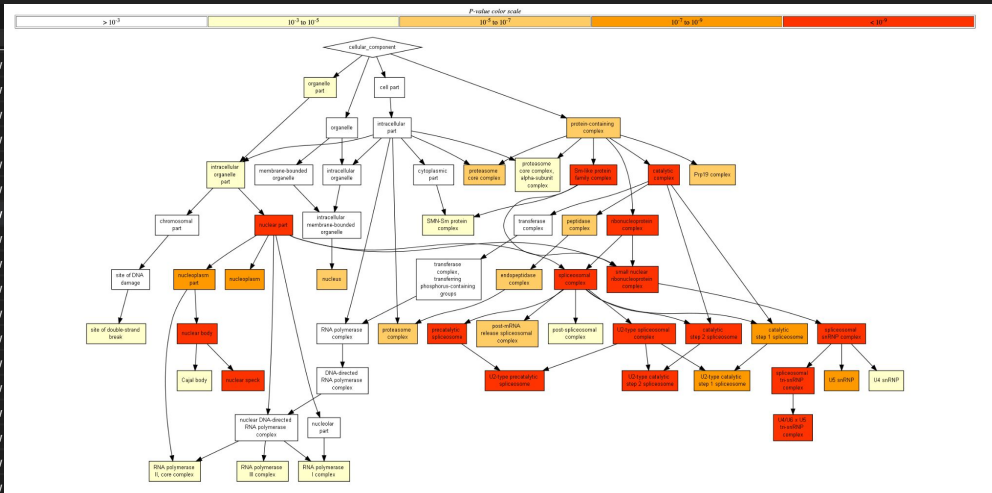


Week 4 - Visualizing GO (Gene Ontology) Analysis Output

I am looking at intron retention data and how it affects RNA metabolism. With the ReLiC screen data, I inputted the gene IDs based on its ranking relating to positive selection (increase of introns). I inserted the output into pandas and at the moment I am plotting a histogram to visually understand the data and how it answers the objective. Hopefully with Christine's and I's findings we can eventually create figures that can be beneficial to the project as a whole.

	index	GO term	Descrip...	P-value	FDR q-v...	Enrichment (N, B, n, b)	Genes
78	78	GO:00442...	cellular me...	0.000416	0.0387	1.29 (2149,1540,67,62)	[+] Show
59	59	GO:00081...	metabolic ...	0.000174	0.0213	1.29 (2149,1562,67,63)	[+] Show
37	37	GO:0071704	organic su...	0.0000162	0.00313	1.34 (2149,1504,67,63)	[+] Show
39	39	GO:00068...	nitrogen c...	0.0000304	0.0056	1.36 (2149,1424,72,65)	[+] Show
25	25	GO:00431...	macromol...	0.00000411	0.00116	1.38 (2149,1421,78,71)	[+] Show
33	33	GO:00442...	primary m...	0.0000155	0.00334	1.38 (2149,1440,67,62)	[+] Show
72	72	GO:0051179	localization	0.000296	0.0298	1.39 (2149,567,291,107)	[+] Show
86	86	GO:00512...	establishm...	0.000793	0.0671	1.40 (2149,513,291,97)	[+] Show
56	56	GO:1903311	regulation ...	0.000145	0.0187	1.53 (2149,250,439,78)	[+] Show
79	79	GO:1901360	organic cy...	0.000444	0.0408	1.55 (2149,1152,47,39)	[+] Show
70	70	GO:00464...	heterocycl...	0.000285	0.0295	1.57 (2149,1137,47,39)	[+] Show
71	71	GO:00067...	cellular ar...	0.000292	0.0299	1.57 (2149,1138,47,39)	[+] Show
67	67	GO:00061...	nucleobas...	0.000241	0.0261	1.58 (2149,1130,47,39)	[+] Show
51	51	GO:00903...	nucleic aci...	0.0000934	0.0132	1.63 (2149,1093,47,39)	[+] Show
77	77	GO:00301...	protein cat...	0.000404	0.0381	1.69 (2149,41,995,32)	[+] Show
87	87	GO:00485...	negative r...	0.000813	0.0679	1.82 (2149,155,313,41)	[+] Show
38	38	GO:0016070	RNA meta...	0.000028	0.00528	1.86 (2149,1016,33,29)	[+] Show
88	88	GO:00062...	nucleotide...	0.000842	0.0696	10.41 (2149,24,43,5)	[+] Show



