أكاديمية كاوست
KAUST ACADEMY

جامعة الملك عبدالله للعلوم والتقنية
King Abdullah University of Science and Technology

The tRNA Repair Machinery Co-enriches with Rlig1: Extending a Two-Enzyme Observation to the Broader Pathway

The tRNA-Repair Machinery Co-Enriches with Rlig1

Hanyn A. Alturki
haneenalturkii@gmail.com

Mentors: **Victor Lopez**, **Thomas Krueger**

Date: 23 / 8 / 2026
Code: https://github.com/haneenalturkii/Rlig1/blob/main/tRNA.R
Project page: https://github.com/haneenalturkii/Rlig1

Abstract

Human RNA ligase 1 (Rlig1, encoded by the *C12orf29* gene) is a cytoplasmic, T4-like RNA-repair ligase that rejoins broken tRNA fragments to preserve translation during oxidative stress, among other suspected functions. While biochemical studies established that Rlig1 operates independently of canonical tRNA-splicing machinery, its broader functional context remained largely unmapped beyond reported associations with CLP1 and ANGEL2.

Reanalysis of a published Rlig1 affinity-enrichment mass spectrometry dataset (Stumpf *et al*., 2026) using an R/Bioconductor pipeline (`QFeatures` and `limma`) extends these initial observations to the broader tRNA pathway. Targeted testing demonstrates that 19 of 22 detected human tRNA end-processing and repair enzymes co-enrich with active Rlig1 (Fisher's exact $p = 3.9 \times 10^{-11}$).

STRING network analysis was applied to the co-enriched tRNA enzymes, confirming they form known functional sub-complexes; however, Rlig1 and ANGEL2 formed no interaction connections, so a complementary interrogation of the IntAct database was done for Rlig1 which resulted in four curated physical interactors, providing candidate anchors to investigate potential mechanistic links to tRNA repair.

These findings support an association between Rlig1 and the human tRNA repair machinery.

Contents

1. Introduction 4

2. Related Work 5

3. Methodology 5

4. Results 11

5. Discussion 15

6. Limitations and Future Work 15

7. Conclusion 16

References 17

1. Introduction

Transfer RNAs (tRNAs) are essential for protein synthesis, and their damage under cellular stress can stall translation. Unusually of what has been known for human cells, where damaged RNA is typically targeted for degradation, the dedicated RNA-repair ligase Rlig1 can rejoin broken RNA fragments to preserve translation during oxidative stress. Because disturbances in tRNA levels or functional integrity affect cellular homeostasis, fueling the pathogenesis of neurological disorders and cancer, mapping the functional context of Rlig1 holds direct biomedical relevance (Orellana *et al*., 2022).

While the original co-enrichment study profiled Rlig1 broadly across multiple cellular systems, primarily highlighting ribosomal and exosomal pathways, its association with tRNA processing was noted only narrowly through CLP1 and ANGEL2. Prompted by emerging evidence hypothesizing that tRNAs serve as the primary physiological substrate for Rlig1 (Hu *et al*., 2024), this project performed a targeted reanalysis specifically focused on the tRNA pathway. Using a diligent statistical pipeline, we tested this hypothesis on the mass spectrometry dataset to quantitatively validate whether Rlig1 physically co-enriches with the broader tRNA end-processing and repair machinery.

The analysis rests on affinity-enrichment mass-spectrometry (AE-MS) data, which measure *co-enrichment* (physical association), not direct binding or catalytic function. Interpretations therefore assume the “guilt by association" logic of interactomics, that proteins acting in the same process tend to co-purify. The experiment uses three biological replicates per condition (a small sample size), and missing values require imputation, both of which constrain statistical certainty.

This project is a fully scripted, reproducible reanalysis pipeline built in R/Bioconductor (QFeatures for data handling, limma for differential enrichment). It independently reprocesses the published Rlig1 co-enrichment and systematically tests the enriched proteins against the complete roster of human tRNA end-processing and repair enzymes, replacing the original study’s two-enzyme observation with a pathway-wide result.

2. Related Work

Rlig1 was first identified as a human RNA ligase by chemoproteomic screening (Yuan *et al*., 2023). Subsequent biochemical and structural work established it as a bacteriophage-T4-like 5′→3′ RNA-repair ligase that acts on tRNA *in vitro*, but showed it is not a tRNA-splicing ligase and is predominantly cytoplasmic (Hu *et al*., 2024). Hu *et a*l. further proposed that CLP1, ANGEL2, and Rlig1 might form a human tRNA-repair module analogous to the T4-phage system, but stated explicitly that this coordination was untested. It also distinguished Rlig1 from the canonical human tRNA-splicing ligase RTCB, a GTP-dependent 3′→5′ ligase with opposite end-chemistry (Hu *et al*., 2024). The dataset reanalysed here comes from the co-enrichment study of Stumpf *et al*. (2026), which reported Rlig1's role in ribosome function under oxidative stress and noted enrichment of CLP1 and ANGEL2.

