


Excel is your Frenemy

Deanna M. Church
Senior Director of Genomics and Content
Personalis, Inc



 @deannachurch

Short Course in Medical Genetics 2014

Typical response from your local bioinformatician

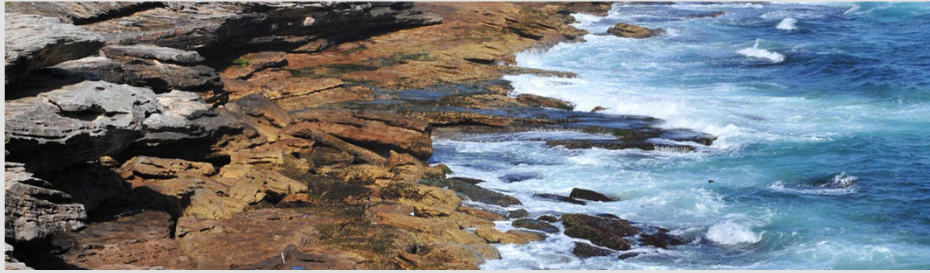


"I just emailed my data as an Excel workbook..."

Typical response from your local bioinformatician

dontuseExcel

Big data analysis, bioinformatics, open science



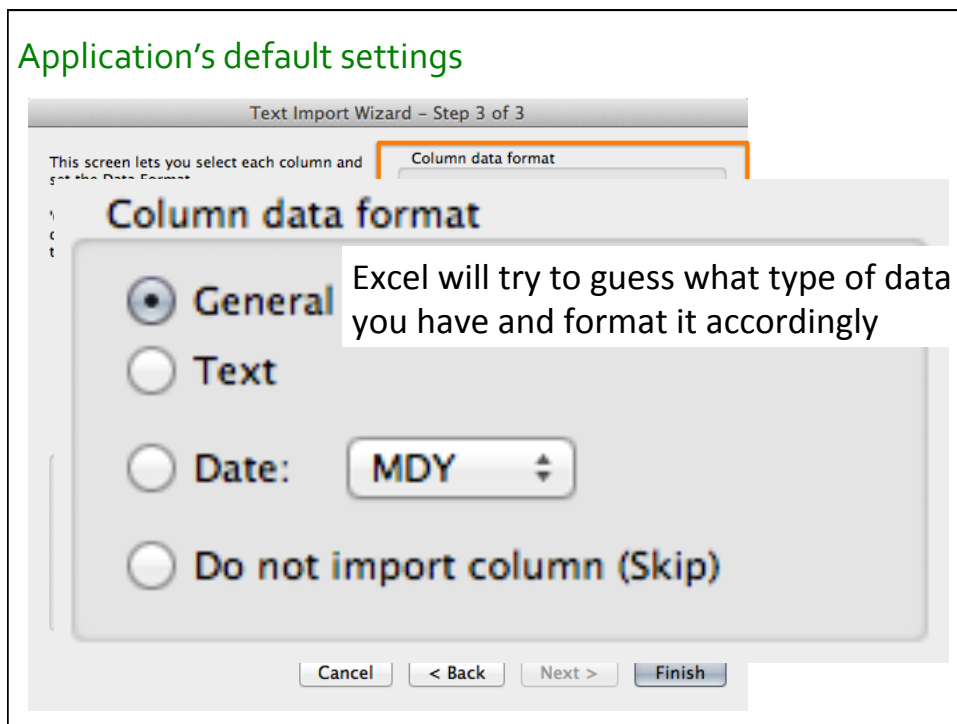
<http://dontuseexcel.wordpress.com/>



It is a poor workman, Malcolm, that blames his tools

New Yorker Cartoon (May 14, 1955), Robert J. Day

Application's default settings



Application's default settings

Oct4 → 4-Oct

Note: The official HGNC name for this gene is POU5F1, but Oct4 is a commonly used alias

Change default settings as needed

Column data format

☐ General
☒ Text
☐ Date: MDY
☐ Do not import column (Skip)

Change default settings as needed

BLAST® Basic Local Alignment Search Tool

Home Recent Results Saved Strategies Help

NCBI/ BLAST/ blastn suite **Standard Nucleotide BLAST**

blastn blastp blastx tblastn tblastx

Enter Query Sequence BLASTN programs search nucleotide databases using a nucleotide query. [more...](#)

Enter accession number(s), gi(s), or FASTA sequence(s) [Clear](#) [Query subrange](#)

From
To

Or, upload file No file chosen [?](#)

Job Title

Enter a descriptive title for your BLAST search [?](#)

☐ Align two or more sequences [?](#)

Choose Search Set

Database ☐ Human genomic + transcript ☐ Mouse genomic + transcript ☒ Others (nr etc.):
Nucleotide collection (nr/nt) [?](#)

Organism [Optional](#)
 Enter organism name or id—completions will be suggested ☐ Exclude [+](#)
Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown. [?](#)

Exclude [Optional](#)
☐ Models (XM/XP) ☐ Uncultured/environmental sample sequences

Entrez Query [Optional](#)
 [Limit search](#) [?](#)

<http://www.ncbi.nlm.nih.gov/blast>

Change default settings as needed

Only some of the available BLAST parameters

General Parameters

Max target sequences	<input type="text" value="100"/>	
Select the maximum number of aligned sequences to display		
Short queries	<input checked="" type="checkbox"/> Automatically adjust parameters for short input sequences	
Expect threshold	<input type="text" value="10"/>	
Word size	<input type="text" value="28"/>	
Max matches in a query range	<input type="text" value="0"/>	

Default BLAST page parameters tuned for highly related sequences

Take home messages

- ⌘ Understand how your tools work
 - ⌘ Understand default parameters
 - ⌘ Adjust your parameters to address your question
- ⌘ Do the exercises to see this in action!