Module 3.1: Learn - Biology

3.1 Biology/Stats

Chromosome X genes

Having gotten a brief introduction to the X chromosome in Module 1, let's learn more about the genes on the X chromosome so we can better understand its role in health and disease.

Fun Facts about the X chromosome

- The X chromosome accounts for about 5% of the total human genome and contains upwards of 1200 genes, about 800 of which are protein coding
- About 1/5th of X chromosome genes have been associated with roles in cognition and brain development, muscle function, immune function, sex and reproduction, and many others
- The X chromosome is necessary for healthy human development
- X and Y chromosomes share evolutionary history so there are areas of sequence similarity between both
    - There are two regions of 100% sequence similarity between X and Y called pseudoautosomal (PAR) regions
    - There are genes called gametologs which have 92-95% sequence similarity between the X and Y versions (ex, DDX3X and DDX3Y) and expression of both versions has been observed

Dosage Compensation (X chromosome inactivation and XIST)

In mammals, including humans, the sex determination system is XX/XY. This means that females that can carry an XX genotype and males typically have XY genotype. This means that cells in XX females have double the nu chromosomes than cells in XY males. If there are twice as many X chromosomes, the two X chromosomes in fe as many gene products as the one X in males and having more than 1 active X chromosome is thought to be pro developed a mechanism to make the number of active X chromosomes equal in males and females (called dosa This mechanism is called X chromosome inactivation (XCI) occurs early in embryonic development and involves but one X chromosome by condensing it into a structure called a Barr body. This holds In cases of sex chromosc well; two X chromosomes are inactivated in Triple X Syndrome (XXX females) and one X chromosome is inactiva Syndrome (XXY males).

This short video shows how X chromosome inactivation works: **X Chromosomes Inactivation (https://youtu.b**

01:48

(**X Chromosomes Inactivation Video Transcript**
(https://docs.google.com/document/d/12wlpb3j0xl02B80CBlCol9B_Sp21iGRontR2oHgeK1Y/edit?usp=sharing))

The video shows how in XX females, one of the X chromosomes reads a long noncoding transcript off a region o
called the X inactivation center (XIC). This transcript called XIST begins a cascade of actions that leads to the gr
one X chromosome into a Barr body structure. The Barr body is replicated and distributed just like the active X c
embryo cell divides and develops into all the parts of the body, but gene expression from the Barr body is genera
addition to compaction driven by XIST, X chromosome inactivation is also associated with methylation of the histo
DNA is wrapped, which is also associated with turning off gene expression. It has been observed that some gen
expressed from the inactive X chromosome; this is called X inactivation escape. The active X inactivation center
gene called TSIX which is antisense to the XIST transcript; this transcript is expressed on the activated X chromo
XIST from also compacting the other X chromosome. For an additional video that explains sex determination, X
expression of XIST and TSIX, look in the additional resources below.

X chromosome in human cancer

The X chromosome has been implicated in cancer in many contexts. In the additional resources for this module,
article that goes over these mechanisms, but we will highlight a few here.

Gain of whole X chromosomes

Gain of whole X chromosomes have been observed in many types of solid and hematopoietic (blood) cancers inc
cancer, leukemia, and prostate cancer. Gains of one arm of the X chromosome have also been observed in som
liver cancer. Gains of X chromosomes associated with cancer are gains of the active X chromosome; increases
chromosomes would not be expected to cause problems since most if not all gene expression from inactivated X

Dysregulated X chromosome inactivation

One of the reasons why dosage compensation is thought to be necessary is that studies have shown that when X
inactivation is altered, that can cause problems. For example, one study in ovarian cancer identified two subtype

properly regulated XCI and one with dysregulated XCI (Winham et al, Human Molecular Genetics 2019, link belo Resources). Patients with dysregulated XCI were shown to have a significantly shorter time to recurrence (time i regrow after treatment) and shorter overall survival time.

