Module 1.1: Learn - Biology

1.1 Biology Stats

Research Question

In this research CURE, we will be using gene expression to assess which sex chromosomes are present in cancer cell lines used throughout biomedical science. Sex differences in tumor incidence and mortality have been observed throughout the world, across a wide age range, and many different cancer types including those of non-reproductive tissues, usually affecting males more than females. The sex chromosome complement (how many X and Y chromosomes are present in tumor cells) is a key genetic factor that determines molecular mechanisms underlying sex differences in tumor cells. We will be using a resource called the Cancer Cell Line Encyclopedia (CCLE) which has performed and stored a collection of genome-wide measurements for 1019 cell lines grown from tumor samples across a wide range of cancer types. A portion of these cell lines have reported sex from the patient whose tumor was used to derive the cell line, but considering that not all cell lines are annotated with patient sex and that tumors can sometimes lose sex chromosomes, we would like to predict sex for each cell line based on gene expression. We will examine the expression of sex chromosome genes across the CCLE and determine a predicted sex chromosome complement for each cell line. This will be published so that other researchers that are using cell lines in the CCLE for their research can consider genetic sex as an additional variable in their analysis.

So we are asking the following questions in this CURE:

1. What is the predicted sex chromosome complement of cell lines in the CCLE based on expression of genes on the X and Y chromosomes?
2. How does this relate to reported sex of the cell lines (the sex of the patient from which the cancer cell line was derived)?
3. Can we assign a sex to cell lines that do not have a reported sex annotated?
4. Which sex chromosome genes have the most distinct distribution in cancer cell lines predicted to have a Y chromosome (ex, XY males) versus those that do not (ex, XX females)?

In answering these questions, you will learn about the impact of sex chromosomes on cancer biology, how we can use transcriptomic assays such as RNA sequencing to examine gene expression in cancer cells, and how to use computer programming (coding) to view,

analyze, and interpret transcriptomic data to ask important biological questions.

Introduction to the Human Sex Chromosomes

At the center of the research question for this CURE are the sex chromosomes. To help you learn about and communicate results about sex differences, it is important to use the appropriate language when discussing them. For this research question, we will be assaying gene expression differences due to genetic sex (sex chromosome genotype) – not differences associated with gender, which are based on how a person self-identifies and actions a person might take as a result of that self-identified gender. As you learn about sex and sex chromosomes in genetics, you will likely encounter publications or descriptions where someone is speaking about genetic sex and saying 'gender' and vice versa; sex and gender are understudied as a whole so people sometimes do not realize the distinctions between these words.

The sex chromosome complement is the set of sex chromosomes contained in the nucleus. Humans have an XY sex determination system, meaning that individuals with a Y chromosome will typically develop male reproductive organs and individuals without a Y chromosome will develop female reproductive organs, and both are required to create offspring. Sex chromosomes are inherited with one from each parent just like autosomes (non-sex chromosomes). While expression of specific Y chromosome genes in specific cells will trigger the development of testes and specific physical changes during puberty, sex chromosomes are part of the karyotype (complete set of chromosomes) of all cells in the body and expression of genes from the sex chromosomes is subject to a complex system of gene regulation just as genes on autosomes are.

Although genetic males typically have one X chromosome and one Y chromosome, and genetic females typically have two X chromosomes, the number of X and Y chromosomes can vary. Common sex chromosome variations include Kleinfelter syndrome (47,XXY) and Turner syndrome (45,X). Further, sex chromosome content can vary over time. For example, XY individuals can lose their Y chromosome within a proportion of their cells over time as they age and loss of Y chromosome has also been observed in cancer cells.

We will learn a lot more about the X and Y chromosomes as we go through this research project, but this video describes summarizes important sex chromosome biology and genes that sex chromosomes carry:
Secrets of the X chromosome - Robin Ball (https://www.youtube.com/watch?v=veB31XmUQm8)

(https://www.youtube.com/watch?v=veB31XmUQm8)
(Secrets of the X Chromosome Video Transcript (https://docs.google.com/document/d/12rxqgZDmFBmRa5ks4ErR3np5qC5pv7TDsZeamJqYHv4/edit?usp=sharing)).

