

Excel Tips and Tricks

March 12, 2020

Bioinformatics and Research Computing

Whitehead Institute

http://barc.wi.mit.edu/hot_topics/



WHITEHEAD INSTITUTE



You will Learn:

- General tips and tricks in navigation, filtering, and selection
 - Keep input data types
 - Keep headers visible as you scroll
 - Keyboard navigation shortcuts
 - Quickly identify patterns within a column
 - Split items within single column
 - Highlight cells by their values
- Using formulas to work with data
- Merge multiple tables with vlookup



Problem: Excel changes my input data

Month: JAN to DEC

a number contains the letter E or
number contains more characters
than can be displayed

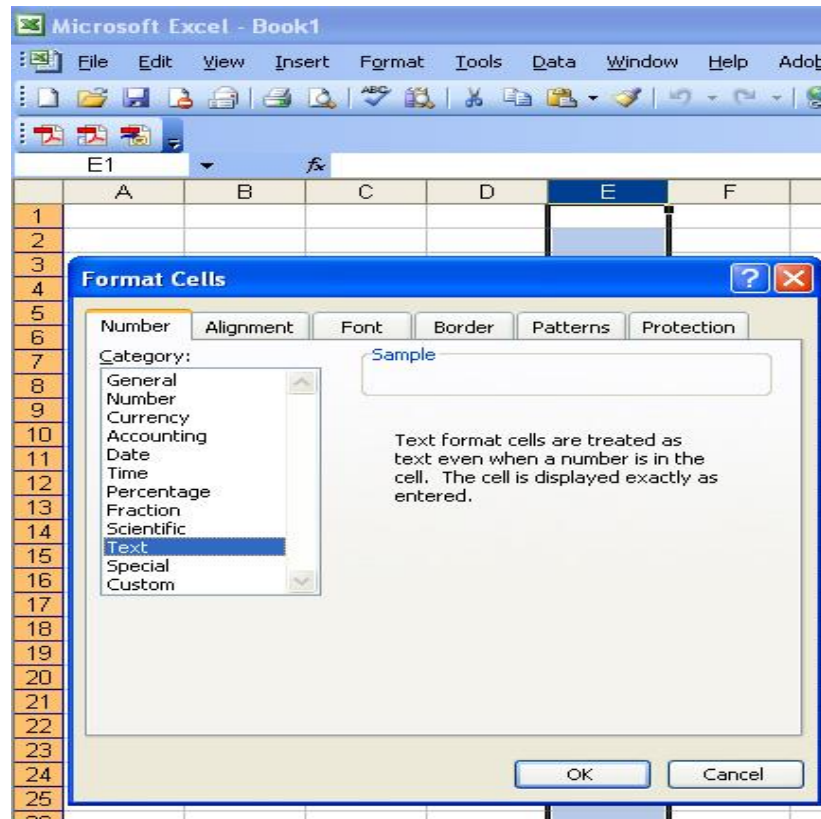
	A	B
1	ID	Default format
2	SEP2	2-Sep
3	2310009E13	2.31E+19
4	0223345	223345
5		
6		

a number contains leading zeros

Note: Copy from HTML page: use paste special -> Unicode text

Solution: Change formatting to "Text"

1. Open file with plain text
2. Before you populate the data into Excel, select the cells that will contain the gene symbols, right-click and choose **Format Cells > Text > OK**
3. Copy and paste data



Sort & Filter button: Easy to spot patterns

- Temporarily hide repetitive elements
- Sort the rest of the elements in order

Sort & Filter

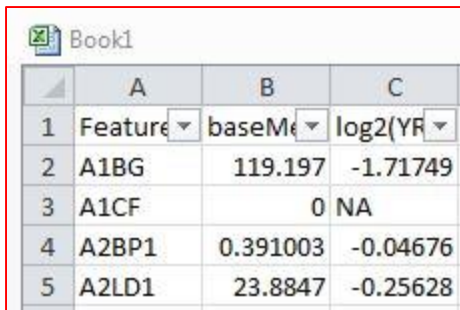
The screenshot shows the Microsoft Excel interface. The ribbon is set to the 'Home' tab, and the 'Sort & Filter' button is circled in red. The task pane on the left is open, showing the 'Sort & Filter' options. A red arrow points to the 'December' filter in the 'Text Filters' section. The spreadsheet data is visible below the ribbon, with columns labeled A through S and rows containing numerical and text data.