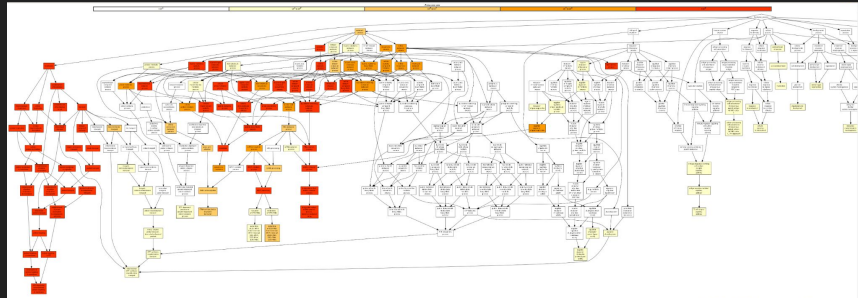
Week 4

Also starting to developing a rough draft of a presentation into markdown by noting all the significant finding along my studies.

```
notes > ● gg_methodology.md > 📄 # Objective > 📄 ## Output
1  ● Objective
2  - RNA metabolism is influenced by various factors, including processes like splicing and nonsense-mediated decay (NMD). Assessing the impact
  of these factors is crucial. In the context of CRISPR screens, direct measurement of RNA levels becomes important. RNA sequencing provides a
  quantitative approach to evaluate RNA molecules, offering insights into the effects of gene perturbations on transcriptional regulation, key
  genes, and pathways. The Relic screening strategy, focused on splicing and NMD, holds promise in identifying functional consequences
  resulting from CRISPR-mediated gene modifications. Such investigations contribute to a deeper understanding of the impact of RNA-related
  processes on gene function.
3  - **GO** - What are the factors that affects RNA metabolism specifically splicing?
4  ## Research
5  - **What is CRISPR Screening?**
6    - CRISPR screening is a technique used to identify and understand the function of specific genes in an organism. It involves using
      CRISPR-Cas9, a powerful gene editing tool, to systematically target and modify different genes within a cell or organism. By observing
      the resulting changes in cellular behavior or characteristics, researchers can gain insights into the roles and importance of different
      genes in various biological processes.
7  - **Relic Screenings** (Notes from Presentation)
8    - CIBER-seq is a powerful tool for studying gene expression and identifying gene pathways related to protein synthesis. It utilizes a
      barcode system to measure the effects of gene changes, revealing varying impacts on cells. By integrating CRISPR interference and
      barcoded expression reporters, it allows measurement of gene expression and understand how different genetic changes influence gene
      interactions. CIBER-seq uncovers the connections between genes and their roles in biological processes.
9    - Splicing Relic: As learned in started Biology classes, the relation between DNA and RNA is known as the central dogma of molecular
      biology where DNA makes mRNA makes protein through a strategic process. The focus of this process includes a complex and regulated step
      that includes transcription, pre mRNA splicing, editing, degradation, polyadenylation, etc. known as RNA metabolism. Splicing plays a
      vital role in which the cutting of introns, noncoding, for exons, coding, to covalently connect for the exons to exit the nucleus and be
      expressed. The exons acts as a random barcode that is counted by a deep screening to find the effect of sgRNA on reporter mRNA levels.
      Pooled CRISPR screens are widely used for forward genetic screening in human cells; however, the disrupted RNA metabolism does not
      manifest with growth and viability defects. For instance, the presence of a mutation can change a phenotype completely or not change it
      at all. FACS-based CRISPR screening uses protein level readouts for mRNA phenotypes; however, RNA molecules may not alter protein
      levels. RNA molecules can undergo modifications with specific proteins and the production can be independently performed.
10   - **Insert Figures**
11  - **MaGeCK**
12    - Model-based Analysis of Genome-wide CRISPR/Cas9 Knockout: a method used to prioritize single guide RNA, genes, and specific pathways
      in CRISPR screenings. This method is commonly used in comparison to other methods, for it identifies positive and negatively selected
      genes under robust and extreme conditions, noting which genes survived and which genes didn't. Though the use and integration of public
      data sets knocked out data sets are easily identifiable.
13    - Pandas is a powerful Python library that specializes in data manipulation and analysis was used to focus on the genetic screening data
      analysis. Pandas is also used to manage tabular data, which helped with analyzing which column to use for the input on GOrilla.
14    - Utilizing pandas, I compared the variables from MaGeCK and visualized the data as a graph. Through this analysis, I identified the
      most suitable column for GOrilla.
15  - **GO(Gene Ontology)rilla**
16    - GOrilla is a web-based application that performs gene ontology (GO) enrichment analysis on ranked gene lists, without the need for
      explicit target and background sets. GOrilla provides efficient and rigorous statistical analysis, taking threshold multiple testing
      into account without the need for simulations. The results are presented in a hierarchical structure, offering a clear visualization of
      the relationships between enriched GO terms. GOrilla is a valuable addition to the existing repertoire of GO enrichment tools, offering
      unique features, fast running time, and effective graphical representation.
```