Stumpf *et al*. (2026) study processed its data in Perseus and reported the two repair enzymes qualitatively. This project's advantages are (i) an independent, fully reproducible R/Bioconductor pipeline (a robustness check on the original findings), and (ii) a systematic, pathway-level analysis. The gap it fills is the untested candidate flagged by Stumpf *et al*. (2026) no prior work quantified the co-enrichment of the tRNA end-processing and repair machinery with Rlig1, nor provided the physical-association evidence for the proposed repair module. A limitation shared with prior co-enrichment work is that co-enrichment shows association, not function, which this project acknowledges rather than over claims.

3. Methodology

Data and software The analysis was performed in R using Bioconductor packages. The input was a Spectronaut v16 library-free directDIA report at the protein-group level (~3,271 protein groups across 18 samples), from the affinity-enrichment mass-spectrometry experiment of Stumpf *et al*. (2026). The report was read with **read.csv**`(check.names = FALSE)` to preserve the exact Spectronaut column names, and the sample metadata was read from a tab-separated file with **read.delim**. Only the per-sample **.PG.Quantity** columns were used as protein abundances.

Data infrastructure (QFeatures) The abundance matrix, protein annotation, and sample metadata were assembled into a single QFeatures object using **readQFeatures**, with the 18 quantity columns as the assay and **PG.ProteinGroups** as feature identifiers. QFeatures was chosen because it keeps the quantitative data, feature annotation, and sample metadata synchronised through every processing step.

Contaminant removal 57 protein groups flagged by Spectronaut with a Cont_ prefix were removed before any quantitative processing, leaving 3,214 protein groups.

Missing-value assessment Missingness was quantified per sample and per protein with **nNA**, then visualised as a bar chart of per-sample missing proportion.

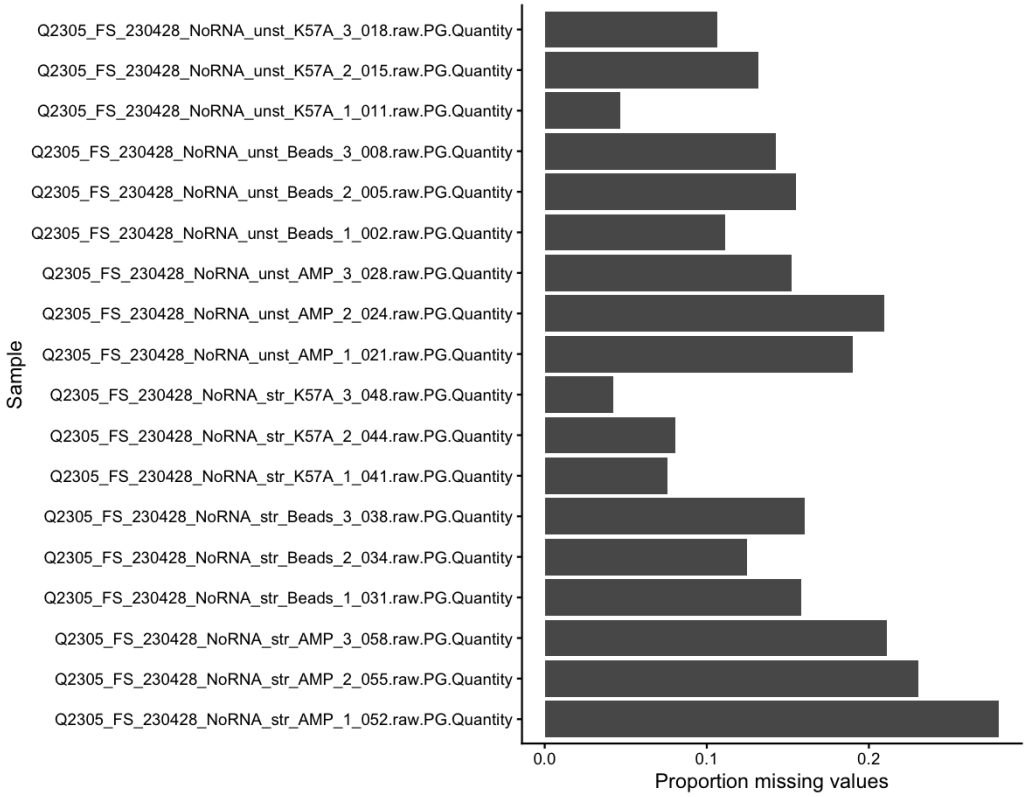


Figure 3.1 Proportion of missing protein values across all 18 proteomics samples. Missing data remained low overall, staying under 28% across all experimental conditions. The K57A mutant samples showed the highest data completeness with the lowest missingness rates (4–13%), while bead controls were consistently moderate at 11–16%. AMP-Rlig1 samples exhibited the highest proportion of missing values (15–28%), which peaked under stressed conditions.