In many cancer types, males have a higher incidence. One idea of why this might be happening is a protective e multiple X chromosomes such that if there is a cancerous variant of a gene on one X chromosome, there is a cha chromosome might have the non-cancerous variant of that gene and expression can take place from that X chror there is typically only one X chromosome, so if there is a cancerous variant of a gene, it will certainly be expresse this idea is that females have X chromosome genes with even higher expression because they can escape X chr and a recent study showed that some of these escape genes are tumor suppressors (genes that prevent the dev of tumors) (paper below). In this study, the researchers looked for mutations X chromosome genes and measure difference between the sexes in a large cohort of cancer samples. They found six genes that escape X inactivati function mutations in a significant proportion of male cancer samples and were previously demonstrated to tumor

Journal Club: Tumor suppressor genes that escape from X-inactivation cor cancer sex bias

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5206905/ (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC52

When reading this paper, ask yourself the following study questions:

1. What is a tumor suppressor gene?
2. Which tumor suppressor genes did they study?
3. What was the data set they used to study the function of these genes in cancer?
4. What kinds of alterations of these genes did they look for?
5. Which cancers were these alterations associated with?
6. How did they measure X chromosome inactivation escape?
7. What was their overall conclusion or model for how they think this mechanism works in cancer?

Please go to the Journal Club assignment for this module to participate in the discussion. Here are those Perusa if you need them:

- Accessing Perusall through Canvas (https://www.youtube.com/watch?v=bs_Z_3wqib4) (Accessing Peru Canvas Video Transcript (https://docs.google.com/document/d/1ql6li6Au6ccO-xoTpQRM_ilF5Z6FMeGtbRG usp=sharing))
- Intro to Perusall (https://www.youtube.com/watch?v=M8bOP7yF_6I) (Perusall Introduction Video Transc (https://docs.google.com/document/d/1OPT_i7YrembK3518QiKaYcgClgsM-BRbuCCc7Y-BQXU/edit?usp=sharing)

Visualizing differences in gene expression: Plotting distributior

Once we get our list of differentially expressed genes, we often want to visualize the expression differences for in find interesting. These genes can be selected based on previously published studies about the biology of a giver example, we might choose to plot the expression of genes known to be involved in metastasis in primary tumors tumors. Genes can also be selected from the data set itself, such as plotting genes that are identified as differen software tools designed for that purpose. There are many plots that are used to compare values in sample group some popular ones we have used in this study. Typically, several plots are placed side by side to compare the dis points among several groups.

Box plot

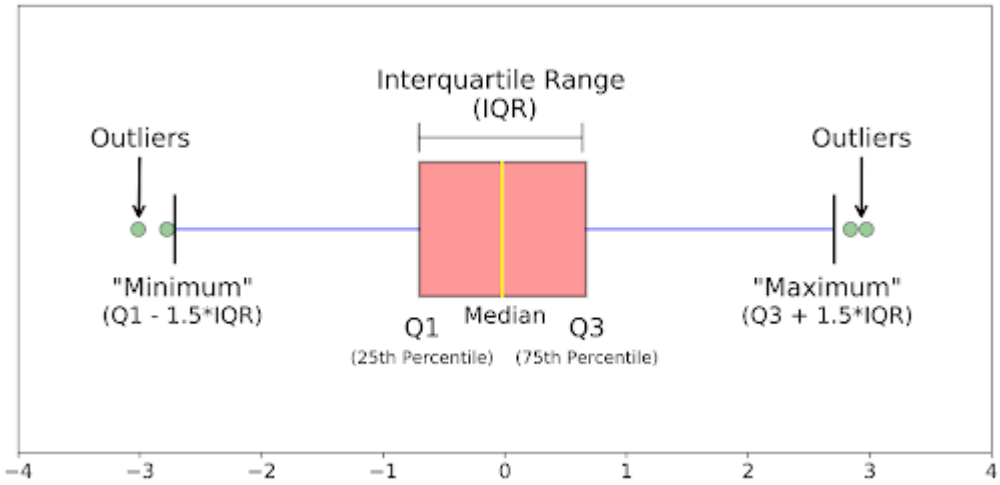

Figure: Parts of box plot. [Source: Towards Data Science, Understanding Boxplots (https://www.kdnuggets.com/2019/11/understanding-boxplots.html)]

A box plot displays a 5 number summary of the distribution of a specific measurement across a set of samples. A median of the values. A box is drawn between the 25th percentile (Q1) and the 75th percentile (Q3). Lines calle to show the distance between the minimum and maximum values of the range. In some cases, whiskers are dra minimum and maximum, while in other cases, outliers are detected and the minimum and maximum are determir those outliers. This type of visualization presents the distribution of the values without showing the specific obse sample. This visualization is often used when sample numbers are very large and values are evenly dispersed th between the minimum and maximum of the distribution.