Week 1-4 Update

- Previous weeks focus: python, data interpretation/organization, and gene ontology
- NMD ReLiC: The analysis of NMD (nonsense-mediated mRNA decay) insertion and deletion events in CRISPR data screening yields valuable insights into the functional consequences and potential regulatory effects of gene modifications within the nonsense-mediated decay pathway.
- This week we are analyzing, interpreting, and putting the gene ontology data into a figure.



GO term	Description	Rank	FDR value	Enrichment (N, R, n, k)	Genes
GO:0001184	nucleus-transcribed mRNA catabolic process, nonsense-mediated decay	2.85E-28	2.09E-24	4.48 (2150,115,203,60)	[1] Show genes
GO:0019083	viral transcription	2.1E-27	7.72E-24	5.02 (2150,88,203,54)	[1] Show genes
GO:0060614	retromer-mediated protein targeting to membrane	2.65E-27	6.3E-24	4.63 (2150,90,294,57)	[1] Show genes
GO:0060444	translational initiation	2.99E-27	5.3E-24	4.12 (2150,133,203,07)	[1] Show genes
GO:0060614	GTP-dependent retromer-mediated protein targeting to membrane	5.96E-27	8.7E-24	4.65 (2150,88,294,50)	[1] Show genes
GO:0070722	protein localization to endoplasmic reticulum	6.37E-27	7.8E-24	4.58 (2150,91,294,57)	[1] Show genes
GO:0072359	establishment of protein localization to endoplasmic reticulum	6.37E-27	6.69E-24	4.58 (2150,91,294,57)	[1] Show genes
GO:0045042	protein targeting to ER	6.37E-27	5.85E-24	4.58 (2150,91,294,57)	[1] Show genes
GO:0060614	protein targeting to membrane	1.5E-26	1.2E-23	4.55 (2150,92,294,57)	[1] Show genes
GO:0060603	protein targeting	3.30E-25	2.4E-22	4.23 (2150,102,294,50)	[1] Show genes
GO:0001150	establishment of protein localization to membrane	7.72E-25	5.10E-22	4.24 (2150,100,294,58)	[1] Show genes
GO:0072632	protein localization to membrane	7.58E-22	4.6E-19	3.86 (2150,110,294,58)	[1] Show genes
GO:0074394	establishment of protein localization to organelle	2.34E-21	1.27E-18	3.52 (2150,133,294,54)	[1] Show genes
GO:0004956	nucleus-transcribed mRNA catabolic process	5.08E-20	2.67E-17	3.22 (2150,175,203,60)	[1] Show genes
GO:0004112	translation	6.15E-19	3.02E-16	3.02 (2150,172,294,71)	[1] Show genes
GO:0043604	amide biosynthetic process	8.29E-19	3.81E-16	2.93 (2150,181,294,78)	[1] Show genes
GO:0060318	peptide metabolic process	1.24E-18	5.50E-16	2.93 (2150,182,294,78)	[1] Show genes
GO:0043618	peptide biosynthetic process	1.41E-18	5.70E-16	2.98 (2150,171,294,71)	[1] Show genes
GO:0043603	cellular amide metabolic process	1.54E-18	5.53E-16	2.87 (2150,191,294,75)	[1] Show genes
GO:0008408	mRNA catabolic process	3.93E-18	1.43E-15	3.00 (2150,191,203,70)	[1] Show genes
GO:0033363	protein localization to organelle	5.53E-18	1.87E-15	2.99 (2150,169,294,69)	[1] Show genes
GO:0014566	organonitrogen compound biosynthetic process	1.17E-16	3.9E-14	2.69 (2150,217,298,70)	[1] Show genes
GO:0004957	macromolecular catabolic process	1.87E-16	5.99E-14	2.32 (2150,279,315,95)	[1] Show genes
GO:0044303	cellular macromolecular catabolic process	1.97E-16	6.03E-14	2.35 (2150,275,315,90)	[1] Show genes
GO:0008866	intracellular protein transport	6.21E-16	1.83E-13	2.92 (2150,182,203,65)	[1] Show genes
GO:0004401	RNA catabolic process	7.78E-16	2.2E-13	3.04 (2150,215,204,62)	[1] Show genes
GO:0034414	cellular protein localization	2.5E-15	6.81E-13	2.68 (2150,194,294,71)	[1] Show genes
GO:0070722	cellular macromolecular localization	5.85E-15	1.54E-12	2.92 (2150,201,294,72)	[1] Show genes
GO:0043601	organic cyclic compound catabolic process	1.74E-14	4.43E-12	2.51 (2150,248,203,70)	[1] Show genes
GO:0046700	heterocycle catabolic process	1.74E-14	4.28E-12	2.51 (2150,248,203,70)	[1] Show genes
GO:0014572	organic substance catabolic process	3.35E-14	5.39E-12	2.11 (2150,335,315,102)	[1] Show genes
GO:0016032	viral process	2.76E-14	6.33E-12	2.47 (2150,238,203,70)	[1] Show genes
GO:0044103	resolvent process	2.76E-14	6.16E-12	2.47 (2150,238,203,70)	[1] Show genes
GO:0040902	intracellular transport	3.65E-14	7.9E-12	3.08 (2150,317,310,50)	[1] Show genes
GO:0044291	cellular nitrogen compound catabolic process	3.86E-14	8.11E-12	2.48 (2150,251,203,70)	[1] Show genes
GO:0044653	indole-3-ol-containing compound catabolic process	5.42E-14	1.11E-11	2.48 (2150,247,203,75)	[1] Show genes
GO:0014575	aromatic compound catabolic process	5.42E-14	1.09E-11	2.48 (2150,247,203,75)	[1] Show genes

