NMD and Splicing ReLiC (RNA-linked CRISPR Screening) Gene Ontology Analysis

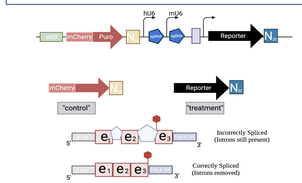
Christine Bynum and Grace Quarterman
Subramaniam Lab Mentor: Katharine Chen

Fred Hutch Cancer Center
Formerly FHCRC

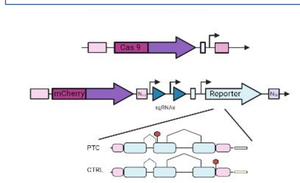
Background

CRISPR screening is an efficient technique utilized to simultaneously disrupt numerous genes by employing expansive libraries of sgRNAs in collaboration with Cas9. This targeted approach enables the comprehensive assessment of a specific gene set, allowing for systematic targeting and modification of various genes within cells or organisms. Through careful observation of these modifications, a deeper understanding of gene roles in different processes is enhanced. This project investigates RNA-linked CRISPR screening (**ReLiC**), a custom screening approach developed in the Subramaniam lab, to concentrate on factors affecting RNA metabolism. Typical CRISPR screens rely on either growth and viability differences or protein-level output, while ReLiC directly measures mRNA levels. Identical ReLiC screens were studied with different "Reporter" locations, either having a splicing or NMD reporter. The **Splicing ReLiC** screening utilized a reporter to assess intron retention, aiding in the identification of factors influencing splicing. Meanwhile, **Nonsense Mediated Decay (NMD) ReLiC** screening employed a reporter to measure PTC-containing mRNA counts, facilitating the identification of factors associated with NMD.

Splicing ReLiC



NMD ReLiC



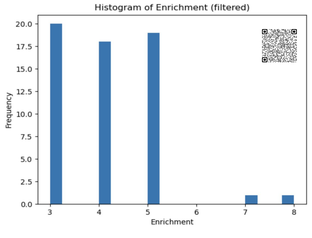
Objective

The project's main goal is to uncover novel factors that influence RNA metabolism, particularly splicing and nonsense-mediated decay (NMD). Using the ReLic screening strategy, which focuses on these processes, the study aims to identify functional consequences resulting from CRISPR-mediated gene modifications. This investigation will contribute to a deeper understanding of how RNA-related processes impact gene function. By utilizing Gene Ontology, the project hopes to pinpoint the specific processes affecting splicing and NMD within RNA metabolism.

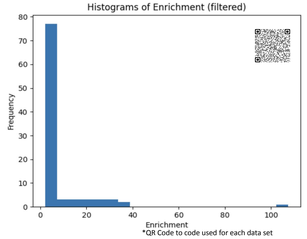
Methodology

Focused analysis on CRISPR data obtained from a custom screening conducted by the Subramaniam Lab, targeting gene targets related to intron retention and nonsense-mediated decay (NMD). Python (pandas) was used for data processing, and GOrilla was utilized for Gene Ontology (GO) analysis. The analysis identified significant results for various biological processes associated with splicing, RNA processing, and NMD, based on color-coded p-values. The False Discovery Rate (FDR) was employed to control for false associations, ensuring q-values below 0.05. The enrichment filtered to numbers that are greater than 2. Histograms were used to visualize data, and the enrichment values were rounded to better understand their significance. The lowest q-values in the table were linked to crucial screening areas, supporting the research objectives.