Log2 transformation Intensities were log2-transformed with **logTransform**(`base = 2`). Raw mass-spectrometry intensities are strongly right-skewed so the log2 transform renders the distribution approximately symmetric and makes fold changes directly interpretable. Transformation was applied before imputation so that all subsequent steps operate on a scaled data.

Imputation (k-nearest-neighbours) Missing values were imputed with **QFeatures::impute**(`method = "knn", MARGIN = 1`), with **set.seed**(118) to ensure full reproducibility. kNN imputes a missing value from other proteins with similar abundance profiles across samples. The result was checked by overlaying the distributions of observed and imputed values as a histogram, confirming the imputed values were placed sensibly relative to the measured data.

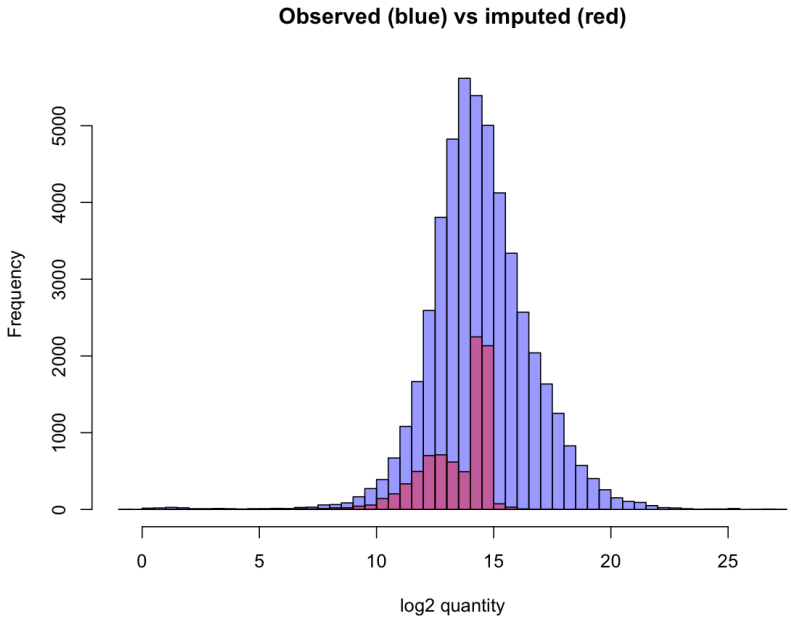


Figure 3.2 Distribution of observed (blue) and imputed (red) log2 protein quantities across the dataset. Observed values follow a normal distribution centered around a log2 quantity of 14–15. Imputed values are concentrated in the lower-to-mid abundance range (log2 quantity 10–15), replacing missing values primarily near lower detection limits without distorting the overall intensity profile.

Normalisation (median) The imputed assay was normalised with **normalize** (`method = "diff.median"`), which aligns each sample's median log-intensity to a common value. Normalisation was verified with before/after boxplots and group-coloured density plots, confirming the sample medians aligned while the group-specific distribution shapes were preserved.

