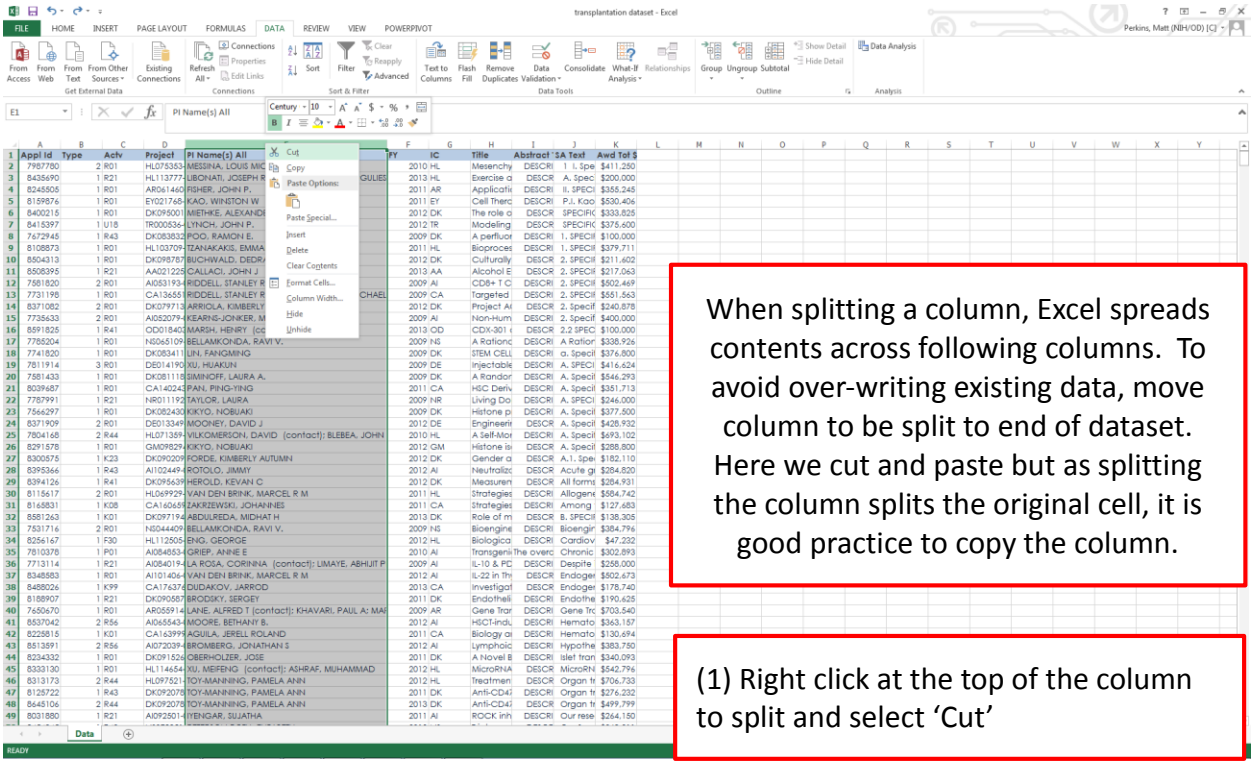
OPA Excel Tips: Splitting cell values into columns

Excel recognizes comma separated value (csv) files and automatically places the values in each row into separate columns, however some file formats and data fields use different characters, or a set number of characters to split the file up into different variables.

The 'Text to Columns' function within Excel is very useful for converting a dataset consisting of a single column into multiple variables, and for splitting single variables. Another option for the latter is to use the LEFT and RIGHT functions (see separate OPA Excel Tip).

Example 1: Splitting PI names into separate columns.

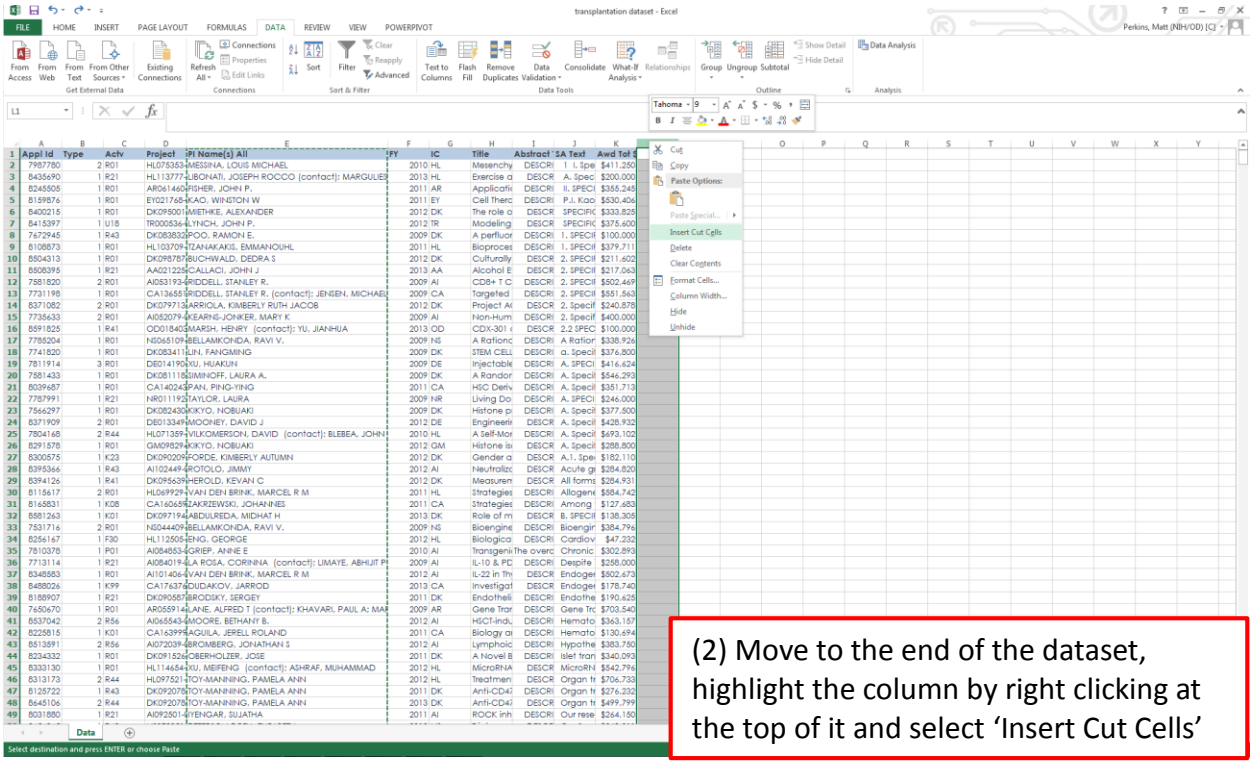
In this example the Transplantation dataset is used. The data is already in columns but the PI column contains multiple PIs. The 'Text to Columns' tool can be used to split the column up into individual PIs.