Feature.ID	baseM	log2(YR)	lfcSE	stat	pvalue	padj	CEU_N	CEU_N	YRI_N	YRI_N	CEU_N	CEU_N	YRI_N	YRI_N	CEU_N	CEU_N	YRI_N	YRI_N	CEU_N	CEU_N	YRI_N	YRI_N
1	119.197	-1.71749	0.365668	-4.67696	2.91E-06	0.000354	236	312	34	21	160.74	243.99	39.2	32.86	6.98	7.33	6.03	5.94	0	0	0	0
0	NA	NA	NA	NA	NA	NA	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0.391003	-0.04676	0.125607	-0.33845	0.735027	NA	NA	0	2	0	0	1.56	0	0	0	-1.09	-1.07	-1.09	-1.09	0	0	0	0
23.8847	-0.25628	0.390629	-0.66022	0.509112	0.824137	36	43	8	18	24.52	33.63	9.22	28.17	4.56	4.67	4.33	4.4	0	0	0	0	
10.2322	0.321001	0.30687	1.11826	0.263457	0.649452	9	3	1	20	6.13	2.35	1.15	31.3	3.08	3	2.98	3.43	0	0	0	0	
1.46147	-0.00963	0.171605	-0.05646	0.954972	NA	NA	0	4	1	1	0	3.13	1.15	1.56	0.5	0.56	0.53	0.53	0	0	0	0
591.392	-0.25942	0.335129	-0.79365	0.427401	0.77935	208	1911	178	335	141.67	1494.41	205.21	524.27	7.71	10.02	8.03	8.92	0	0	0	0	
0	NA	NA	NA	NA	NA	NA	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0.782006	-0.07462	0.127875	-0.54963	0.582575	NA	NA	0	4	0	0	0	3.13	0	0	-0.37	-0.32	-0.37	-0.37	0	0	0	0
384.481	0.469684	0.314927	1.49144	0.135846	0.488014	431	413	487	230	293.55	322.97	561.45	359.95	8.32	8.41	8.93	8.51	0	0	0	0	
408.579	-0.53154	0.323797	-1.64102	0.100794	0.425495	731	648	348	146	497.88	506.74	401.2	228.49	8.85	8.86	8.64	8.13	0	0	0	0	
0.391003	-0.04676	0.125607	-0.33845	0.735027	NA	NA	0	2	0	0	1.56	0	0	0	-1.09	-1.07	-1.09	-1.09	0	0	0	0
0.195502	-0.014	0.124404	-0.12732	0.898691	NA	NA	0	1	0	0	0.78	0	0	0	-1.74	-1.73	-1.74	-1.74	0	0	0	0
0	NA	NA	NA	NA	NA	NA	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0.391003	-0.04676	0.125607	-0.33845	0.735027	NA	NA	0	2	0	0	1.56	0	0	0	-1.09	-1.07	-1.09	-1.09	0	0	0	0

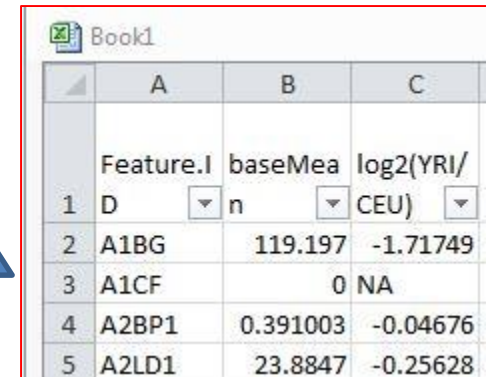


Problem: End of cell contents is missing

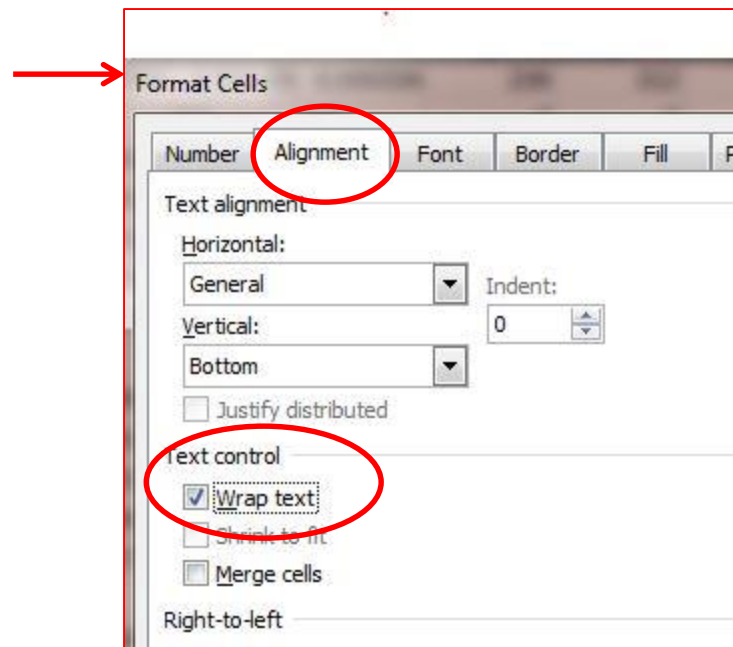
Solution: Wrap text



	A	B	C
1	Feature	baseMea	log2(YR
2	A1BG	119.197	-1.71749
3	A1CF	0	NA
4	A2BP1	0.391003	-0.04676
5	A2LD1	23.8847	-0.25628



	A	B	C
	Feature.I	baseMea	log2(YRI/
1	D	n	CEU)
2	A1BG	119.197	-1.71749
3	A1CF	0	NA
4	A2BP1	0.391003	-0.04676
5	A2LD1	23.8847	-0.25628



Split items within single column: example 1

Text to Columns

Convert Text to Columns Wizard - Step 2 of 3

This screen lets you set the delimiters your data contains. You can see how your text is affected in the preview below.

Delimiters

- Tab
- Semicolon
- Comma
- Space
- Other: -

Treat consecutive delimiters as one

Text qualifier: -

Data preview

TCGA	A8	A07B	01A	11R	A00Z	07
TCGA	A8	A08B	01A	11R	A00Z	07
TCGA	A8	A08P	01A	11R	A00Z	07
TCGA	A8	A094	01A	11R	A00Z	07
TCGA	AO	A03O	01A	11R	A00Z	07
TCGA	A2	A04P	01A	31R	A034	07
TCGA	A2	A04Q	01A	21R	A034	07
TCGA	A2	A04T	01A	21R	A034	07
TCGA	A2	A04V	01A	21R	A034	07
TCGA	AN	A041	01A	11R	A034	07
TCGA	AN	A046	01A	21R	A034	07
TCGA	AN	A04A	01A	21R	A034	07
TCGA	AN	A04C	01A	21R	A034	07
TCGA	AN	A04D	01A	21R	A034	07
TCGA	AO	A03T	01A	21R	A034	07
TCGA	AQ	A04J	01A	02R	A034	07
TCGA	A7	A0CH	01A	21R	A00Z	07
TCGA	A8	A06T	01A	11R	A00Z	07
TCGA	A8	A07F	01A	11R	A00Z	07
TCGA	A8	A081	01A	11R	A00Z	07