Violin plot

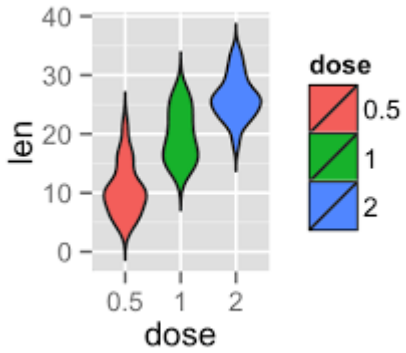

Figure: Violin plots showing the length of teeth after a dose of drug treatment [Source: (https://docs.google.com/document/d/1OPT_i7YrembK3518QiKaYcgCIgsM-BRbuCCc7Y-BQXU/edit?usp= ggplot2 violin plot : Quick start guide - R software and data visualization (http://www.sthda.co violin-plot-quick-start-guide-r-software-and-data-visualization)

Violin plots are similar to box plots in that they show the distribution, but the width of the plot is smoothed by a ke which shows the range at which there are many samples data points as wider. The example above shows only t

plots can also be marked with the median and percentiles. This visualization can be helpful when two distribution
range, but the distribution of where data points are found within the range. In the dose violin plot, the length at dc
similar, but you can see that the points are more evenly distributed in dose 1, but in dose 2 most of the points are

Jitter plot

A jitter plot is used to visualize the distribution and show the individual values, instead of just hte summary of the
values are plotted as dots along one axis and the dots are then shifted randomly along the other axis, which has
data-wise, allowing the dots not to overlap.

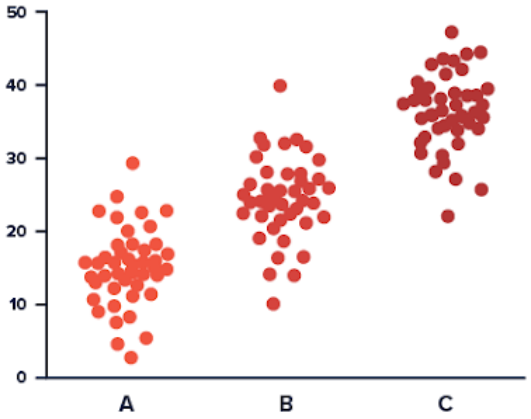

Figure: Example of jitter plot. [Source: **datavizproject, Jitter Plo (https://datavizproject.com/data-t**

Combination plots

Depending on what you want to convey with your visualizations, different kinds of plots are often combined to sho
of the distribution of values within a group. You can mix and match easily using the ggplot function in R as you w
section of this module.

Violin box plot

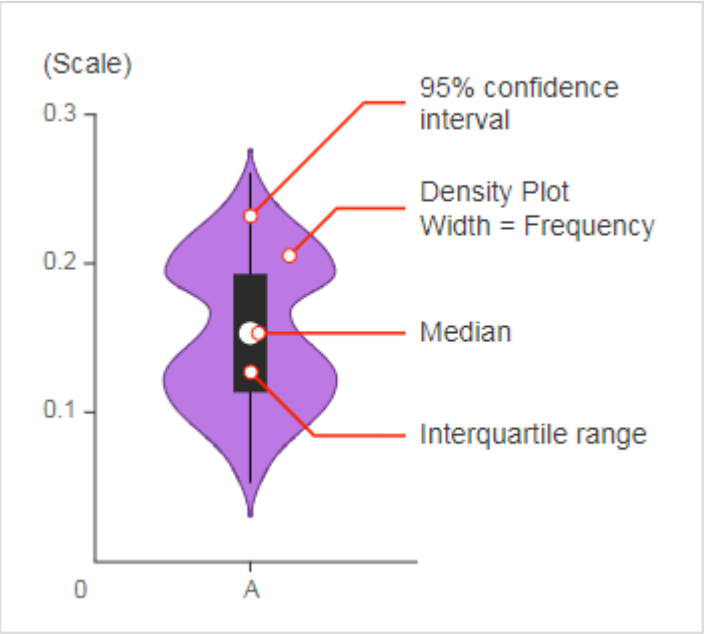

Figure: Parts of a violin plot [Source: **Infinity Insight, Violin Plots: What They Are and Why You Sh**
(https://datavizcatalogue.com/methods/violin_plot.html)]