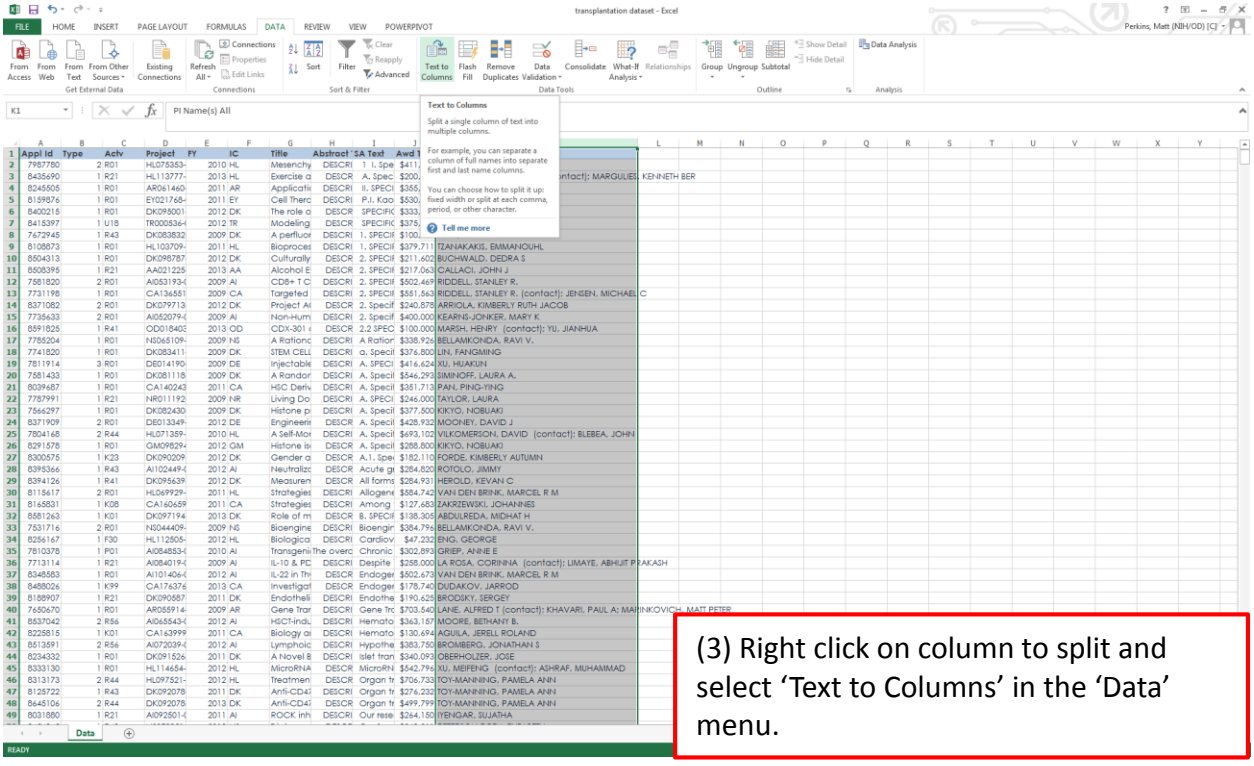
When splitting a column, Excel spreads contents across following columns. To avoid over-writing existing data, move column to be split to end of dataset. Here we cut and paste but as splitting the column splits the original cell, it is good practice to copy the column.

(1) Right click at the top of the column to split and select 'Cut'