What have I learned during my first week at Fred Hutch?

I learned ...

- About the application **terminal** on my mac computer and the similarities and difference between using the app on a mac computer and windows computer.
- **Markdown** and its markup language to transfer to a text document, website, etc.
- **GitHub** and **VSCode** and the connection to transfer text using the terminal on my mac
- About **Slack** and its use for messaging
- And refreshed my memory on **CRISPR** and its strategy to edit genomes
- The routes of the campus

```
racequarterman@graces-mbp notes % git commit
error: pathspec 'screen' did not match any file
error: pathspec 'notes' did not match any file
racequarterman@graces-mbp notes % git add gq
racequarterman@graces-mbp notes % git commit
error: pathspec 'notes' did not match any file
racequarterman@graces-mbp notes % git push o
git: 'credential-manager-core' is not a git c
o https://github.com/kychen37/rasilab_spelma
! [rejected]        main -> main (fetch firs
error: failed to push some refs to 'https://g
int: Updates were rejected because the remot
int: not have locally. This is usually cause
int: to the same ref. You may want to first
int: (e.g., 'git pull ...') before pushing a
int: See the 'Note about fast-forwards' in '
racequarterman@graces-mbp notes % git pull o
error: cannot pull with rebase: Your index co
error: please commit or stash them.
racequarterman@graces-mbp notes % git add gq
```

rasilab_spelman_2023 / gq_crispr_notes.md

Preview Code Blame 98 Lines (87 loc) · 4.94 KB

What is CRISPR-Cas9?