Cancer

In this course we are studying cell lines from tumor cells from patients suffering from cancer. Cancer is one of the most common illnesses and an incredible amount of research has been done and continues to be done in hopes of treating patients.

To get an idea of what cancer is and how it develops, see the National Cancer Institute's summary here: https://www.cancer.gov/about-cancer/understanding/what-is-cancer (https://www.cancer.gov/about-cancer/understanding/what-is-cancer)

You will read that cancer is a genetic disease that causes changes to genes that control the way cells function and leads to uncontrolled growth of cells with genetic mutations in organs causing those organs to malfunction. The body has mechanisms in place to produce new cells only when needed and replace old or damaged cells, but sometimes this process of checks and balances breaks down. These cells can become lumps of tissue called tumors, some of which can become cancerous (continue to grow and divide) and some of which can invade or travel to other tissues to form new tumors (metastasis). These changes occur as a result of genetic mutations that occur as a result of errors during cell replication, DNA damage, or inherited from our parents.

To better understand how cancer cells are different from healthy cells, we can reference a famous review article which summarizes how cancer cells acquire new abilities to circumvent the checks and balances the body employs to prevent mutations, defects, and overgrowth of cells. This paper called the "Hallmarks of Cancer" (link to this paper and follow up papers are in Additional Resources below) starts with this figure:

Figure: Acquired capabilities of cancer from the Hallmarks of Cancer review by Hanahan and Weinburg.

This article describes how cells undergo one or more of the following to transform into cancerous cells:

1. Self-sufficiency in growth signals: cells no longer need specific stimuli before starting to grow or divide
2. Evading apoptosis: cells are no longer naturally turned over when they are supposed to be
3. Insensitivity to anti-growth signals: cells no longer respond to stimuli that normally signal that growth can not be supported
4. Sustained angiogenesis: tumors need oxygen and nutrients to grow larger so they hijack signaling pathways that cause the development of new blood vessels and recruit a variety of cell types such that new stimuli circulate in their microenvironment
5. Tissue invasion and metastasis: tumors develop the ability to spread into surrounding tissues
6. Limitless replicative potential: cells naturally limit the amount of times they can be replicated, but cancers develop the ability to become immortal

This review was written over two decades ago and so much remains relevant to what is known about cancer biology today. Additional cancer mechanisms have been observed since, see additional resources below for links to follow-up articles.

Sex as a Biological Variable in Studying Cancer

The reason we care about the sex chromosome complement in cancer cells is that it is the most obvious genetic sex difference between females and males, which have been shown to exhibit differences in cancer susceptibility, progression, survival, and response to therapies. The sex chromosome complement (genotype) determines what sex chromosome genes are available in the genome to be expressed, which mechanisms regulate that gene expression (for example, X chromosome inactivation in cells with two X chromosomes), and ultimately contributes to which genes are differentially expressed in tumors of different sex which leads to differences in clinical features of cancer between the sexes.

Differences in how cancer forms differently in different sexes is a result alterations at many levels of gene expression regulation: at the DNA level with genetic variation between individuals, at the gene expression level with which genes are translated and which are further expressed as proteins, and at the post-translational levels with how different protein complexes and signaling pathways operate in the cell.

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7719817/pdf/fonc-10-597788.pdf (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7719817/pdf/fonc-10-597788.pdf)

For this CURE, we are going to scratch the surface of this by using the expression of genes on the X and Y chromosomes to predict which sex chromosomes are present in cancer cell lines. Researchers that use these cell lines in their research should know which sex chromosomes are present when determining how good a model the cell lines are for tumor cells in a patient.

Cell lines

Cell lines are cells that can be maintained in laboratory cell culture equipment. They are used in the lab as a model system that people using them hope will retain some features of the original tumor they came from, and studies that use cell line models have to demonstrate that that is the case. Cell lines are typically clonal, meaning they originate from one cell and gain the ability to grow inside a petri dish with growth medium in an incubator. It is not easy to create cell lines; most cells will die when you remove them from the environment they have been growing in since you are taking away all kinds of signals and connections in their environment. If you are interested in learning more about how cell lines are generated and how to use them, there is a page on this in the Additional Resources section.