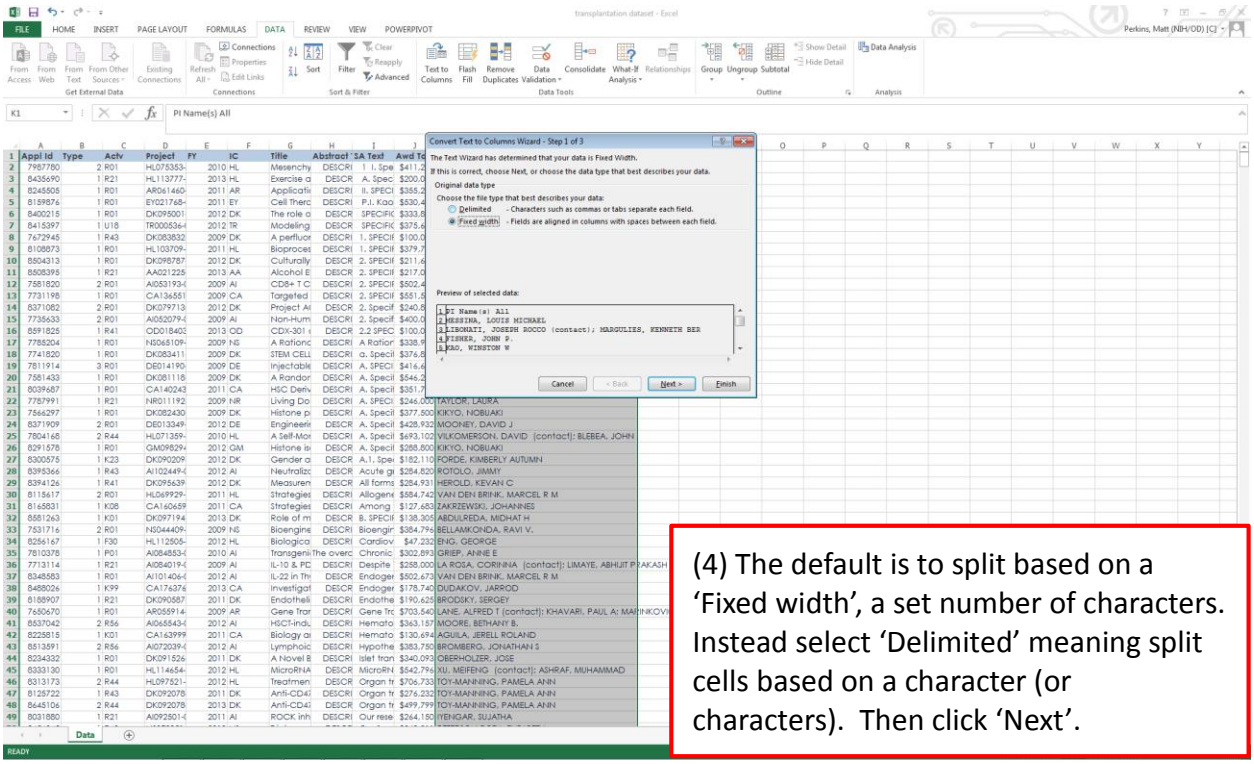
National Institutes of Health
Office of Portfolio Analysis



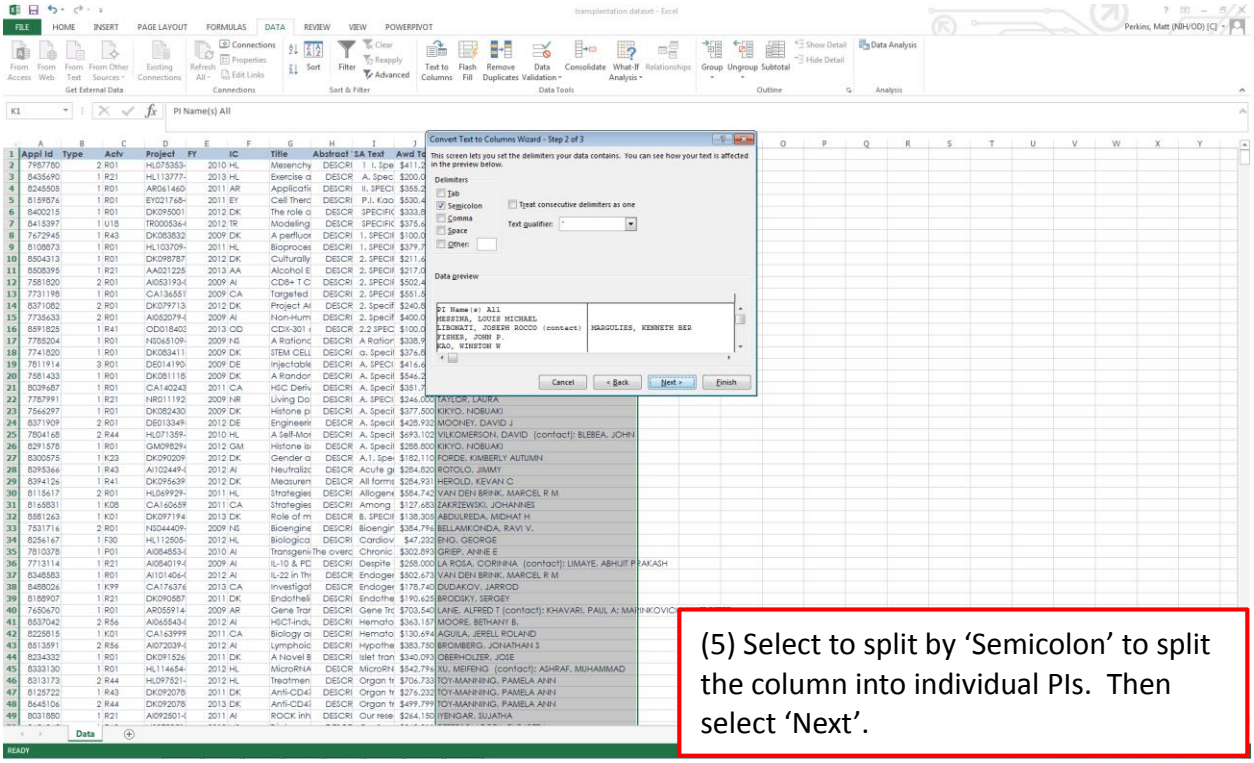
(2) Move to the end of the dataset, highlight the column by right clicking at the top of it and select ‘Insert Cut Cells’