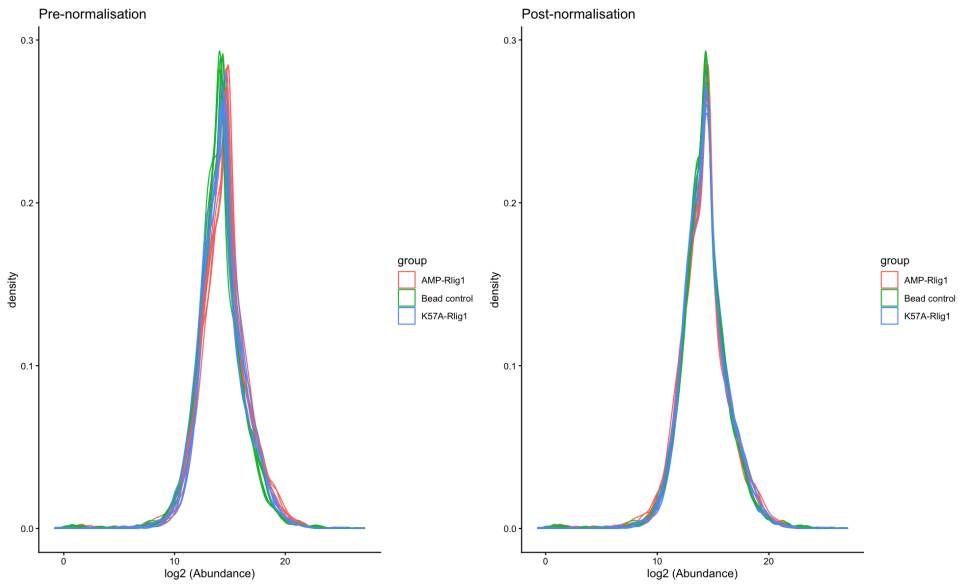


Figure 3.3. Density distributions of log2 protein abundances before (left) and after (right) normalization across AMP-Rlig1 (red), K57A-Rlig1 (blue), and Bead control (green) samples. Subtle baseline shifts prior to normalization were corrected, aligning signal intensity distributions across all samples into a uniform profile centered around log2 abundance ~14.

Quality control (PCA) Principal component analysis was performed on the normalised matrix with **FactoMineR::PCA**(`scale.unit = TRUE`), and the variance structure was inspected with factoextra::fviz_screeplot. The first two components (35.6% and 19.7% of variance) were plotted, coloured by bait and shaped by stress state, to confirm that replicates clustered by their group and stress condition.

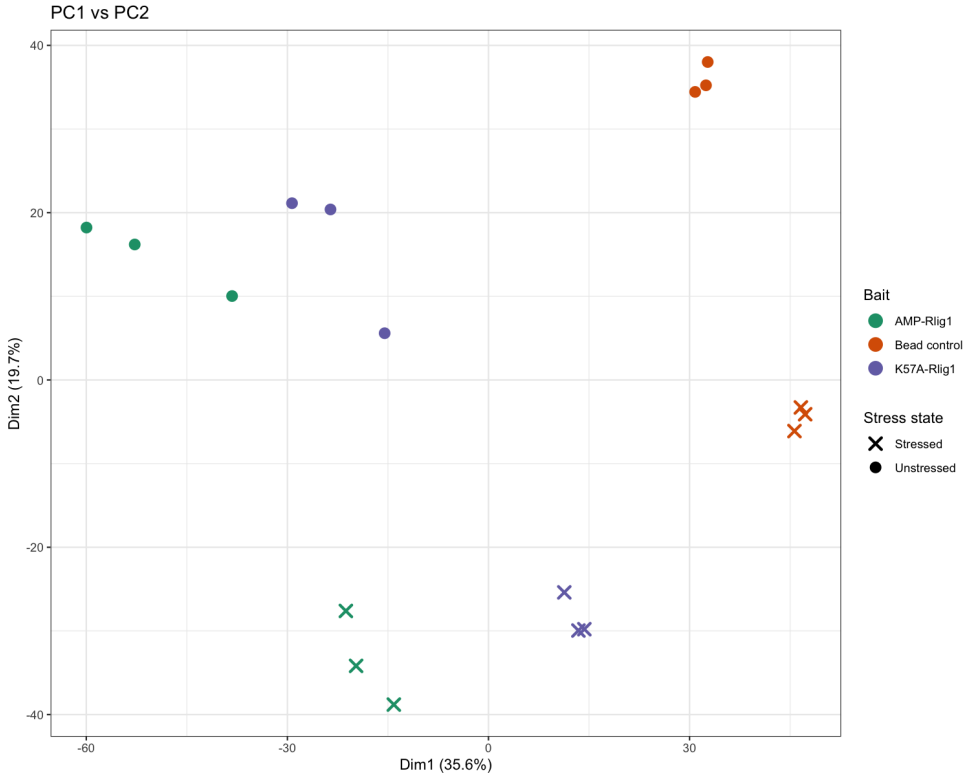


Figure 3.4. Principal component analysis (PCA) of proteomic profiles across bait conditions and stress states. Dim1 (35.6% of variance) primarily separates samples by bait type, distinguishing AMP-Rlig1 (green) and K57A-Rlig1 (purple) from Bead controls (orange). Dim2 (19.7% of variance) reflects the impact of cellular stress, cleanly separating unstressed (circles) from stressed (crosses) conditions. Replicates within each group cluster together, demonstrating high experimental reproducibility.

