 minutes).

First your supervisor asks you if you know or can describe the concepts of **necessary and sufficient?**

No, you reply. You may have had some genetic courses but have never encountered these concepts. Your supervisor tells you not worry and asks you if you are familiar with Cystic Fibrosis?
Oh, now you are on a familiar territory - Cystic Fibrosis is an autosomal recessive genetic disease meaning that two disease alleles are passed to the child one by each of the child's parents, thus the recessive model of inheritance. The gene involved is the Cystic fibrosis transmembrane conductance regulator gene (*CFTR),* of which delta508 is the most common mutation carried by ~80% of individuals. CF affects many systems of the body but the most notable are the lungs and pancreas.

"Excellent! Your supervisor praises! Do you know of any way to get CF without a mutation in the *CFTR* gene?"

Hmmm, you think for a while…. Is there a way? No, you can't think of one, but you wonder is this a trick question? Finally, after considerable thought you say no, it is a genetic disease! You are correct there is no other way to get CF inheritance of two CF alleles, are both necessary and sufficient to result in the phenotype of cystic fibrosis.

Question 1: Do you know of any other diseases that fit the definition of necessary and sufficient?

Question 2: What about inborn errors of metabolism and immunity?

Question 3: What is the newborn screening program?

Question 4: Do all newborn screening conditions meet the definition of necessary and sufficient?

Let's take a closer look at a condition that is a part of the screening panel but does not meet the definition of **necessary and sufficient**. It is **necessary but insufficient and would be a classic example of a gene environment interaction.** It is a genetic condition called Phenylketonuria or (**PKU**). PKU it is an inborn error of metabolism that results in decreased metabolism of the essential amino acid phenylalanine. The condition like CF is a recessive genetic disorder and untreated it leads to severe irreversible intellectual disorders. Inheritance of two PKU alleles is necessary to the disease or phenotype, but ingestion of phenylalanine is also necessary. Without the genetic mutation you can ingest phenylalanine and it produces no symptoms. Phenylalanine is obtained through the diet. It is found in all proteins and in some artificial sweeteners.

Thus, both the genetic mutation **and** ingestion of phenylalanine are necessary to produce the phenotype.

Phenylalanine ubiquitous, as it is present in many foods and drinks, removing it from the diet avoids most of the symptoms of the disease including the severe and irreversible intellectual impairment. Phenylalanine must be removed from the diet as soon as possible as intellectual impairment will result within a few months and is irreversible, thus waiting for symptoms to appear results in irreversible damage.

PKU is a classic example of a gene-environment interaction.

Question 5: Why were the newborn screening panels developed?

Question 6: How many conditions are screened for?

Question 7: What are the criteria for inclusion on the newborn screening panel?

Question 8: Do all provinces screen for the same conditions?

Next your supervisor asks you if you know what is effect modification? You respond that no you are unfamiliar with the concept of effect modification. Effect modification occurs when the magnitude of the effect of the primary exposure on an outcome (i.e., the association) differs depending on the level of a third variable. So, to better understand this concept let's take the example of Chronic Obstructive Pulmonary Disease **(**COPD**)** and α_1-Antitrypsin.

COPD is, as the name implies, a chronic obstructive lung disease that is generally characterized by long-term breathing problems and poor airflow. The main symptoms include shortness of breath and cough with sputum production. COPD is a progressive disease and is generally diagnosed in older adults with a long-term history of smoking. **Smoking is sufficient** but not **necessary** to the development of COPD.