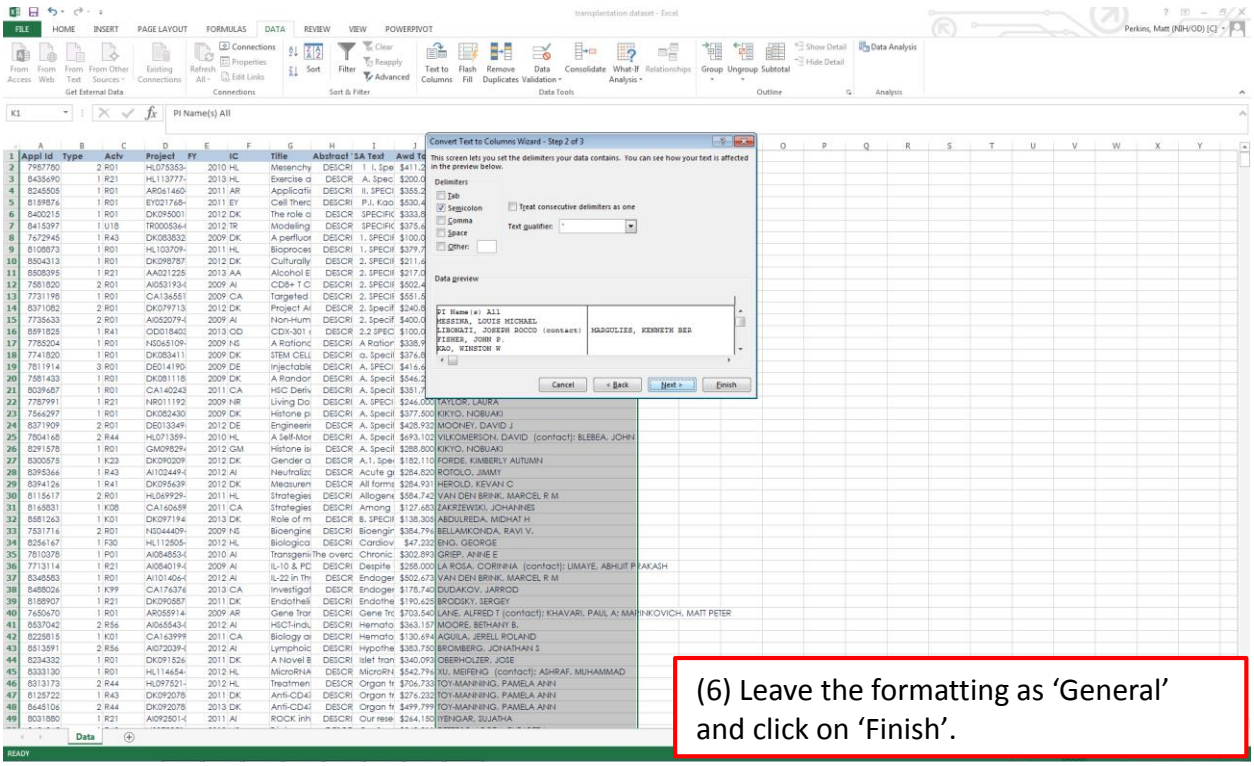
(3) Right click on column to split and select ‘Text to Columns’ in the ‘Data’ menu.



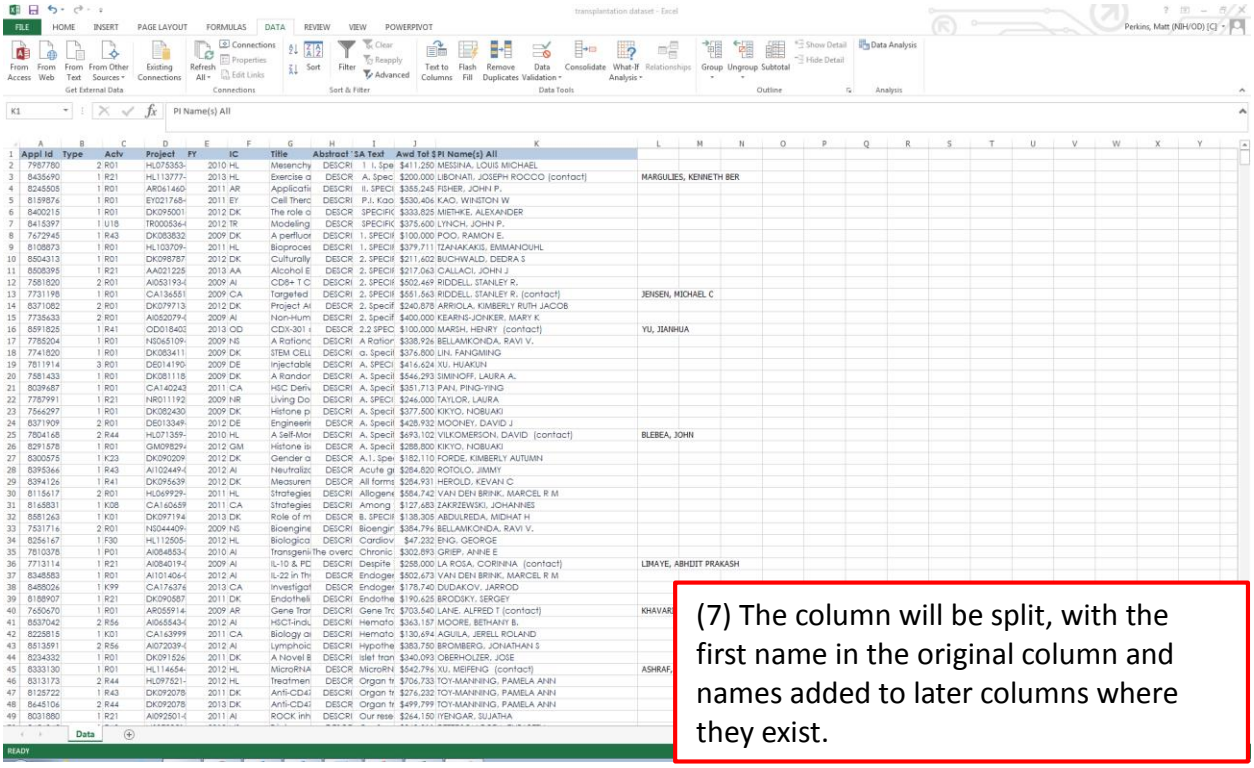
(4) The default is to split based on a 'Fixed width', a set number of characters. Instead select 'Delimited' meaning split cells based on a character (or characters). Then click 'Next'.