	A	B	C	D	E	F	G	H
1	TCGA	A8	A07B	01A	11R	A00Z	07	
2	TCGA	A8	A08B	01A	11R	A00Z	07	
3	TCGA	A8	A08P	01A	11R	A00Z	07	
4	TCGA	A8	A094	01A	11R	A00Z	07	
5	TCGA	AO	A03O	01A	11R	A00Z	07	
6	TCGA	A2	A04P	01A	31R	A034	07	
7	TCGA	A2	A04Q	01A	21R	A034	07	
8	TCGA	A2	A04T	01A	21R	A034	07	
9	TCGA	A2	A04V	01A	21R	A034	07	
10	TCGA	AN	A041	01A	11R	A034	07	
11	TCGA	AN	A046	01A	21R	A034	07	
12	TCGA	AN	A04A	01A	21R	A034	07	
13	TCGA	AN	A04C	01A	21R	A034	07	
14	TCGA	AN	A04D	01A	21R	A034	07	
15	TCGA	AQ	A03T	01A	21R	A034	07	
16	TCGA	AQ	A04J	01A	02R	A034	07	
17	TCGA	A7	A0CH	01A	21R	A00Z	07	
18	TCGA	A8	A06T	01A	11R	A00Z	07	
19	TCGA	A8	A07F	01A	11R	A00Z	07	

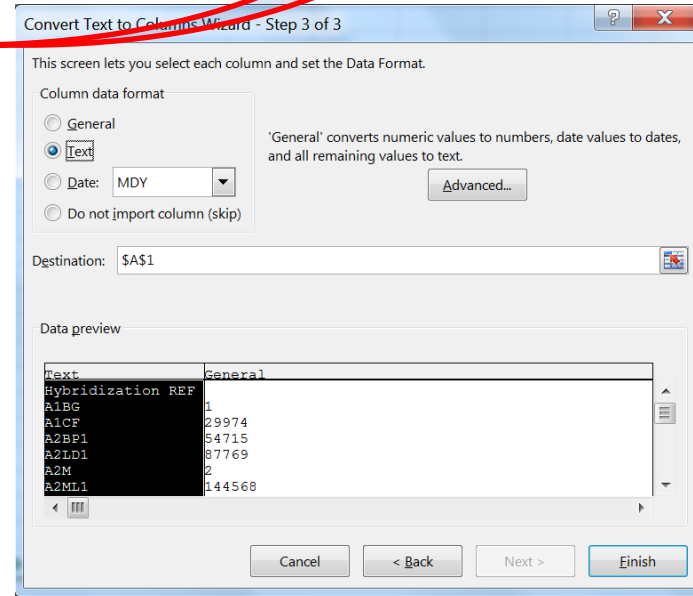
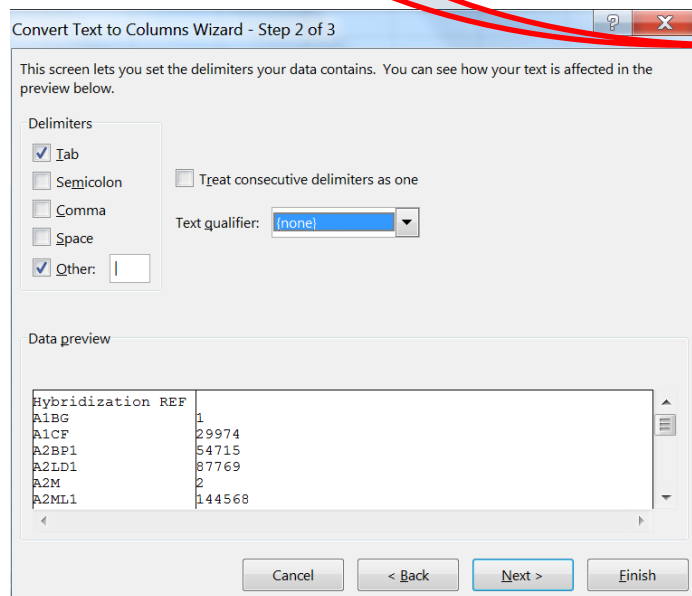


Split items within single column: example 2

	A
15728	SEPN1 57190
15729	SEPP1 6414
15730	SEPSECS 51091
15731	SEPT10 151011
15732	SEPT11 55752
15733	SEPT1 1731
15734	SEPT12 124404
15735	SEPT14 346288
15736	SEPT2 4735
15737	SEPT3 55964
15738	SEPT4 5414
15739	SEPT5 5413

	A	B	C
15728	SEPN1	57190	
15729	SEPP1	6414	
15730	SEPSECS	51091	
15731	SEPT10	151011	
15732	SEPT11	55752	
15733	SEPT1	1731	
15734	SEPT12	124404	
15735	SEPT14	346288	
15736	SEPT2	4735	
15737	SEPT3	55964	
15738	SEPT4	5414	
15739	SEPT5	5413	