Differential enrichment (limma) A means-model design matrix (`model.matrix(~ 0 + condition)`) was used so that each of the six conditions received its own coefficient. Six biologically-motivated contrasts were defined with makeContrasts: **AMP_vs_Beads**, **K57A_vs_Beads**, and **AMP_vs_K57A**, each in both stress states. The model was fitted with lmFit, re-expressed in terms of the contrasts with **contrasts.fit**, and moderated with **eBayes**(`robust = TRUE, trend = TRUE`). Results were extracted per contrast with **topTable**, using Benjamini–Hochberg (BH) false discovery rate correction to control for multiple testing across ~3,200 proteins, and merged with gene and description annotation into results table. Enrichment was defined as BH-adjusted $p < 0.05$ and log2 fold change > 1 (≥ two-fold).

Targeted pathway over-representation (Fisher's exact test) A gene set of the human tRNA end-processing and repair enzymes (the TSEN endonuclease, RNase P subunits, ELAC1/2, the end-healing enzymes CLP1/RTCA/ANGEL2, the RTCB ligase complex, and TRNT1) was defined. For the AMP_vs_Beads (unstressed) contrast, each testable protein was flagged as tRNA-machinery or not, plus as enriched or not, and a 2 × 2 contingency table was built. A one-sided Fisher's exact test, **fisher.test**(`alternative = "greater"`) tested whether the tRNA machinery was over-represented among the enriched proteins.

Model check The distribution of raw p-values was plotted per contrast (geom_histogram, faceted by contrast) as a standard diagnostic.

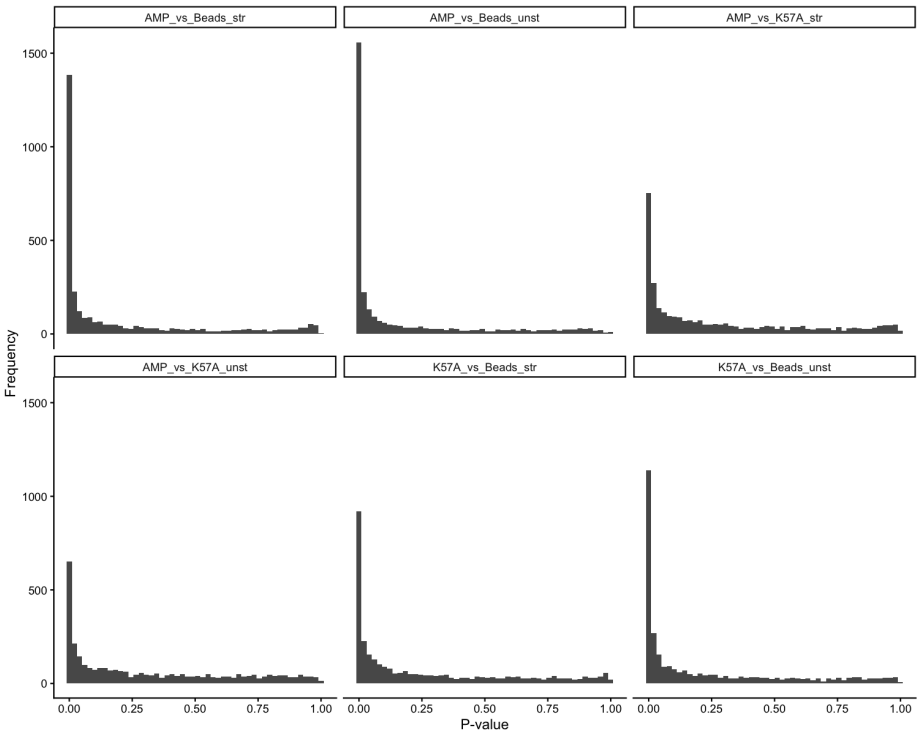


Figure 3.5. Distribution of raw p-values across six pairwise differential abundance comparisons between bait conditions (AMP-Rlig1, K57A-Rlig1, and Bead control) under stressed and unstressed states. All comparisons show a pronounced peak near zero, indicating a strong subset of statistically significant protein changes.

Functional enrichment (Gene Ontology) Gene Ontology over-representation analysis was performed with **clusterProfiler::enrichGO** against the org.Hs.eg.db human annotation, for the four bait_vs_bead contrasts. The enriched proteins (BH < 0.05, logFC > 1) were tested against a background of all proteins quantified in that contrast, rather than the whole genome. All three ontologies were computed (`ont = "ALL"`), but for display only the Molecular Function terms of the primary contrast were shown as a dot plot.

Visualisation The tRNA-machinery results were displayed as pathway-grouped bar charts of log2 fold change coloured by significance (**ggplot2**), and as a row-Z-scored abundance heatmap across all 18 samples (**pheatmap**), with columns fixed in bait × stress order, rows clustered by correlation distance (**ward.D2**), and bait/stress annotation bars. An UpSet plot (**UpSetR**) of proteins shared across the six contrasts was generated to identify contrast-specific versus ubiquitous (background) proteins, as a basis for future work.