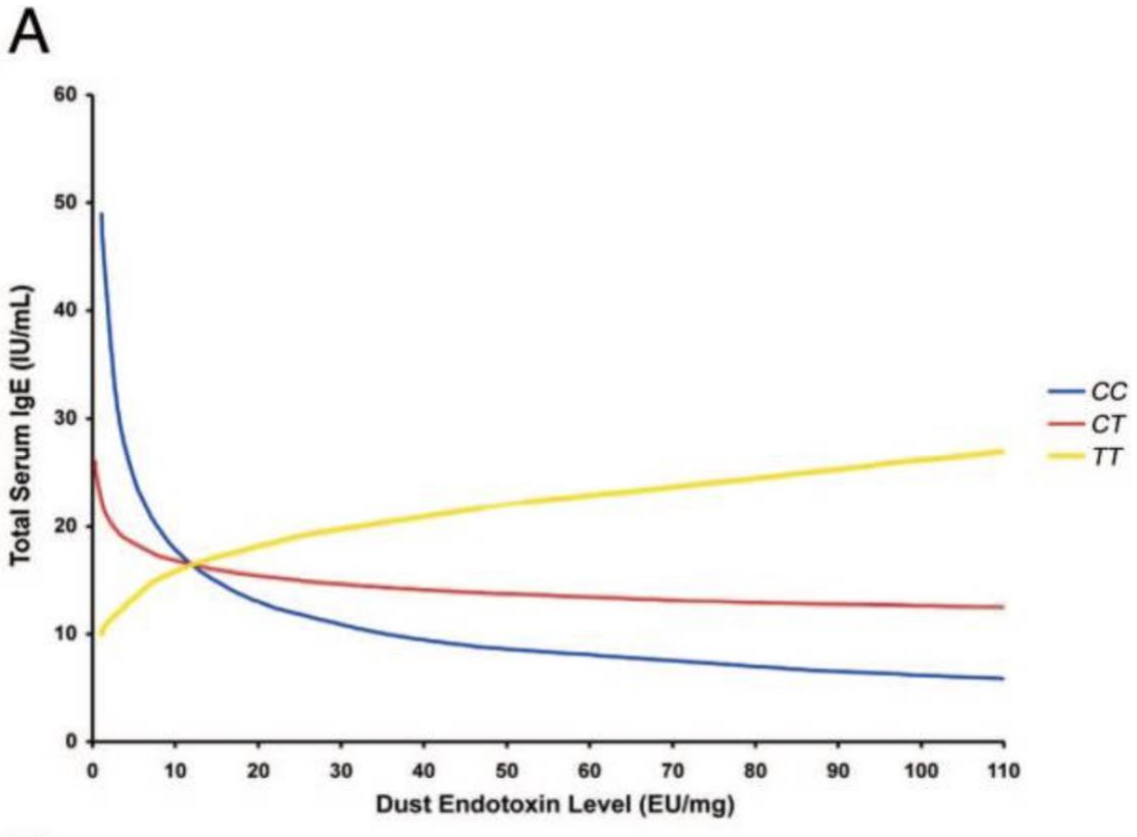
COPD can also be diagnosed in non-smokers, many of whom are subsequently screened and found to be α_1-Antitrypsin deficient (AAT deficiency). Alpha-1antitrypsin is a protease inhibitor encoded by the *SERPINA1* gene on chromosome 14, it has a co-dominant inheritance model. AAT deficiency depends upon the combination of *SERPINA1* alleles and severe deficiency generally results from the inheritance of two Z alleles (PMID: 22761482). Alpha-1 antitrypsin (AAT) is a protein normally found in the lungs and the bloodstream. It helps protect the lungs from the damage caused by inflammation that can lead to emphysema and COPD. People whose bodies do not produce enough of this protein (AAT deficiency) are more likely to develop emphysema and to do so at a younger-than-normal age (30 to 40 years old). Thus, AAT deficiency is not necessary to the development of COPD as it can also be caused by smoking but it is **sufficient** for the development of emphysema and COPD. It is important cause of the disease and it is estimated that up to 5% of people with COPD are AAT deficient (PMID: 22761482).

It is critically important that individuals with AAT deficiency do not smoke, as smoking causes inflammation and damage to the lungs and individuals with AAT have no or impaired protection against. Smoking in AAT deficient individuals results in earlier onset of symptoms (as early as 20) and more severe obstruction (COPD or emphysema).

Smoking modifies the age at onset and severity of disease in AAT deficient individuals, this is an example of **effect modification**. **Effect modification** is present because the primary phenotype in this case emphysema/COPD (statistically dependent variable) and the effect of our primary genetic cause AAT deficiency (predictor variable) differs depending smoking (third variable).

Question 9: How does effect modification differ from the classical gene-environment interaction where both elements are necessary but insufficient?

This is all fascinating but how can I determine if there is an interaction? Can you graph or see interactions? Excellent question! Let's take a look at another example of effect modification that is provided by *CD14* genotypes, endotoxin levels and the risk for developing childhood atopy. *CD14* has a genetic variation that is located in the promoter region of the gene consisting of a cytosine (C) to thymine (T) transition at base pair −260 (C-260T). This region is important in receptor functions that recognise endotoxins. The gene has been extensively studied in relationship to asthma and atopy. Association results are conflicting, likely due to the fact that many studies only examine two factors, atopy status (affected or unaffected) and *CD14* genotypes, but as Williams *et al*. (18320914) demonstrated in 2008 genotype effects are modified by endotoxin levels. To better understand this let's look at a modified figure from the publication. On the Y-axis is total serum IgE a quantitative measure of atopy and on the X-axis is dust endotoxin levels, the different colored lines are for the genotypes. Effect modification is clearly demonstrated as the effect of the genotype differs depending endotoxin level, the CC genotype carries the highest level of risk with lower endotoxin levels and the lowest risk when endotoxin levels are high. Thus, the genotype effect is dependent on not just exposure but the level of exposure.