Text to Columns



Merge multiple columns

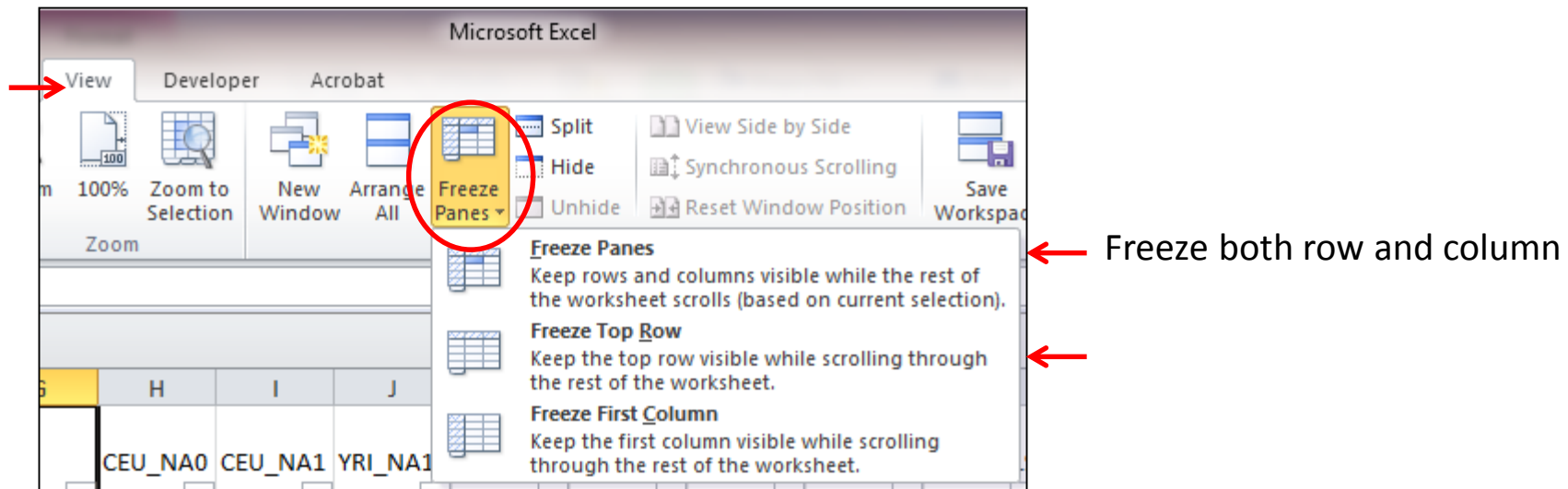
The screenshot shows the Microsoft Excel interface with the Data tab selected. The formula bar displays the formula `=CONCATENATE(A1, "-", B1, "-", C1)`, which is circled in red. The spreadsheet below shows columns A through G with data, and column H contains the concatenated results.

	A	B	C	D	E	F	G	H	I	J
1	TCGA	A8	A07B	01A	11R	A00Z	7	TCGA-A8-A07B		
2	TCGA	A8	A08B	01A	11R	A00Z	7	TCGA-A8-A08B		
3	TCGA	A8	A08P	01A	11R	A00Z	7	TCGA-A8-A08P		
4	TCGA	A8	A094	01A	11R	A00Z	7	TCGA-A8-A094		
5	TCGA	AO	A03O	01A	11R	A00Z	7	TCGA-AO-A03O		
6	TCGA	A2	A04P	01A	31R	A034	7	TCGA-A2-A04P		
7	TCGA	A2	A04Q	01A	21R	A034	7	TCGA-A2-A04Q		
8	TCGA	A2	A04T	01A	21R	A034	7	TCGA-A2-A04T		
9	TCGA	A2	A04V	01A	21R	A034	7	TCGA-A2-A04V		
10	TCGA	AN	A041	01A	11R	A034	7	TCGA-AN-A041		
11	TCGA	AN	A046	01A	21R	A034	7	TCGA-AN-A046		
12	TCGA	AN	A04A	01A	21R	A034	7	TCGA-AN-A04A		
13	TCGA	AN	A04C	01A	21R	A034	7	TCGA-AN-A04C		
14	TCGA	AN	A04D	01A	21R	A034	7	TCGA-AN-A04D		



Keep headers visible as you scroll

- Choose **View > Freeze Panes -> Freeze top Row**
- To unfreeze frozen rows, choose **View > Freeze Panes > Unfreeze Panes**



Keyboard navigation shortcuts

- PC:
 - First cell in the active column: **Ctrl + up arrow**
 - Last cell in the active column: **Ctrl + down arrow**
 - First cell in the active row: **Ctrl + left arrow**
 - Last cell in the active row: **Ctrl + right arrow**
 - The very first cell (A1): **Ctrl+ Home**
 - Bottom right corner: **Ctrl+ End**
- MAC:
 - use either **Command** or **Control**



Keyboard selection shortcuts

- Selecting a large range of cells can be a pain. Here's one trick that could help.
 - To start the selection of a large matrix, **start by selecting the bottom right cell** of the desired matrix (**Ctrl + End**)
 - Use "**Ctrl + Shift + up arrow**" to select everything above the original cell.
 - Use "**Ctrl + Shift + left arrow**" to select everything to the left of the original cell.
 - Use "**Shift + down arrow**" to move down one row (to avoid the header line, if present).
 - Use "**Shift + right arrow**" to move to the right one column (to avoid any ID column(s), if present)