Network connections STRING and IntACT websites were used to see if there is any known connections between Rlig1 and the tRNA enzymes involved the repair pathway.

Additional methodological notes

- **MinProb** and **mixed** imputation approaches were evaluated, but both introduced bimodal normalization profiles and downstream plot noise. **KNN** was chosen as it resolved both artifacts cleanly.
- **DEqMS** was tested alongside standard limma empirical Bayes (**eBayes**). Outputs were similar, so I went with **eBayes** because I am more familiar with it.
- Initial reanalysis agreed with the original paper on the main biological findings, so The workflow was then refined from a global reproduction to a targeted analysis addressing the tRNA pathway.

4. Results

The broader aspect

Gene Ontology analysis of the active-Rlig1 recovers a strongly RNA and translation centric profile, reproducing the original study's functional themes by an independent method and confirming the pipeline is sound.

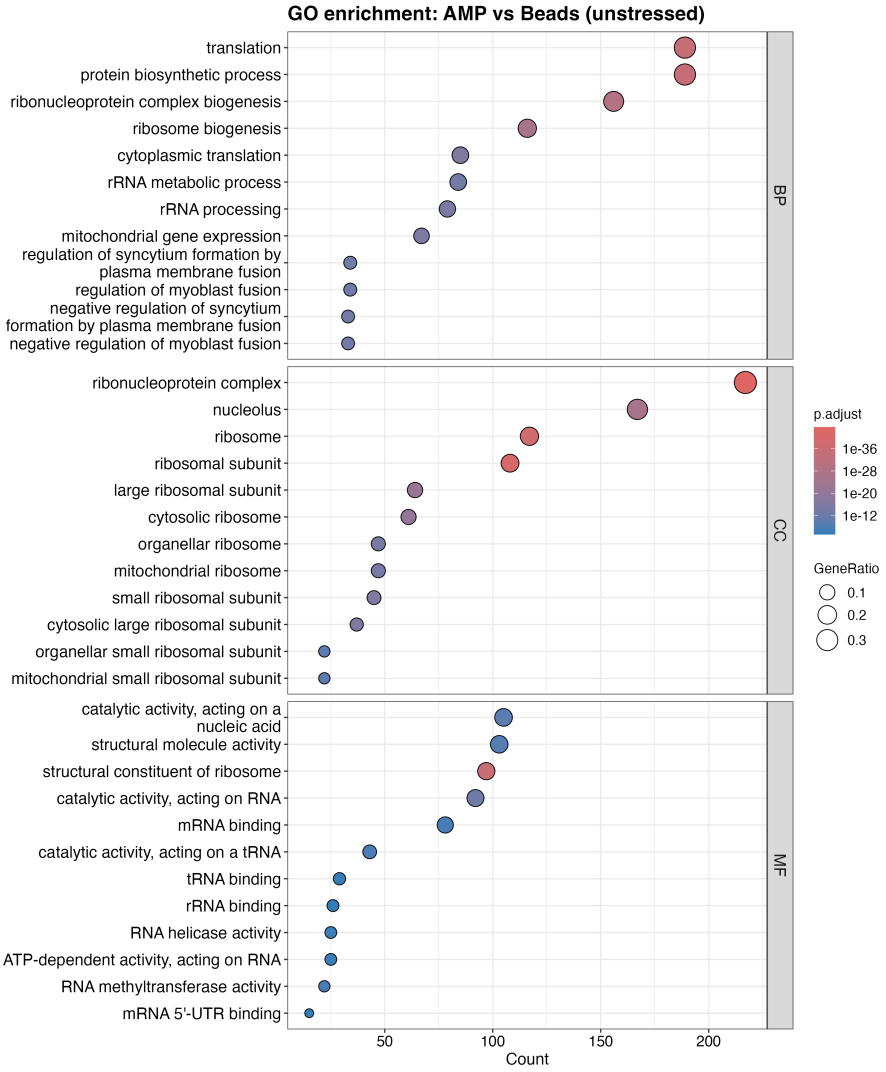


Figure 4.1. Gene Ontology (GO) enrichment analysis of proteins enriched in AMP-Rlig1 versus Bead control under unstressed conditions, categorized into Biological Process (BP), Cellular Component (CC), and Molecular Function (MF). Enriched terms are strongly dominated by translation, ribosome biogenesis, and RNA processing (BP), localized primarily to ribonucleoprotein complexes, the nucleolus, and ribosomal subunits (CC). Corresponding molecular functions prominently feature RNA and tRNA binding, structural ribosomal constituents, and nucleic acid catalytic activities (MF). Dot size reflects the gene ratio, while color indicates adjusted p-value, highlighting extreme statistical enrichment (padjust $<10^{-30}$) across protein synthesis machinery.