Violin jitter box plot

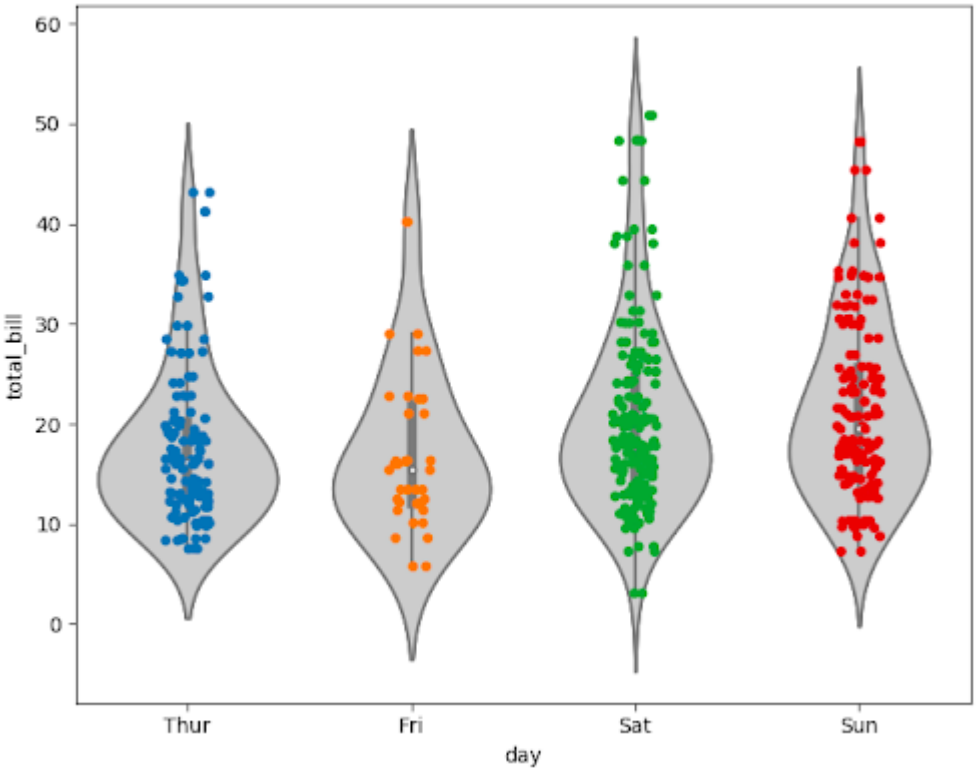

Figure: Violin plot with box plots and jitter plot added in. [Source: **Stack Overflow** (https://stackoverflow.com/questions/55797760/seaborn-stripplot-with-violin-plot-bars-in-front-of-points)

Visualize overlap between cell line groups: Upset Plot

Overlap between two groups of data is often represented as a Venn diagram, where overlap between groups is r
overlap between two overlaid shapes. As you can see in the figure below, as you increase the number of groups
representing overlap using Venn diagrams gets very complicated:

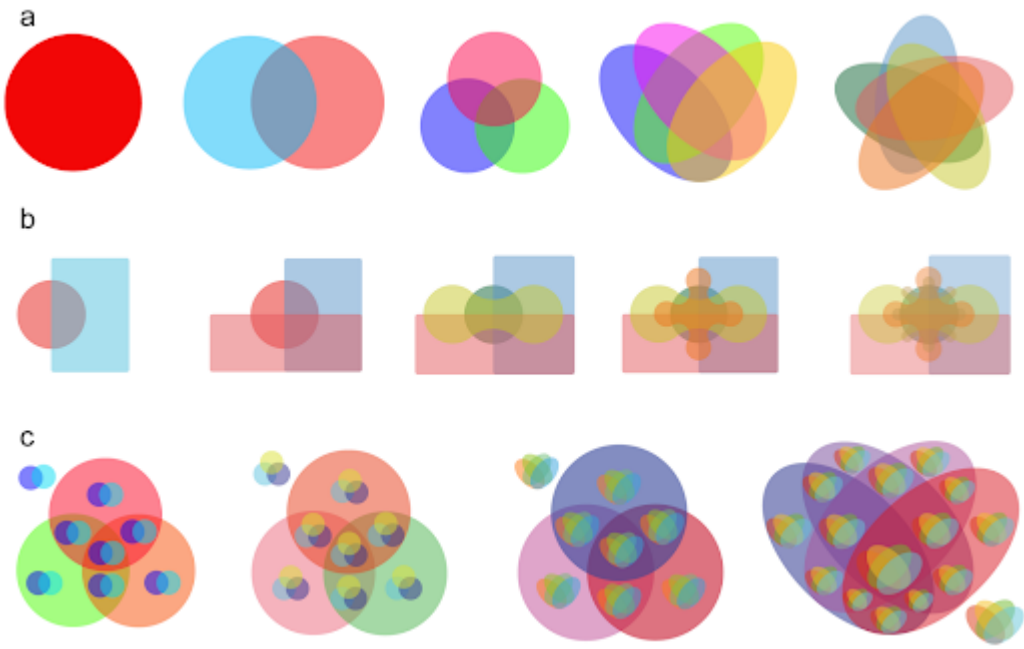

Figure: Increasing complexity of Venn diagrams **PLOS One, VennPainter (https://journals.plos.org id=10.1371/journal.pone.0154315)**

For this reason we are going to use another type of plot called an upset plot to represent overlap between differe lists from our 9 data sets. An upset plot uses bar graphs to show the values of intersections you would put on a \ lines to represent membership between multiple groups. In the example shown below, we have an upset plot tha membership to 5 different genres of a collection of movies: Crime, Thriller, Fantasy, Comedy, and Drama. The m lines to define intersections in groups; a single unconnected dot represents movies uniquely belonging to one ge multiple rows represent movies belonging to multiple genres. In this way, we can make observations like the higl that belong to two genres are Comedy and Drama (163 movies).

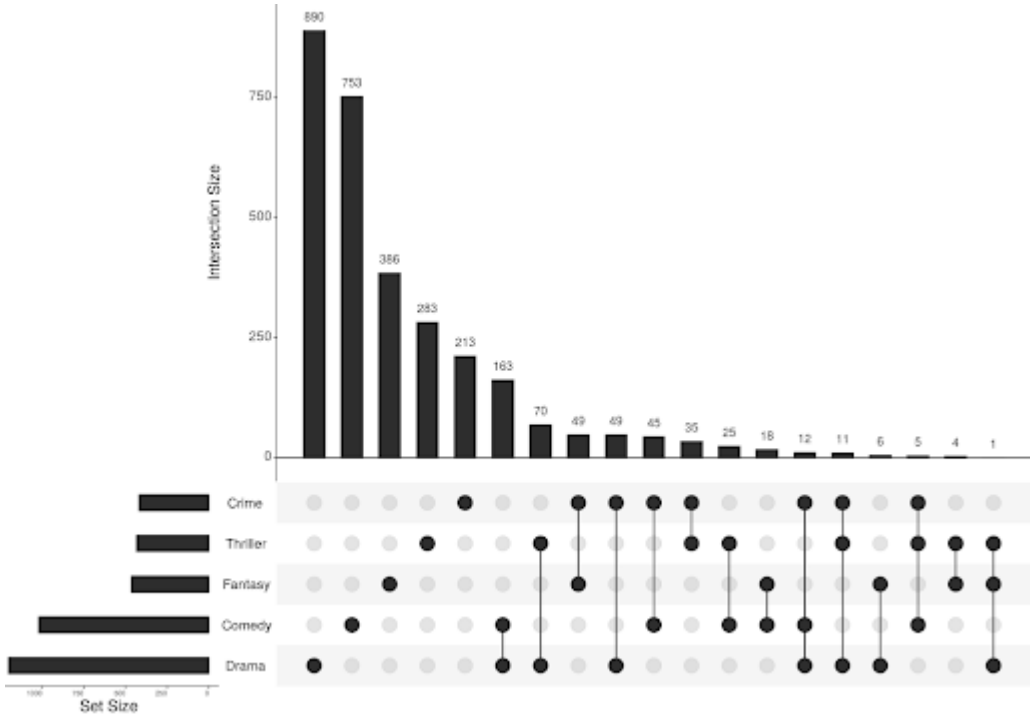

Figure: Upset plot showing the intersection of movie genres in a collection of movies.