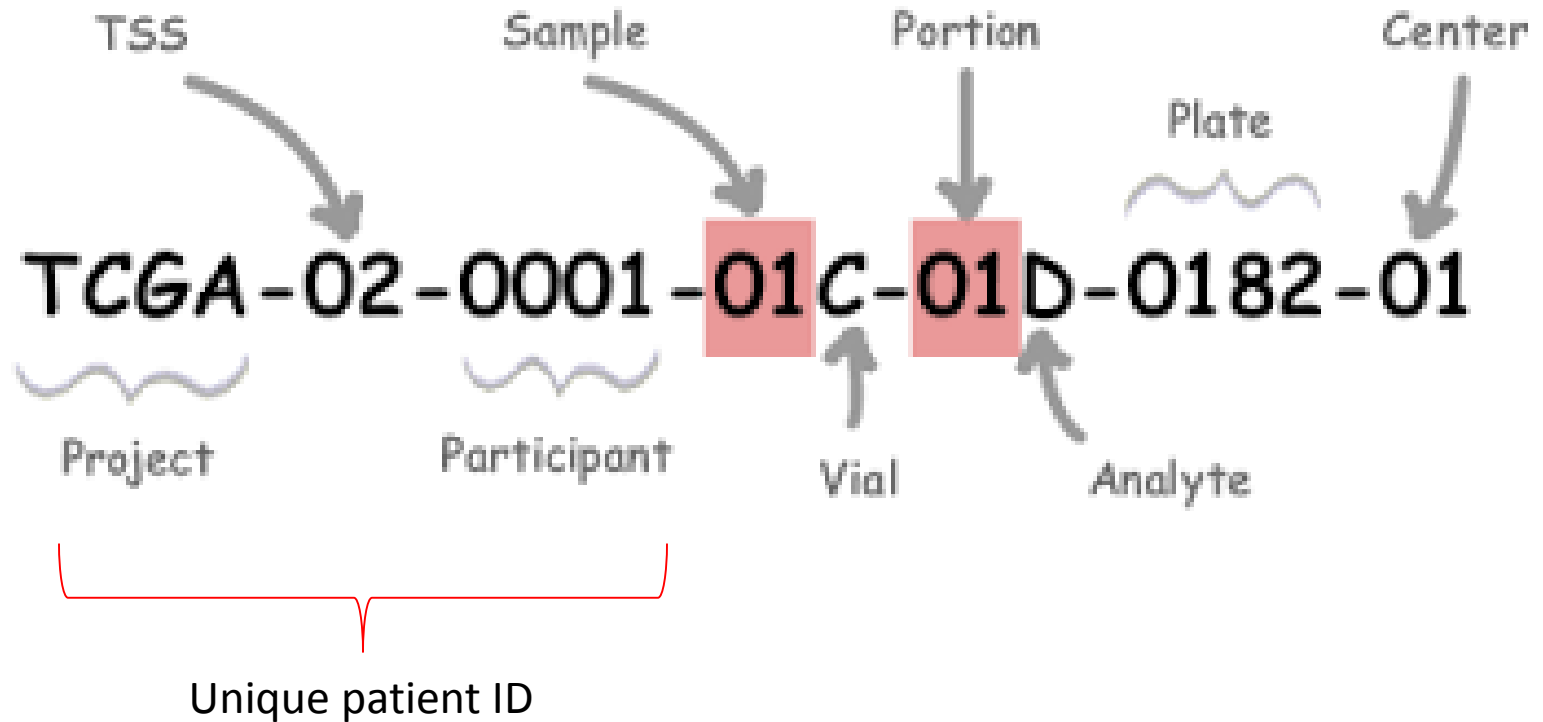


The Cancer Genome Atlas (TCGA)

- NCI / NHGRI project to molecularly characterize
 - Over 20k primary cancer samples
 - Matched normal samples
 - 33 cancer types
- Genomic, epigenomic, transcriptomic, and proteomic assays
- Data we'll use for exercises:
 - Small subset of BRCA RNA-seq data from TCGA
 - Unit: normalized gene expression levels



TCGA barcode IDs



https://docs.gdc.cancer.gov/Encyclopedia/pages/TCGA_Barcode/

TCGA sample codes

	Code	Definition	Short Letter Code
Tumor (01-09)	01	Primary Solid Tumor	TP
	02	Recurrent Solid Tumor	TR
	03	Primary Blood Derived Cancer - Peripheral Blood	TB
	04	Recurrent Blood Derived Cancer - Bone Marrow	TRBM
	05	Additional - New Primary	TAP
	06	Metastatic	TM
	07	Additional Metastatic	TAM
	08	Human Tumor Original Cells	THOC
	09	Primary Blood Derived Cancer - Bone Marrow	TBM
Normal	10	Blood Derived Normal	NB
	11	Solid Tissue Normal	NT
	12	Buccal Cell Normal	NBC
	13	EBV Immortalized Normal	NEBV
	14	Bone Marrow Normal	NBM
	15	sample type 15	15SH
	16	sample type 16	16SH
	20	Control Analyte	CELLC

Complete sample codes can be found at:

<https://qdc.cancer.gov/resources-tcga-users/tcga-code-tables/sample-type-codes>



Ex. 1

- Freeze the portion of the sheet BRCA_from_TCGA, so you can always see the column names (TCGA barcodes) and gene names no matter how you scroll.
 - hint: View -> Freeze panes
- Split the TCGA barcodes and check the sample codes. How many sample types?
 - hint: copy and paste barcodes to another sheet with Paste special -> Transpose, then split barcodes with Data -> Text to Columns -> View sample codes by turning on Sort & Filter button under Home panel)
- Replace sample barcodes with meaningful names:
 - 01 -> Primary
 - 06 -> Metastasis
 - 11 -> Normal
- Add sample codes as additional row to top of the BRCA_from_TCGA sheet.
 - hint: Copy -> Paste special -> Transpose
- Sort the samples by sample codes
 - hint: Data -> Sort -> Options -> Sort left to right



Math formulas

– Absolute value:

=ABS(B4)

– Average:

=AVERAGE(B4:D4)

– Log2 ratio:

=LOG((O4+1)/(Q4+1),2)


Add 1 to avoid denominator being 0



Statistics Formulas

- Example: T test:

=TTEST(B4:D4, I4:N4, 2, 3)

- **Tails** Required. Specifies the number of distribution tails. If tails = 1, T.TEST uses the one-tailed distribution. If tails = 2, T.TEST uses the two-tailed distribution.
- **Type** Required. The kind of t-Test to perform.