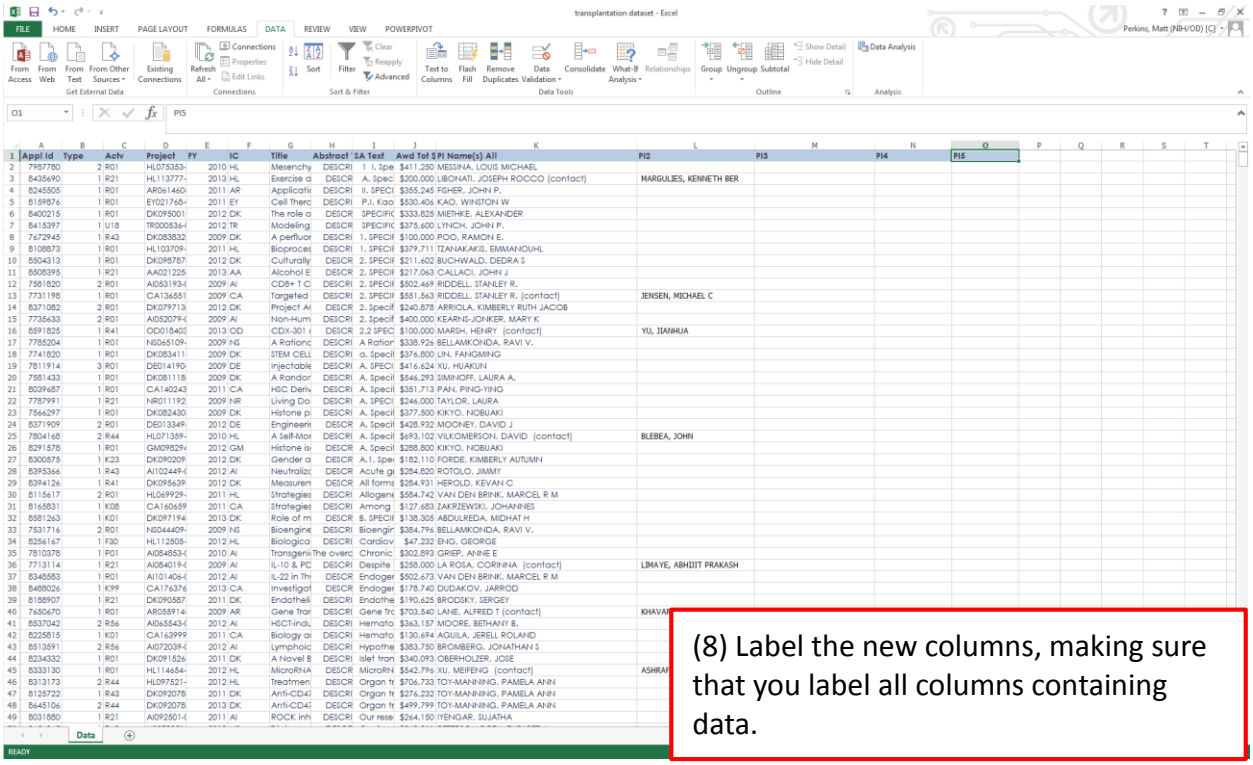
(5) Select to split by 'Semicolon' to split the column into individual PIs. Then select 'Next'.



(6) Leave the formatting as 'General' and click on 'Finish'.



(7) The column will be split, with the first name in the original column and names added to later columns where they exist.



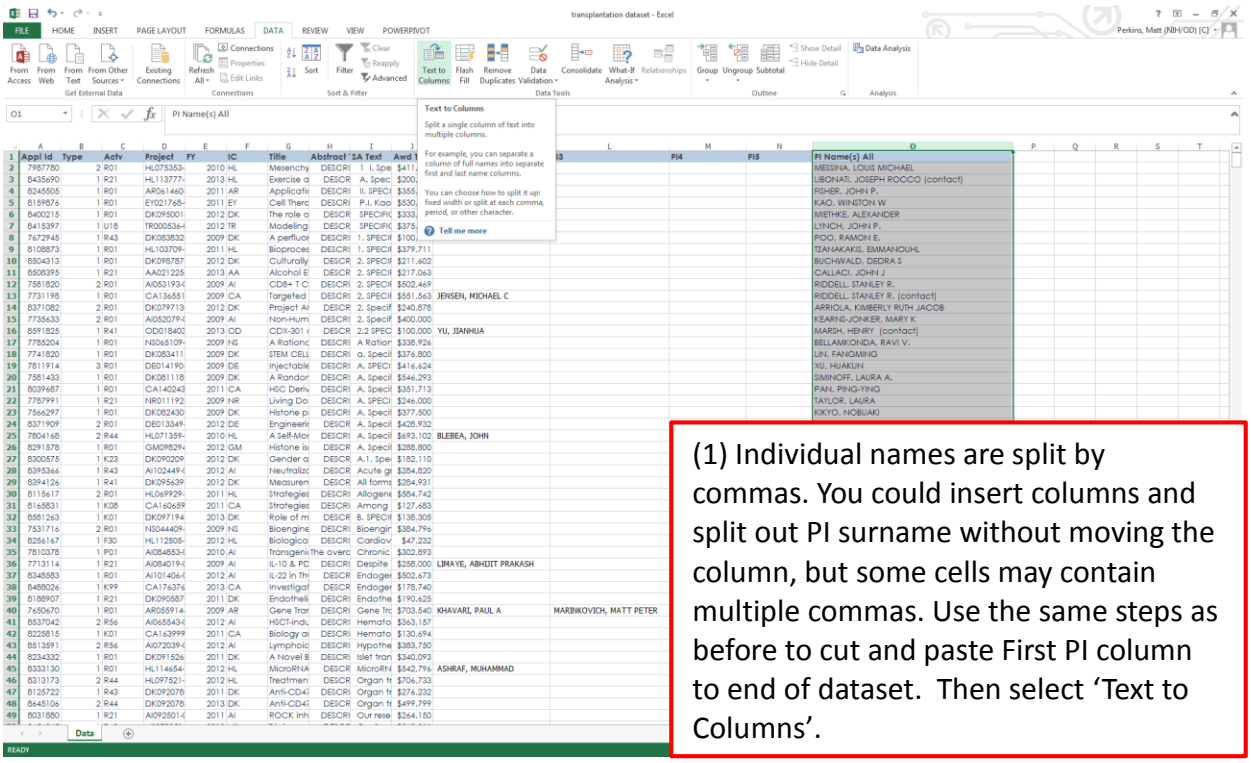
(8) Label the new columns, making sure that you label all columns containing data.