Splicing ReLiC



NMD ReLiC



Results

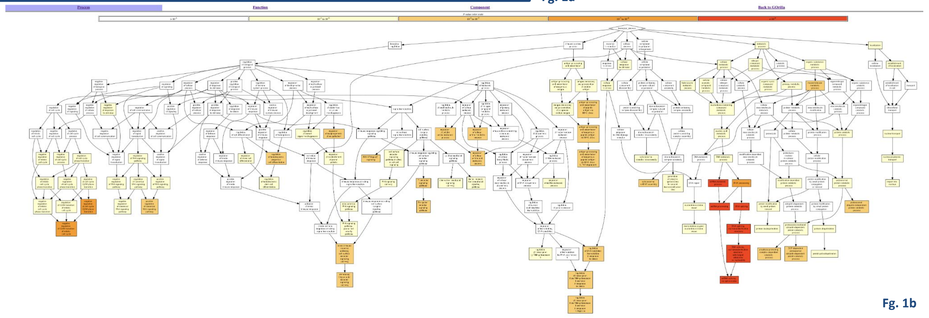
Splicing

	GO term	Description	FDR q-value	enrichment	AssociatedGenes
2	GO:0000377	RNA splicing, via transesterification reactions with bulged adenosine as nucleophile	1.04e-20	5.37	235
3	GO:0000375	RNA splicing, via transesterification reactions	1.59e-20	5.28	239
4	GO:0000398	mRNA splicing, via spliceosome	2.08e-20	5.37	235
5	GO:0008380	RNA splicing	1.5e-16	3.86	321
6	GO:0006397	mRNA processing	9.86e-14	3.57	369
7	GO:0016071	mRNA metabolic process	2.19e-09	3.42	552
8	GO:0006396	RNA processing	2.39e-06	2.66	711
9	GO:0010972	negative regulation of G2/M transition of mitotic cell cycle	1.87e-05	5.31	21
10	GO:1902750	negative regulation of cell cycle G2/M phase transition	4.88e-05	5.07	22
11	GO:0042590	antigen processing and presentation of exogenous peptide antigen via MHC class I	0.000341	5.09	19
12	GO:0002479	antigen processing and presentation of exogenous peptide antigen via MHC class I, TAP-dependent	0.000372	5.09	19
13	GO:0038095	Fc-epsilon receptor signaling pathway	0.000409	5.09	19
14	GO:0038093	Fc receptor signaling pathway	0.000431	4.73	22
15	GO:0031146	SCF-dependent proteasomal ubiquitin-dependent protein catabolic process	0.000505	5.25	17
16	GO:0038061	NIK/NF-kappaB signaling	0.000541	5.25	17
17	GO:0043618	regulation of transcription from RNA polymerase II promoter in response to stress	0.000738	5.08	23
18	GO:1901532	regulation of hematopoietic progenitor cell differentiation	0.000874	3.32	27
19	GO:0002223	stimulatory C-type lectin receptor signaling pathway	0.000879	3.87	18
20	GO:0000387	spliceosomal snRNP assembly	0.000882	3.84	32

Figure 1a presents a table organized by q-value, revealing significant findings related to splicing and RNA processing. Each entry includes a human-readable name and a unique identifier for the Gene Ontology (GO) term. Lower q-values indicate more relevant genes associated with splicing. The "enrichment" column summarizes the functions of differentially expressed genes, and the "associated genes" column shows the total number of genes linked to each specific GO term.

In **Figure 1b**, the directed acyclic graph (DAG) represents broader biological processes derived from Gene Ontology analysis, serving as a roadmap of gene pathways. The low false discovery rate in the table and its prevalence on the DAG further reinforce splicing as a primary factor throughout the ReLiC screening data.

Fg. 1a



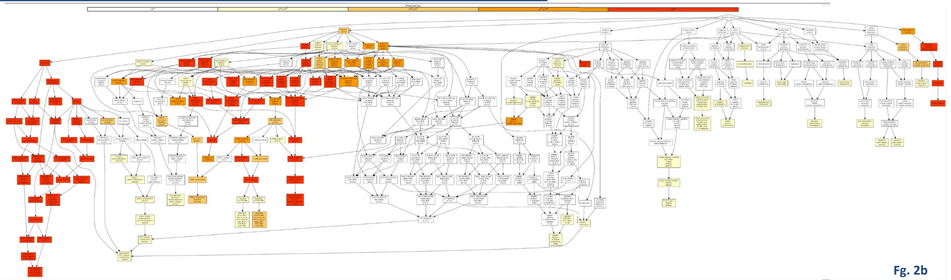
Fg. 1b

Nonsense Mediated Decay

	GO term	Description	FDR q-value	AssociatedGenes	enrichment
2	GO:0000184	nuclear-transcribed mRNA catabolic process, nonsense-mediated decay	2.09e-24	115	4.48
3	GO:0006413	translational initiation	5.5e-24	133	4.12
4	GO:0045047	protein targeting to ER	5.85e-24	91	4.58
5	GO:0006613	cotranslational protein targeting to membrane	6.5e-24	90	4.63
6	GO:0072599	establishment of protein localization to endoplasmic reticulum	6.6899999999999e-24	91	4.58
7	GO:0019083	viral transcription	7.72e-24	88	5.02
8	GO:0070972	protein localization to endoplasmic reticulum	7.8e-24	91	4.58
9	GO:0006614	SRP-dependent cotranslational protein targeting to membrane	8.76e-24	88	4.65
10	GO:0006612	protein targeting to membrane	1.2299999999999998e-23	92	4.53
11	GO:0006605	protein targeting	2.4e-22	102	4.23
12	GO:0090150	establishment of protein localization to membrane	5.16e-22	100	4.24
13	GO:0072657	protein localization to membrane	4.64e-19	110	3.86
14	GO:0072594	establishment of protein localization to organelle	1.27e-18	133	3.52
15	GO:0000956	nuclear-transcribed mRNA catabolic process	2.67e-17	175	3.22
16	GO:0006412	translation	3.02e-16	172	3.02
17	GO:0043604	amide biosynthetic process	3.81e-16	181	2.95
18	GO:0006518	peptide metabolic process	5.36e-16	182	2.93
19	GO:0043043	peptide biosynthetic process	5.76e-16	174	2.98
20	GO:0043603	cellular amide metabolic process	5.95e-16	191	2.87