Investigating tRNA repair pathway

Of 22 detected tRNA end-processing and repair enzymes, **19 are significantly enriched** with active Rlig1 (FDR < 0.05, $\log_2$ FC > 1), spanning across most of the pathway (end-healing, 5′/3′ processing, ligation, maturation). Against the **baseline** enrichment rate of all other quantified proteins (~20%), the tRNA machinery is enriched at **86%** (19/22). The targeted Fisher's exact test gives $\mathbf{p = 3.9 \times 10^{-11}}$, establishing that this over-representation is far beyond chance.

The original study reported only two enriched enzymes (CLP1, ANGEL2), the expected outcome if that were the full picture would be a handful of isolated hits. Instead, most of the pathway co-enriches, which extend the two-enzyme observation to the broader pathway.

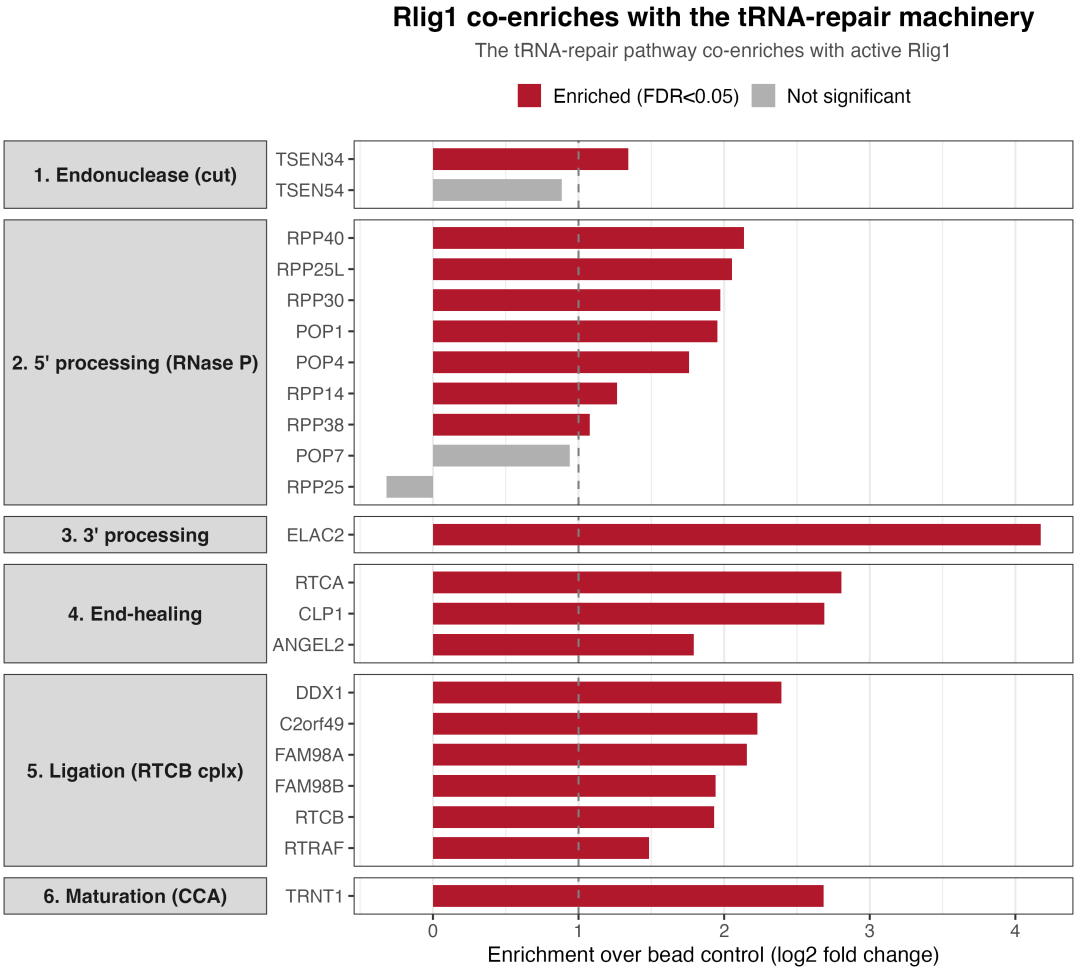


Figure 4.2. Co-enrichment of tRNA processing and repair machinery with active Rlig1 relative to bead controls across six functional pathway stages. Red bars mean statistically significant enrichment (FDR<0.05), while grey bars indicate non-significant proteins. Active Rlig1 broadly co-enriches with key pathway components such as: endonuclease cutting, 5' processing, 3' processing, end-healing, ligation, and CCA maturation. ELAC2 shows the highest enrichment, and notably, there is a complete co-enrichment across the entire RTCB complex as well.