Parameters

If type equals	This test is performed
1	Paired
2	Two-sample equal variance (homoscedastic)
3	Two-sample unequal variance (heteroscedastic)



Logical formulas

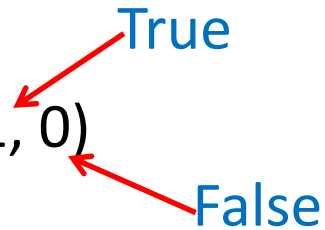
- Examples:

- If, else

=IF(U4<0.05, 1, 0)

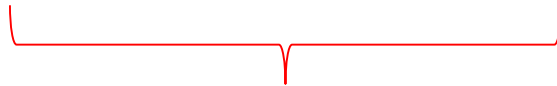
True

False



- and or

=IF (AND (T4<0.05, R4>=1), "yes", "no")



Formula details

- Begin with equal sign: =
- Dollar sign (\$): absolute cell references
- Copy → Paste Special → Values



Apply formula to multiple rows at once

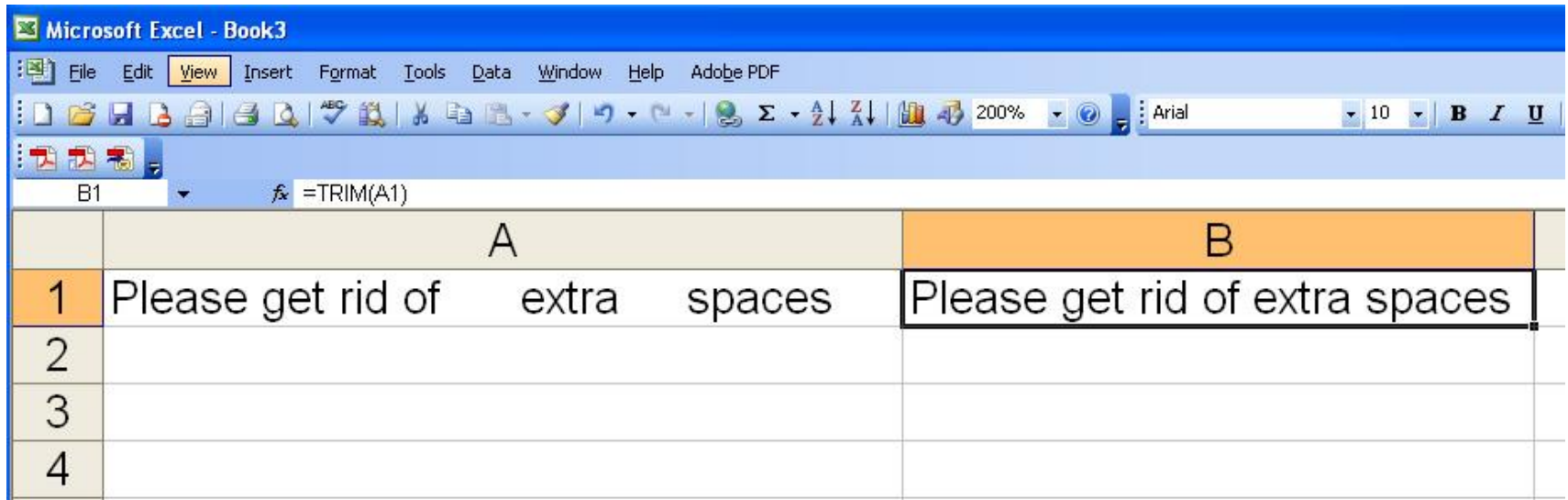
A	N	O	P	Q	R
	Primary		Metastasi:	Metastasi:	Metastasi: Nor
	TCGA-E2- A14V-01A- 11R-		TCGA-BH- A18V-06A- 11R-A213-11R-	TCGA-E2- A15E-06A- 11R-A12P-11A-	TCGA-E2- A15K-06A- A18V-06A- 11R-A12P-11A-
Hybridization REF	A12D-0	Max	07	A12D-0	07
PTPN20A 653129	0	0.0726	0.101112	0	0
CT45A2 728911	0	0.1657			
CT45A4 441520	0	0.1657			
SNORA4 619568	0	0.2651			
SNORA44 677825	0	0.284			
AQP12A 375318	0	0.2904			
C19orf75 284369	0	0.2904			
C3orf79 152118	0	0.2904			
C7orf33 202865	0	0.2904			
C8A 731	0	0.2904			
CCDC105 126402	0	0.2904			
CLEC1B 51266	0	0.2904			
DEFA4 1669	0	0.2904			
DMRTB1 63948	0	0.2904			
DPPA5 340168	0	0.2904			
DRD3 1814	0	0.2904			
GDF7 151449	0	0.2904			

A	N	O	P	Q	R	S
	Primary		Metastasi:	Metastasi:	Metastasi:	Normal
	TCGA-E2- A14V-01A- 11R-		TCGA-BH- A18V-06A- 11R-A213-11R-	TCGA-E2- A15E-06A- 11R-A12P-11A-23R-	TCGA-E2- A15K-06A- 11R-A12P-11A-23R-	TCGA-BH- A18V- 52R-
Hybridization REF	A12D-0	Max	07	A12D-0	07	A12D-0
PTPN20A 653129	0	0.0726	0.101112	0	0	
CT45A2 728911	0	0.1657	0	0	0	
CT45A4 441520	0	0.1657	0	0	0	
SNORA4 619568	0	0.2651	0	0	0	
SNORA44 677825	0	0.284	0	0	0	
AQP12A 375318	0	0.2904	0.367818	0	0	
C19orf75 284369	0	0.2904	0.367818	0	0	
C3orf79 152118	0	0.2904	0.367818	0	0	
C7orf33 202865	0	0.2904	0.367818	0	0	
C8A 731	0	0.2904	0.367818	0	0	
CCDC105 126402	0	0.2904	0.367818	0	0	
CLEC1B 51266	0	0.2904	0.367818	0	0	
DEFA4 1669	0	0.2904	0.367818	0	0	
DMRTB1 63948	0	0.2904	0.367818	0	0	
DPPA5 340168	0	0.2904	0.367818	0	0	
DRD3 1814	0	0.2904	0.367818	0	0	
GDF7 151449	0	0.2904	0.367818	0	0	
LCE3E 353145	0	0.2904	0.367818	0	0	
LOC283999 283999	0	0.2904	0.367818	0	0	

Note: double check several cells to make sure



Remove extra spaces within a cell



The screenshot shows the Microsoft Excel interface with the following details:

- Window title: Microsoft Excel - Book3
- Menu bar: File, Edit, View, Insert, Format, Tools, Data, Window, Help, Adobe PDF
- Toolbar: Standard toolbar with icons for file operations, editing, and formatting. The font is set to Arial, size 10. Bold, Italic, and Underline buttons are visible.
- Formula bar: B1, =TRIM(A1)
- Worksheet grid:
 - Column A: Please get rid of extra spaces
 - Column B: Please get rid of extra spaces
 - Row 1: Header row with columns A and B.
 - Row 2: Data row where the formula result is displayed.
 - Row 3: Empty row.
 - Row 4: Empty row.