Cell lines are used widely in laboratory research because they allow researchers to change specific things and study the effect of that change in a controlled, reproducible environment. Laboratory cell lines are typically immortalized, meaning that they have been growing in the lab for many generations, as opposed to primary, meaning just recently isolated from a tumor growing in the body. Cell lines are used to test the function of specific genes, effectiveness of potential therapies, view changes in cell behavior when you change specific things, and many other aspects of cancer cell biology.

In order for a cell line to be generated, cells have to be able to grow outside the body. In order for cells to become immortalized, they are separated into single cells and allowed to grow and divide for many generations, during which time many genetic mutations can take place that allow the cells to live in this environment. Because cell lines are difficult to generate and sometimes tumors don't yield any stable cell lines, cell lines used in laboratory research are often immortalized cell lines that can be purchased from companies like American Tissue Type Culture (ATCC). In order to do this, scientists can pick cell lines from a catalog that are closest to the cancer they are studying (for example, choosing a breast cancer cell line if you are studying how a particular gene dysregulated in breast cancer tissues functions).

Cancer Cell Line Encyclopedia (CCLE)

In this project, you will be working with a very well curated data set of human cancer cell lines called the Cancer Cell Line Encyclopedia. This set of cell lines includes many of the most commonly used cell lines in laboratory research.

In this figure from the original CCLE paper, you can see that the CCLE includes cell lines from a host of tumor tissues. For all the included cell lines, the researchers conducted a massive molecular profiling effort to thoroughly examine the DNA and messenger RNA in the cell lines in their base state. They demonstrate similarity to primary tumors in terms of chromosomal copy number (B), gene expression (C), and genetic mutations (D) by showing high correlation to profiles observed in primary tumors (tumor cells taken from tumors still inside the body instead of grown in a lab). The idea is that people that use these cell lines to do studies in their lab will have a good estimate of the genetic background of that cell line and will be able to design and interpret their experiments more accurately. We will take that one step further by highlighting the sex chromosome complement of the cell lines and encourage people to consider the sex chromosomes of the cell lines in their interpretation of results.

Figure: A) Proportion of cell lines isolated from various tumor types. B) Correlation of chromosome (DNA) copy number between CCLE cell lines and tumors of matching tissue type. C) Correlation of gene expression (mRNA) between CCLE cell lines and tumors of matching tissue type. D) Correlation of DNA mutation frequency between CCLE cell lines and tumors of matching tissue type. You can see that there are yellow squares along the diagonal, indicating a high correlation in cell lines to the matching primary tumor type. This is why cell lines can be a good model for learning about tumor cell biology.
https://www.nature.com/articles/nature11003 (https://www.nature.com/articles/nature11003)

More work has been done on the CCLE and can be read about in the paper linked in the Additional Resources for this module.

In Module 3 of this course, you will be downloading processed data from the CCLE from their website, but if you take a brief look at it now, there is a description of how the project has grown over the years: **https://sites.broadinstitute.org/ccle/ (https://sites.broadinstitute.org/ccle/)**

Additional Resources

- Hallmarks of Cancer original paper (2000)

    - https://www.cell.com/action/showPdf?pii=S0092-8674%2800%2981683-9 (https://www.cell.com/action/showPdf?pii=S0092-8674%2800%2981683-9)
- Hallmarks of Cancer: The Next Generation (2021)
    - https://www.cell.com/action/showPdf?pii=S0092-8674%2811%2900127-9 (https://www.cell.com/action/showPdf?pii=S0092-8674%2811%2900127-9)
- Hallmarks of Cancer: New Dimensions (2022)
    - https://aacrjournals.org/cancerdiscovery/article-pdf/12/1/31/3052722/31.pdf (https://aacrjournals.org/cancerdiscovery/article-pdf/12/1/31/3052722/31.pdf)
- Additional data in the CCLE
    - https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6697103/ (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6697103/)
- What is a cell line?
    - https://www.researchsquare.com/blog/what-is-a-cell-line (https://www.researchsquare.com/blog/what-is-a-cell-line)