Jennifer Doudna's Ted Talk

- Doudna purpose of creating CRISPR was to invent technology for **editing genomes**
- Changes to genetic cells to cure genetic diseases
- Cas9 is a specific protein that cuts the DNA to change specific infected part on DNA
- Stands for Clustered Regularly Interspaced Short Palindromic Repeats
- Allows the cell to keep track of infection..."genetic vaccination card"
- Trigger cells to repair breaks in DNA at the points near or at the mutation
- Still some complications due to unintended effects but when it happend Doudna believes th
- Enhancement for genetic change..."designer humans"

YourGenome

- Edits parts of the genome by removing, adding, or alternating section of DNA
- CRISPR is beneficial because it is a simple and precise way of genetic manipulation
- Cas9 (enzyme) acts as a "molecular scissor"
- gRNA (guide RNA) guides the enzyme to the right part of the mutated genome
- Makes a cut across both strands

The terminal (mac/linux) or command prompt (windows) is a way to execute sim

GitHub

- Make an account on github and verify your email address
- Install git to your laptops by following [this](https://www.atlassian.com/gi
- Clone our group's remote repository to your local computer by opening the te
- github.com/kychen37/rasilab_spelman_2023.git``
- Since this repo was made in my account (kychen37), I needed to added gquar
- [Settings](https://github.com/kychen37/rasilab_spelman_2023/settings) before
- Each user needs to then generate a personal access token:
- Go to your user settings -> Developer settings -> Personal access tokens -
- Under Note, name the token something descriptive and check 'repo'
- Press Generate token, copy the entire token to a different location like a
- Follow the top comment on https://stackoverflow.com/questions/46645843/where
- access token to the git credential helper so you don't have to keep copy/pasti
- ``git config --global user.name "Katharine Chen"``
- ``git config --global user.email kychen37@uw.edu``
- ``git config --global credential.helper manager-core``



Grace Quarterman

WEEK 1 AT FRED HUTCH CANCER CENTER IN THE SUBRAMANIAM LAB

GITHUB, COMMAND PROMPTS, VSCODE AND MARKDOWN

- LANGUAGE BETWEEN WINDOWS AND MAC COMPUTERS
- COMPUTER INTERACTIONS AND NAVIGATIONS
- TROUBLESHOOTING ERROR MESSAGES WITHIN OUR CODE
 - THERE WERE A LOT OF THEM

CRISPR

- IT USES A SPECIALIZED PROTEIN CALLED CAS9, GUIDED BY A SMALL RNA MOLECULE, TO TARGET SPECIFIC DNA SEQUENCES AND INTRODUCE MODIFICATIONS, WITHIN LIVING ORGANISMS

```
C:\Users\chris\rasilab_spelman_2023\notes>git push origin main
Everything up-to-date

C:\Users\chris\rasilab_spelman_2023\notes>git add cb_crispr_notes.md
fatal: pathspec 'cb_crispr_notes.md' did not match any files

C:\Users\chris\rasilab_spelman_2023\notes>

C:\Users\chris\rasilab_spelman_2023\notes>cb_crispr_screening_notes.md

C:\Users\chris\rasilab_spelman_2023\notes>
cb_crispr_screening_notes.md

C:\Users\chris\rasilab_spelman_2023\notes>
git commit
On branch main
Your branch is up to date with 'origin/main'.

Changes not staged for commit:
  (use "git add <file>..." to update what will be committed)
  (use "git restore <file>..." to discard changes in working directory)
        modified:   cb_crispr_screening_notes.md

no changes added to commit (use "git add" and/or "git commit -a")

C:\Users\chris\rasilab_spelman_2023\notes>git commit -m "edit #4"
On branch main
Your branch is up to date with 'origin/main'.

Changes not staged for commit:
  (use "git add <file>..." to update what will be committed)
  (use "git restore <file>..." to discard changes in working directory)
        modified:   cb_crispr_screening_notes.md

no changes added to commit (use "git add" and/or "git commit -a")

C:\Users\chris\rasilab_spelman_2023\notes>git add cb_crispr_screening_notes.md
C:\Users\chris\rasilab_spelman_2023\notes>git commit -m "edit #4"
[main 5fc9048] edit #4
```