Color cells with Conditional Formatting

Example 2

Highlight extreme values: Color Scales

The screenshot displays the Microsoft Excel interface. The 'Conditional Formatting' menu is open, and the 'Color Scales' option is selected. A dialog box for the 'Red - White - Blue Color Scale' is visible, providing instructions: 'Apply a color gradient to a range of cells. The color indicates where each cell value falls within that range.' The spreadsheet background shows columns P through U and rows 769 through 846. The 'log2(M/P)' column (T) is highlighted in green.

	P	Q	R	S	T	U		
769	3.3143	0	3.3143	0.2271	0.917217	0	0.643761	-0.93901
0	3.3143	0.6593	3.3143	0.358175	0.746617	0	0.362894	-0.80456
0	3.3143	0	3.3143	0.354275	0.721067	0	0.345782	-0.7833
0	3.3143	0.6593	3.3143	0	1.105383	0.1684	1.074083	-0.84955
0	3.3143	0	3.3143	0.117925	0.552383	0	0.473661	-0.63448
692	3.3143	0	3.3143	0.11355	1.088267	0.67365	0.90714	-0.31931
0	3.3143	0	3.3143	0.9615	1.058517	1.11	0.069648	0.035638
0	3.3143	0	3.3143	0	0.679967	1.01045	0.748433	0.259086
0	3.3143	0	3.3143	0	0.552383	0	0.634485	-0.63448
0	0	0.9889	3.3681	0.594075	0.64965	1.68405	0.04944	0.702252
923	0	0	3.3681	0.703175	0.192267	2.43735	-0.51452	1.52759
846	1.3257	1.6481	3.3681	1.44295	1.915183	2.43735	0.254962	0.23771
0	0	1.3185	3.3681	1.542875	0.21975	2.0607	-1.05988	1.327276



Color cells with Conditional Formatting

Example 3

Highlight extreme values: Color Scales -> More Rules

Microsoft Excel interface showing the 'New Formatting Rule' dialog box. The dialog is open over a spreadsheet with columns labeled R, S, and MvsP. The 'New Formatting Rule' dialog has 'Format only top or bottom ranked values' selected. The 'Format values that rank in the:' section shows 'Top' selected, '10' in the input field, and the checkbox for '% of the selected range' checked. A red box in the preview area shows 'AaBbCcYyZz'. Red arrows point from the text 'Top/bottom' to the 'Top' dropdown and the '% of the selected range' checkbox, and from 'Format...' to the 'Format...' button.

log2(M/P	log2(P/	MvsP.
0.384316	-0.34068	0.56
0.093457	0.077152	0.64
-0.30015	-1.06402	0.54
0.205713	0.109701	0.78
-0.97894	-1.33834	0.11
-0.53919	2.732735	0.78
1.403292	-2.04853	0.50
0.448461	-0.88095	0.4
0.008764	-0.00214	0.96
0.48596	-0.07929	0.31

Top/bottom



Ex. 2

1. Drop genes with no expression in all samples

- 1) Hint: Create another column and name it as “Max” -> use formula to calculate maximum value for each gene -> Sort data by “Max”
- 2) How do you easily find out where the non-zero value start?

Hint:

- a) Convert values in “Max” column from formula format to value with Copy -> Paste Special -> Values
- b) After sorting, with Sort & Filter button to find out the value right above 0 -> Search for this value under “Max” column