This video describes another simple example of how to read an upset plot and shows the R function we will be us

UpSet Plots + R Demo (https://www.youtube.com/watch?v=n9MRCZxJOfk)

(https://www.youtube.com/watch?v=n9MRCZxJOfk)

UpSet Plots + R Demo transcript (https://docs.google.com/document/d/1f8LC25TK74VzUHpYG2hvId-KgSj1cX

XIST expression in healthy tissues

In our research project, we are studying XIST expression in cancer cells, but what does XIST look like in non-car
in different organs of the body? To find this out, let's make use of an incredible public genomics resource on hea

called the Genotype-Tissue Expression (GTEx) data set. The website for this project describes that human tissu
collected from 54 non-diseased tissue sites across nearly 1000 individuals and assayed with several types of ger
measurements including RNA sequencing and DNA sequencing (exome, whole genome). They have set up an e
application to allow people to look through the data set to find information relevant to their own studies. We are g
look at expression of XIST across healthy tissues.

Activity: Lookup XIST in GTEx

To see what XIST expression looks like in healthy tissues, look it up on GTEx and report your findings in your we

Instructions:

1. Open a browser and go to: **https://gtexportal.org/home/ (https://gtexportal.org/home/)**
2. Enter “XIST” in the “By gene ID” search box by “Browse”
3. First entry will match XIST exactly by gene ID
4. Scroll down to see how XIST is expressed across various healthy tissues

5. There are several buttons you can push at the top to change the way these violin box plots are displayed

- Switching subset to “Sex” splits the violin box plots by reported sex, showing the females in pink ar
  you can see that the ones that are high on the y-axis are females (all pink) but there is some variat
  really are

- If you switch scale to “Log”, the violin box plots from all the tissues will all be in a more comparable

This reinforces ideas most people don't think about (but not us, we are the Sex Chromosomes Lab!! ⭐):

1. There are sex chromosomes in cells in your whole body, not just your reproductive tissues!
2. Certain tissues have higher expression of sex chromosome genes than others
    - Tissue specific sex chromosome gene expression might play a role in sex differences in diseas tissues
3. Genes on the sex chromosome are regulated by many factors just like genes on the autosomes, so it's impor chromosome genes instead of ignoring them because of the added steps needed to account for sex chromos

Additional Resources

- Video showing more details about X chromosome inactivation
    - **X inactivation | Role of Tsix gene and its mechanism | Genetics | Akash Mitra (https://youtu.be/J_I**

(https://youtu.be/J_IDzKIAWjQ)
(**X Inactivation | Role of Tsix Gene and Its Mechanism Video Transcript**
(https://docs.google.com/document/d/1K5NYTX0bC9nZKBSHnwK1bYqZxmSsRRB8dCRaKlJ0Ae4/edit?usp=s

- X chromosome overview
    - **https://www.sciencedirect.com/science/article/pii/B9780123749840016508**
    (https://www.sciencedirect.com/science/article/pii/B9780123749840016508)
- X-Chromosome Genetics and Human Cancer
    - **https://www.nature.com/articles/nrc1413 (https://www.nature.com/articles/nrc1413)**
- The X chromosome in immune functions: when a chromosome makes the difference
    - **https://www.nature.com/articles/nri2815 (https://www.nature.com/articles/nri2815)**
- Gene content evolution on the X chromosome
    - **https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4590997/ (https://www.ncbi.nlm.nih.gov/pmc/articles/**

- Tumor suppressor genes that escape from XCI contribute to sex bias
  - https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5206905/ (https://www.ncbi.nlm.nih.gov/pmc/articles/
- Molecular signatures of X chromosome inactivation and associations with clinical outcomes in epithelial ovari
  - https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6625007/ (https://www.ncbi.nlm.nih.gov/pmc/articles/
- X chromosome reactivation
  - https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3293375/ (https://www.ncbi.nlm.nih.gov/pmc/articles/
  - https://www.cell.com/cell-systems/fulltext/S2405-4712(22)00403-3?_returnURL=https%3A%2F%2Flinkinghub.elsevier.com%2Fretrieve%2Fpii%2FS2405471222004033 (https://www.cell.com/cell-systems/fulltext/S2405-4712(22)00403-3?_returnURL=https%3A%2F%2Flinkinghub.elsevier.com%2Fretrieve%2Fpii%2FS2405471222004033%3Fshowa