Figure 2a and **Figure 2b** share similar variables with **Figures 1a** and **Figure 1b,** but they focus on nonsense-mediated decay (NMD). As in **Figure 1a**, lower q-values in **Figure 2a** indicate desired output, specifically genes associated with NMD. "enrichment" in **Figure 2a** summarizes the functions of differentially expressed genes related to NMD, and the "associated genes" column highlights significant numbers. The low false discovery rate in the table and its prevalence on the DAG , **Figure 2b,** further validate NMD as a primary factor throughout the ReLiC screening data, similar to splicing in **Figure 1b**.

Fg. 2a



Fg. 2b

Conclusions

We conducted gene ontology analyses on both datasets and ranked outputs by descending q-value so most significantly affected biological processes are at the top. This confirmed that splicing and NMD are the biological processes most affected by the ranking of the FDR (false discovery rate) q-value. In essence, the low rates of false positives indicate that the ReLiCs effectively target the specific factors of splicing and NMD. This reduces the risk of mistakenly attributing the effects to other factors, confirming the relevance of NMD and splicing in the ReLiC screening process.

Future work

Explore potential factors beyond what is currently known about splicing and NMD, specifically focusing on the statistically significant processes revealed by the NMD diagram, which have not been examined in this context before. Findings prompt further examination to understand their implications and applicability to other relevant studies, particularly within the context of CRISPR technology. Analysis of the most significant factors and q factors, supported by visualizations, to understand their impact on splicing and NMD. Additionally, conduct further investigations from Gene Ontology (GOrilla) output to gain deeper insights into the contributing factors in these biological processes.

Clinical Impact on Oral Health

1. CRISPR allows researchers to make precise changes to genes in laboratory models, like animals or cell cultures. By doing so, they can study the genetic factors involved in dental diseases such as tooth decay, gum disease, and oral cancers. This research can lead to a better understanding of these conditions and the development of new treatments.

2. CRISPR-based techniques can be used to create advanced diagnostic tools for oral diseases. By leveraging CRISPR's gene-editing abilities, scientists can design tests that detect specific genetic markers in saliva. This enables early detection of oral conditions and diseases, improving the chances of successful treatment.

Acknowledgments & Sources

Special thanks are extended to the Fred Hutchinson DEI core team, Subramaniam Lab, Morehouse School of Medicine Undergraduate Health Sciences Academy, Tavarius Horne, Katharine Chen, Dr. Christopher Garrett, and Dr. Rahmelle Thompson for providing us with the invaluable opportunity to intern with the Fred Hutchinson computational biology department during the Summer of 2023.

1. Innovative Genomics Institute. (n.d.). What is CRISPR? Retrieved from https://innovativegenomics.org/education/digital-resources/what-is-crispr/

2. Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., & Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science, 337(6096), 816-821. doi:10.1126/science.1225829

3. Khan Academy. (2015, June 1). CRISPR-Cas9 [Video]. YouTube. Retrieved from https://www.khanacademy.org/college-careers-more/bjc/2015-challenge/2015-life-science/v/bjc-crispr-cas-9

4. Shalem, O., Sanjana, N. E., & Zhang, F. (2015). High-throughput functional genomics using CRISPR-Cas9. Nature Reviews Genetics, 16(5), 299-311. doi:10.1038/nrg3899

5. SourceForge. (n.d.). MAGeCK - Analysis of CRISPR/Cas9 knockout and knockdown screens. Retrieved from https://sourceforge.net/p/mageck/wiki/Home/#gene_summary_txt

6. Terns, M. P., & Terns, R. M. (2011). CRISPR-based adaptive immune systems. Current Opinion in Microbiology, 14(3), 321-327. doi:10.1016/j.mib.2011.03.005

7. Wei, W., Gilbert, H. N., & Virgil, W. G. (2009). Gibbs motif sampler: de novo motif discovery. BMC Bioinformatics, 10(1), 1-17. doi:10.1186/1471-2105-10-48