Example 2: Extracting surname from PI name

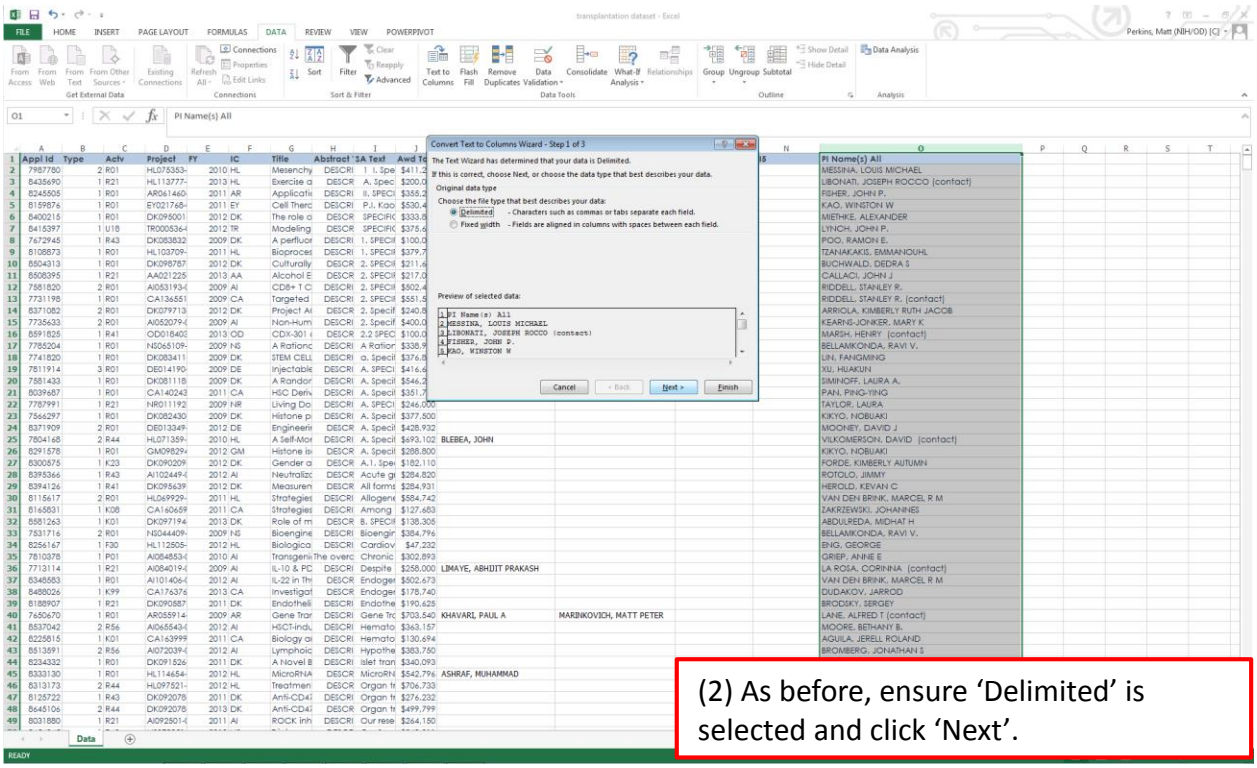
In this example the Transplantation dataset is used.

In the example above, the PI names were split into separate columns. But it may be that you are interested in just the PI surname. 'Text to Columns' can be used for this too.

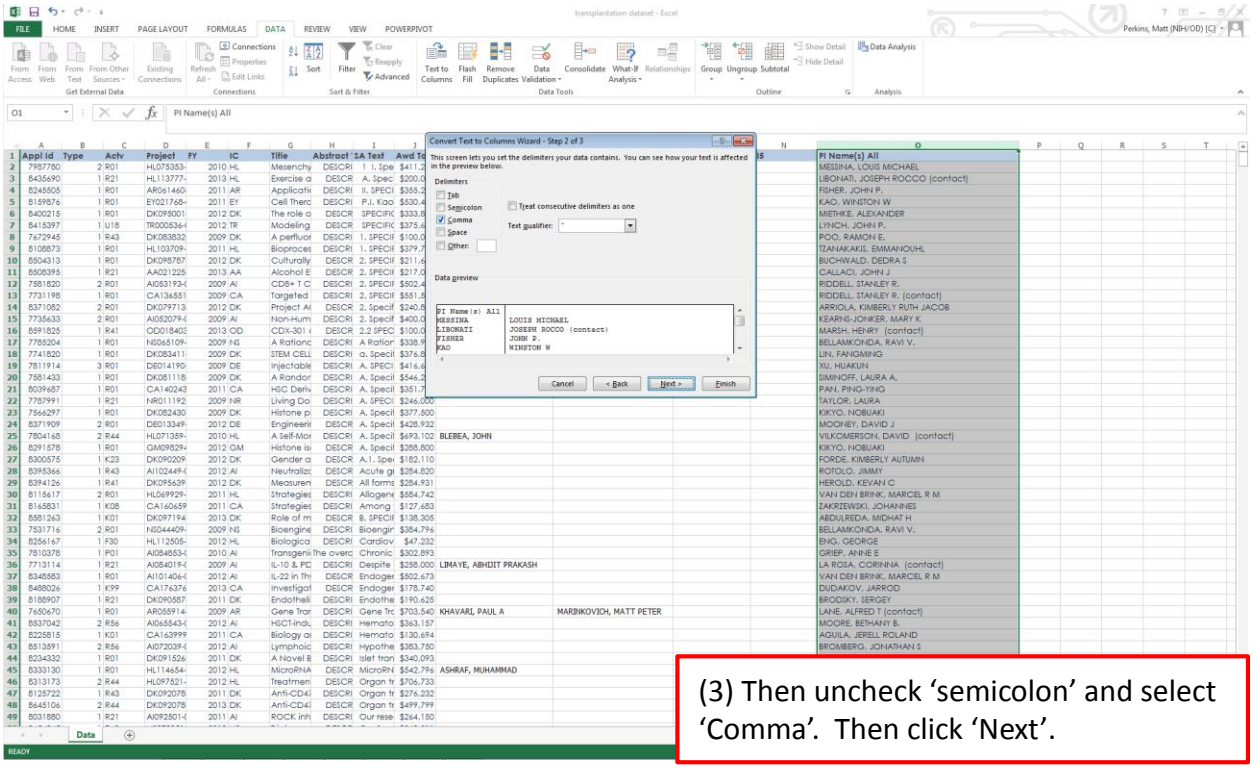
An alternative would be to use LEFT and SEARCH to extract just the text up to the ',' which would leave the existing cell intact. (see help for more details)



