1. Check out GSE189685 on nih.gov
   1. In which cell line, which organism are we?
   2. Which type of sequencer did they use?
   3. How many samples is data available for?
   4. Describe their experiment design. Which samples/conditions are available to us?
   5. Has the article been published?
2. Explore the counts dataframe you have
   1. Use head(), colnames(), rownames(), str()
      1. What’s wrong with the dataframe that is not compatible with transforming it into a matrix?
      2. Solve these problems (2)
      3. What is the extra steps the author did but was not a great idea ?

Clue/help:

* A matrix is an object that can only contain which type of data?
* Is it practical to have dashes in variable names? Try using the counts$ sign to access a column, see what happens
* Check the column class with str(counts)
* You can automatize change column class by using the type.convert() function
* counts <- type.convert(counts, as.is = TRUE) what is this function doing?
* Re-check the column class again, what changed?
* (iii) What are the usually is the name of the rows of a count matrix?

1. Create a high-level data structure using the DESeq2 package3.
   1. Create your coldata dataframe that contains your experiment design
   2. Create your DESeq object. What are we trying to compare? What should be in the design argument?
   3. Generate the lineal model

Clue/help:

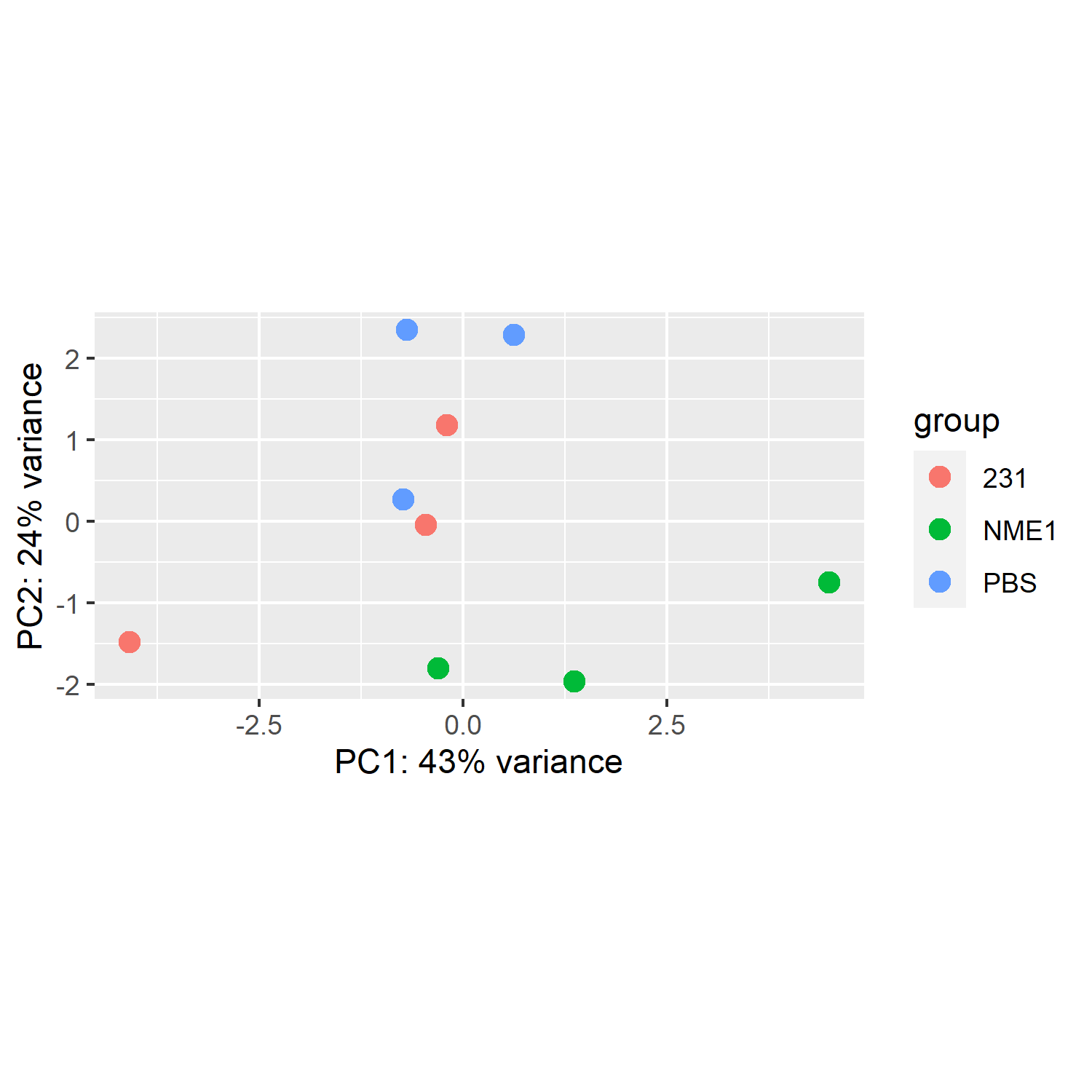
* Coldata contains sample condition, cell type, timepoint and other useful information
* Use the function rep() when you have to repeat the same string of character a lot of times
* Don’t forget that the row names of the coldata must match the column names of your count matrix. Verify that they do.

1. Check the quality of your data
   1. What can you say about the size factors?
   2. Obtain this graph and comment it.

Chart, histogram

Description automatically generated

1. Try resultsNames(), what is the output? Why? Do we need to relevel your data comparison?
2. Obtain these graphs:



Chart, scatter chart

Description automatically generated

1. What can you predict based on the different treatment vs ctrl?
2. What can you say about the replicates? Is it useful to represent them in a PCA?
3. Extract the results for the NME1 vs PBS comparison and obtain these graphs of normalized raw counts of the two isoforms of the catalytic subunit of AMPK

Chart, bar chart

Description automatically generatedChart, bar chart

Description automatically generated

* 1. What can you say about the difference in treatment between the samples for AMPKa1 expression?
  2. What can you say about the AMPKa2 expression in the different samples?

1. Save your dds object and your res\_tbl