Question 10: Can you think of any other examples of effect modification, perhaps as related to tropical and/or infectious disease?

Many genetic variants are protective against parasitic and infectious diseases. Classic examples include the β-thalassemia's and Sickle Cell Disease which involve mutations of the β-globin gene which are the most common cause of genetic disorders in humans. There are more than 350 mutations and the worldwide distribution varies but common distribution includes Africa, India, and Asia. The frequency is high, and it is believed to be the result of selective genetic pressure as they convey protection against malarial infection.

Another notable genetic mutation known as CCR5-delta 32 is responsible for the two types of Human immunodeficiency virus (HIV) resistance. CCR5-delta 32 hampers HIV's ability to infiltrate immune cells. The mutation results in the co-receptor on the outside of the cells to develop smaller than usual so that it no longer sits outside of the cell. This CCR5 co-receptor is like door that allows HIV entrance into the cell. The CCR5-delta 32 mutation in a sense locks "the door" which prevents HIV from entering into the cell. 1% of people descended from Northern Europeans, particularly Swedes, are immune to HIV infection. These are homozygous carriers – i.e. recessive inheritance model. While 10-15% of people with European heritage inherited one copy of the variant. One copy of the variant does not prevent against infection. It does however reduce the probability of infection and delays the onset of AIDS. The CCR5-delta 32 variant is tied primarily to the Eurasia region, and has not been found in Africans, East Asians, or Amerindians.

Question 11: Are there any other diseases (infectious or not) where disease susceptibility, severity or survival where the effect modification may be genetically mediated?

In December of 2019 an illness of unknown origin breaks out in China and quickly spreads around the country. In early 2020 it is identified as a coronavirus similar to Severe Acute Respiratory Syndrome (SARS), by March of 2020 it has been declared a word-wide pandemic, the first in modern times. Scientists around the world are working day and night to develop rapid tests and vaccines and understand the epidemiology of the outbreak. You are going to join this effort.

You and your group are tasked with identifying the most informative databases tracking this outbreak. Here are a few to get you started:

RESOURCES:

- John's Hopkins University https://coronavirus.jhu.edu/map.html Very nice graphical map of the outbreak
- US CDC https://www.cdc.gov/coronavirus/2019-ncov/cases-updates/cases-in-us.html
- ECDC https://www.ecdc.europa.eu/en/novel-coronavirus-china
- BC Centre for Disease Control — http://www.bccdc.ca/about/news-stories/stories/2020/information-on-novel-coronavirus
- Public Health Agency of Canada — https://www.canada.ca/en/public-health/services/diseases/2019-novel-coronavirus-infection.html
- World Health Organization — https://www.who.int/emergencies/diseases/novel-coronavirus-2019
- The Centre for Evidence-Based Medicine, University of Oxford — https://www.cebm.net/oxford-covid-19/
- The Lancet — https://www.thelancet.com/coronavirus
- Journal of the American Medical Association — https://jamanetwork.com/journals/jama/pages/coronavirus-alert
- New England Journal of Medicine — https://www.nejm.org/coronavirusAnnals of Internal Medicine — https://annals.org/aim/pages/coronavirus-contentCochrane evidence relevant to critical care — https://www.cochrane.org/special-collection-coronavirus-covid-19-evidence-relevant-critical-care
- Read by QxMD — https://read.qxmd.com/keyword/22047/1

Questions that need to be addressed by the group and addressed in your presentation:

Question 12: Which of these websites provides a dataset for download and breaks down cases/deaths by gender, age, and other pre-existing conditions?

Question 13: What pre-existing conditions increase the risk for death from COVID-19, and by how much?

Question 14: What commonly used group of medications may increase the risk for death from COVID-19?

Question 15: What effect does smoking have on COVID-19?

Question 16: What is the mechanism whereby smoking may be an effect modifier for COVID-19?

Question 17: Expression of what gene may be important in the COVID-19 outbreak?