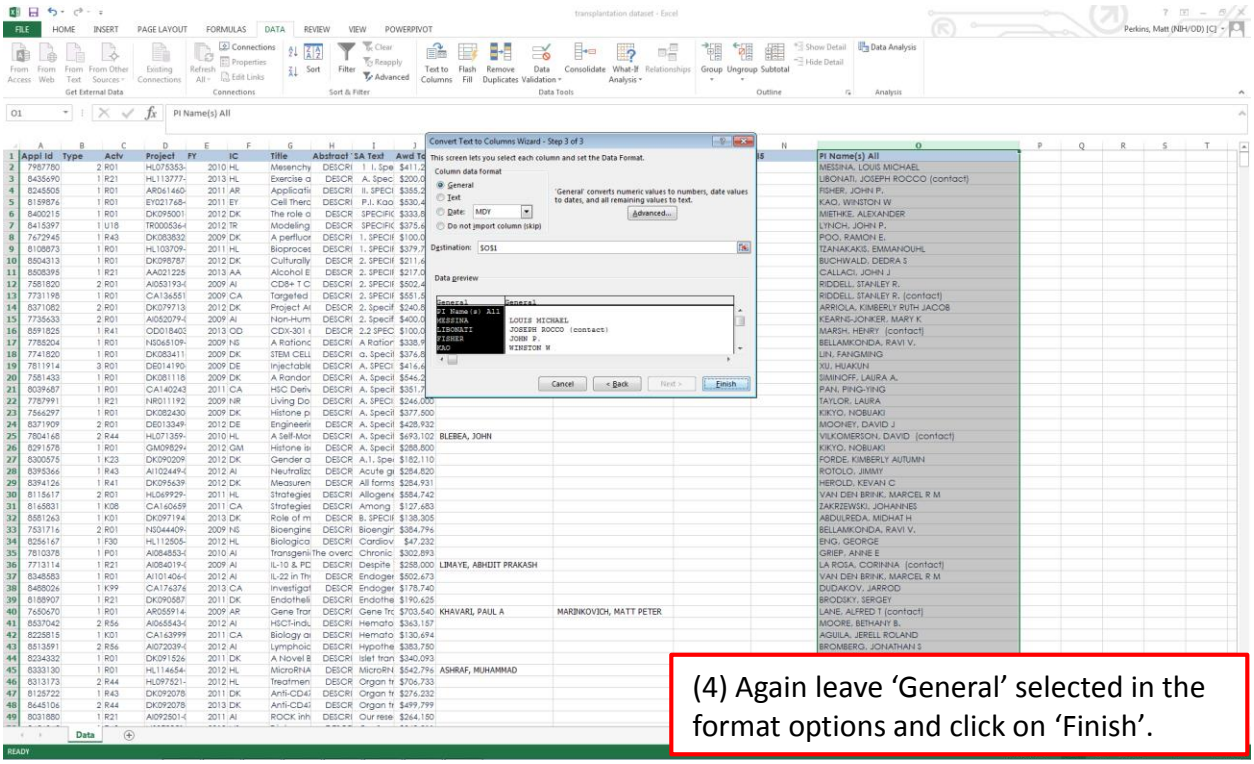
(1) Individual names are split by commas. You could insert columns and split out PI surname without moving the column, but some cells may contain multiple commas. Use the same steps as before to cut and paste First PI column to end of dataset. Then select 'Text to Columns'.