2. Compare primary tumor vs normal, and metastatic cancer vs primary tumor:

1. Convert values to log2 format:

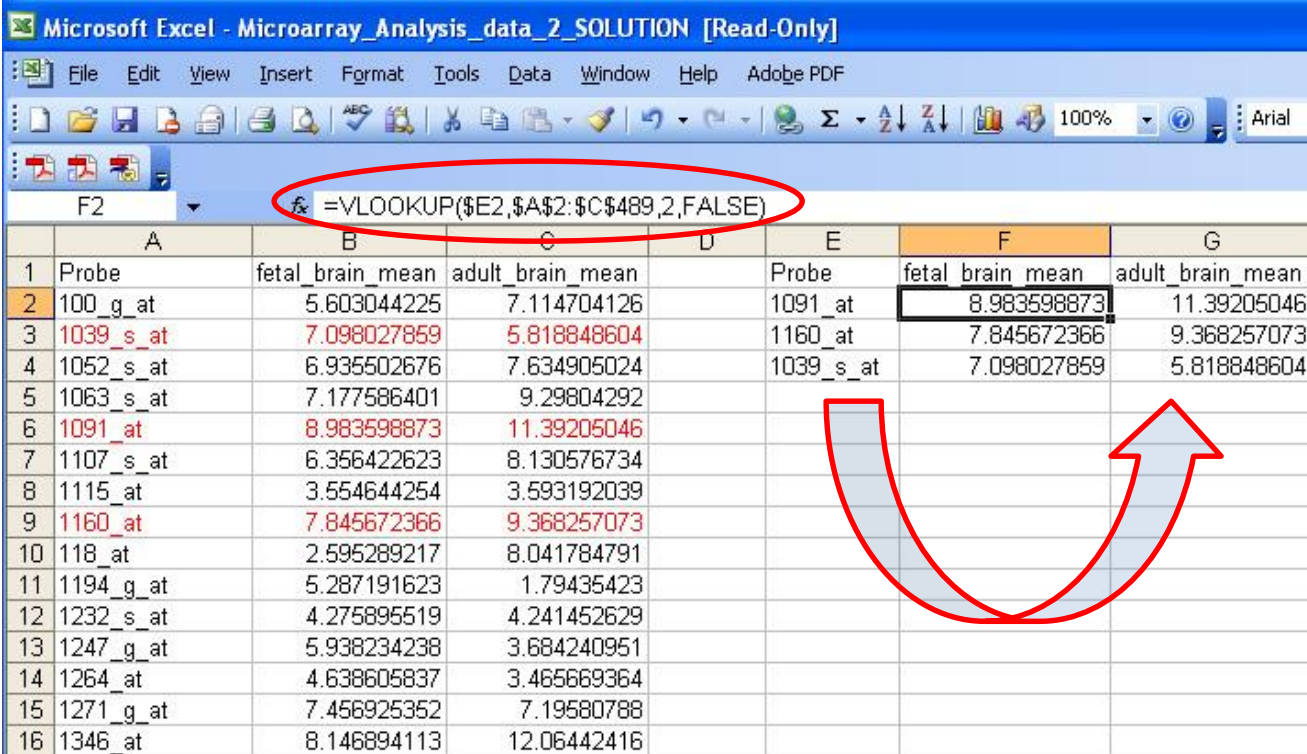
- 1) To avoid dividing by 0, you can add 1 to each value
- 2) Start to get the log2 expression for the top gene in the first patient sample, use our tricks (see slides) to fill out the rest of the cells

2. Calculate log2 average for each gene in normal samples, and do the same for primary tumor and metastasis samples.
3. Calculate log2 fold changes (log2FC) between primary tumor and normal, and between metastatic cancer and primary tumor, and highlight the log2FC with Conditional Formatting
4. Calculate p-values, find out FDR adjusted p-value with the BaRC web tool (<http://bioinfo.wi.mit.edu/bio/tools/fdr/>), and highlight FDR adjusted p-values < 0.05 with Conditional Formatting.

3. We consider a gene is “differentially expressed” if meeting the cutoffs (FDR adjusted p-value < 0.05 + at least 2 fold difference). Add two columns named as Sig(PvsN) and “Sig(MvsP)”, fill in with “y” and “n” to indicate if a gene is “differentially expressed” or not.

Hint: with “if” plus “and” formula

Lookup and retrieve data



Microsoft Excel - Microarray_Analysis_data_2_SOLUTION [Read-Only]

File Edit View Insert Format Tools Data Window Help Adobe PDF

100% Arial

F2 =VLOOKUP(\$E2,\$A\$2:\$C\$489,2,FALSE)

	A	B	C	D	E	F	G
1	Probe	fetal_brain_mean	adult_brain_mean		Probe	fetal_brain_mean	adult_brain_mean
2	100_g_at	5.603044225	7.114704126		1091_at	8.983598873	11.39205046
3	1039_s_at	7.098027859	5.818848604		1160_at	7.845672366	9.368257073
4	1052_s_at	6.935502676	7.634905024		1039_s_at	7.098027859	5.818848604
5	1063_s_at	7.177586401	9.29804292				
6	1091_at	8.983598873	11.39205046				
7	1107_s_at	6.356422623	8.130576734				
8	1115_at	3.554644254	3.593192039				
9	1160_at	7.845672366	9.368257073				
10	118_at	2.595289217	8.041784791				
11	1194_g_at	5.287191623	1.79435423				
12	1232_s_at	4.275895519	4.241452629				
13	1247_g_at	5.938234238	3.684240951				
14	1264_at	4.638605837	3.465669364				
15	1271_g_at	7.456925352	7.19580788				
16	1346_at	8.146894113	12.06442416				

Too complicated?

Try BaRC Submatrix Selector

http://bioinfo.wi.mit.edu/bio/tools/submatrix_selector/



Look up values in data list: vlookup

- **VLOOKUP(lookup_value,table_array,col_index_num,range_lookup)**
 - **lookup_value**: The value to search in the first column of the table array
 - **table_array**: The table to search (containing the value to search for in the first column)
 - Note: use “\$” to fix the table
 - **col_index_num**: the column number from which the matching value is returned
 - **range_lookup**:
 - True/omitted, an exact/approximate match
 - False: an exact match

Dollar sign (\$): absolute cell references



Ex.3

1. Extract gene symbols from the cells under column “Hybridization REF”. How do you check gene symbols to make sure none of them is converted to data format?

Hint:

Copy cells in this column and Paste them next to the rightmost column -> use “Text to Columns” to split cells

2. Use vlookup to get the expression levels of a list of genes, such as genes in the KEGG_ECM_RECEPTOR_INTERACTION gene set from [Broad's MSigDB](#).

3. Are any of these genes "differentially expressed" as defined by

- 1) adjusted p-value < 0.05
- 2) absolute log2 fold change >= 1

Hint: Use “Sort & Filter” button to quickly narrow down the genes



References

- Excel Annoyances - Curtis Frye
- Excel Hacks - O'Reilly
- <http://www.exceltip.com> (Joseph Rubin)
- <http://excel.tips.net> (Allen Wyatt)
- Some Excel Basics as well as formula basics
- http://barc.wi.mit.edu/education/bioinfo2005/arrays/Excel_help.html
- Zeeberg BR, Riss J, Kane DW, Bussey KJ, Uchio E, Linehan WM, Barrett JC, Weinstein JN. BMC Bioinformatics 5:80 2004