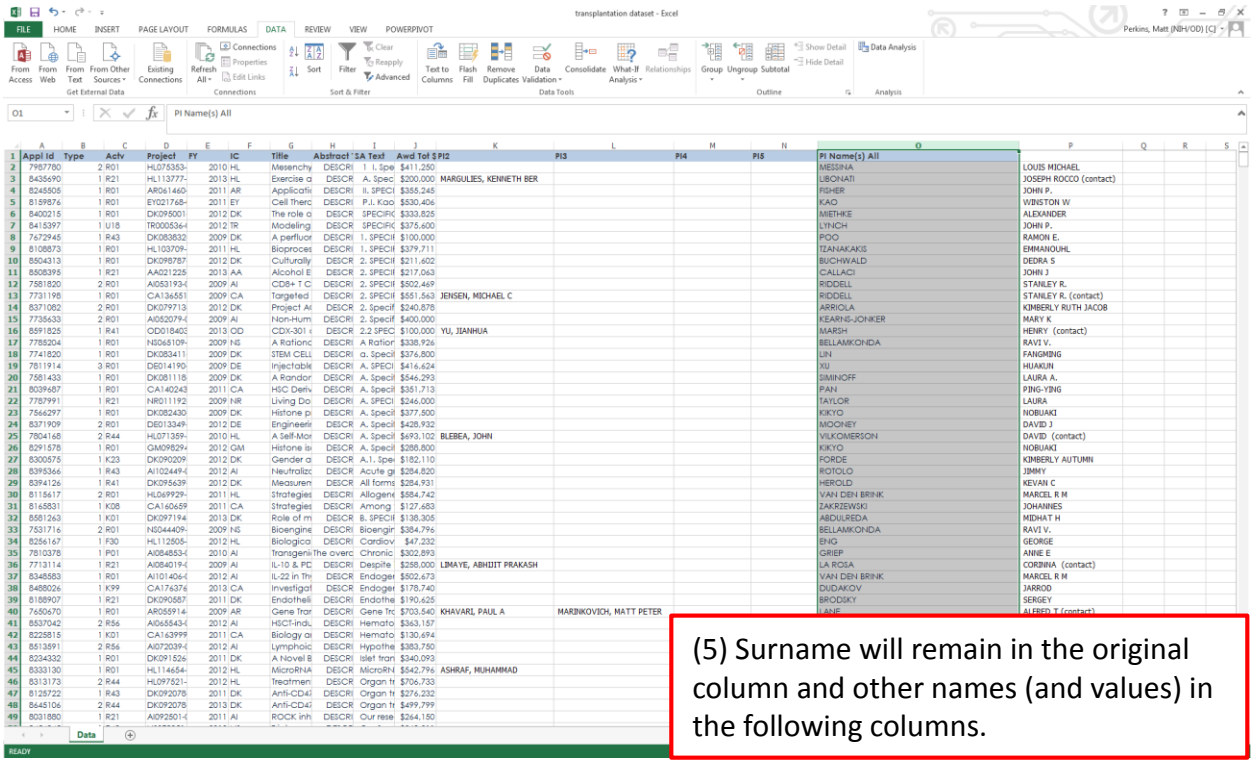


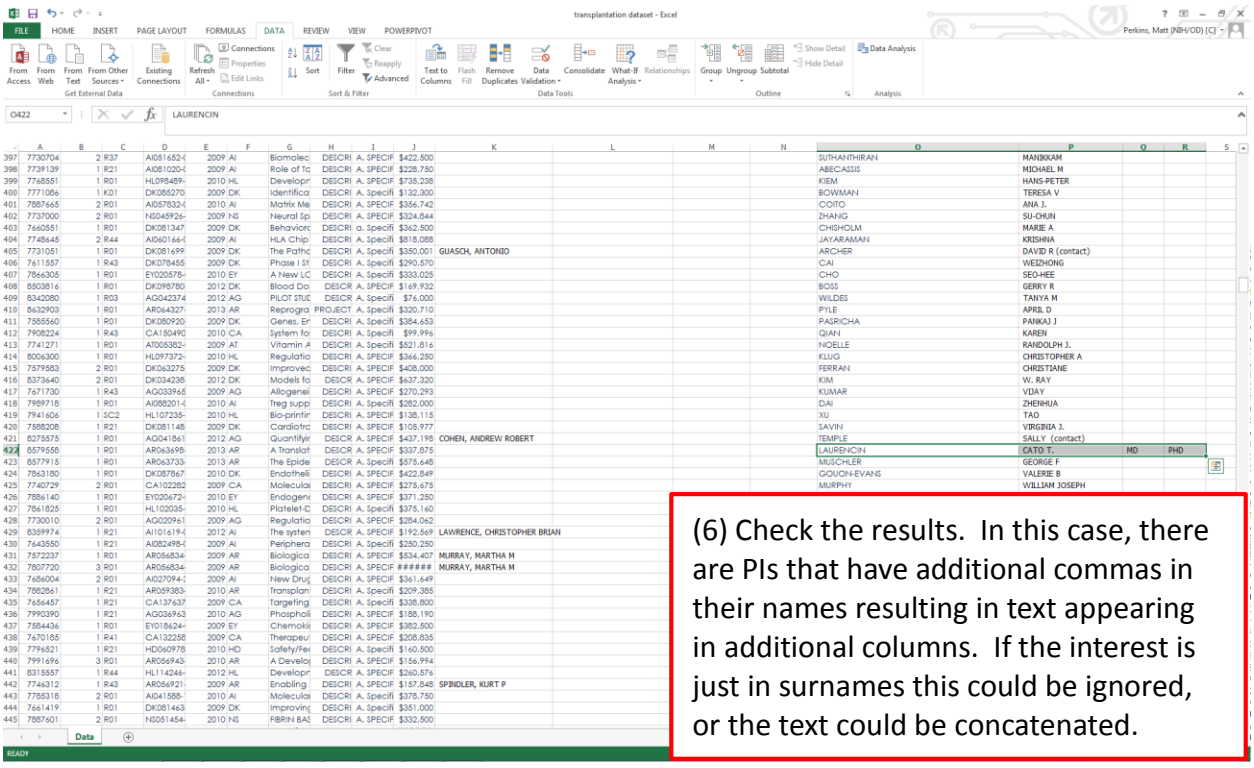
(2) As before, ensure 'Delimited' is selected and click 'Next'.



(3) Then uncheck 'semicolon' and select 'Comma'. Then click 'Next'.


Convert Text to Columns Wizard - Step 3 of 3
This screen lets you select each column and set the Data Format.
Column data format
General
Text
Date: MDY
Do not import column (skip)
'General' converts numeric values to numbers, date values to dates, and all remaining values to text.
Advanced...
Destination: O1
Data preview
General
PI Name(s) All
MESSINA
LIBONATI
FISHER
KAO
LOUIS MICHAEL
JOSEPH ROCCO (contact)
JOHN P.
WINSTON W
Cancel
< Back
Next >
Finish
(4) Again leave 'General' selected in the format options and click on 'Finish'.


(5) Surname will remain in the original column and other names (and values) in the following columns.



(6) Check the results. In this case, there are PIs that have additional commas in their names resulting in text appearing in additional columns. If the interest is just in surnames this could be ignored, or the text could be concatenated.

More help

Click on the '?' in the top left (as below) and search for 'Text to Columns'.