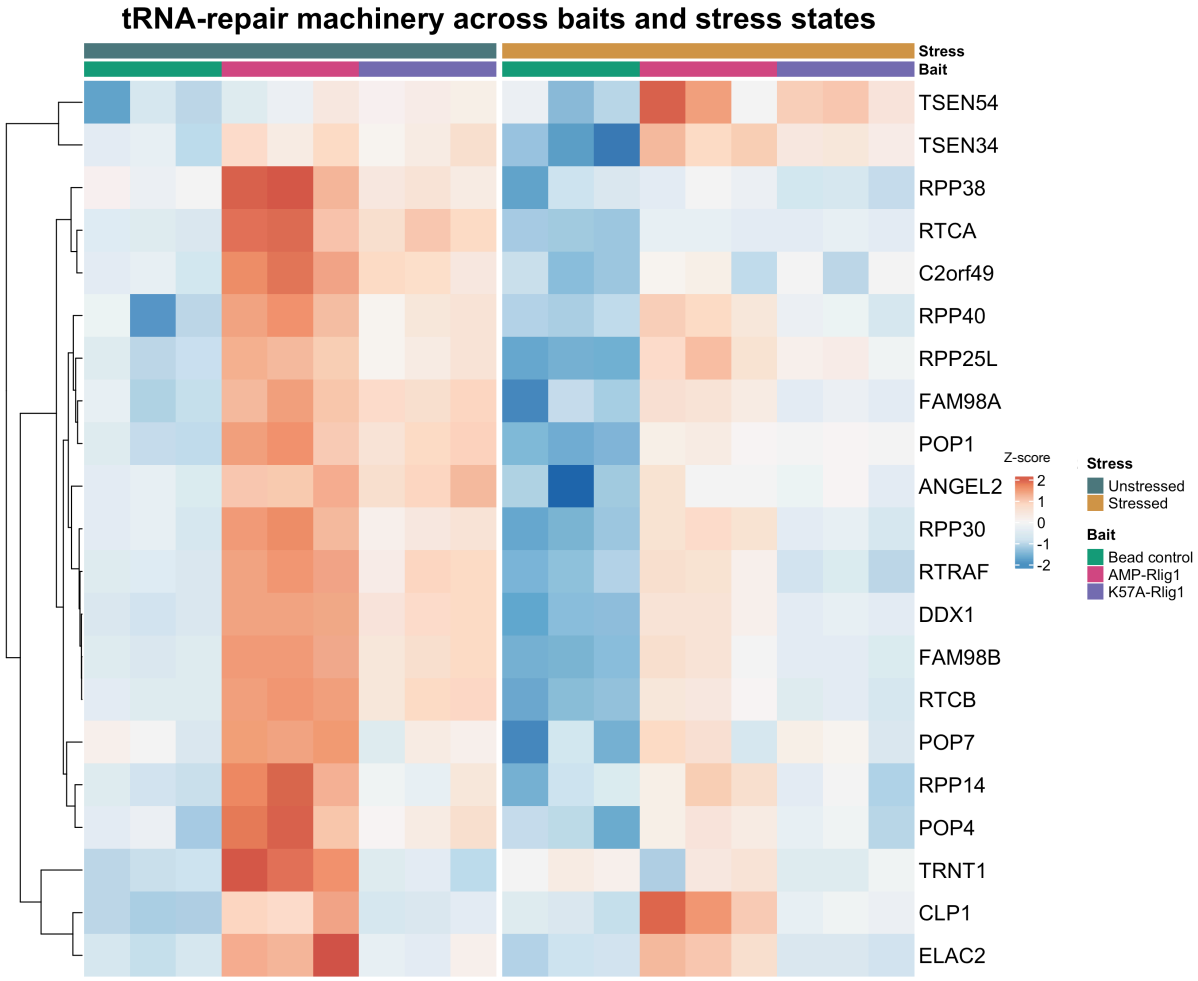


Figure 4.3. Row Z-scored abundance (log2, normalised) of the detected tRNA end-processing and repair enzymes across all 18 samples. Columns are ordered by stress state (unstressed, then stressed) and within each by bait (bead control, AMP-Rlig1, K57A-Rlig1), rows are clustered by correlation distance (ward.D2). Enzymes show high relative abundance in the AMP-Rlig1 and K57A-Rlig1 baits compared to bead controls, in both stress states. Color indicates row Z-score (red = high, blue = low).

As shown in **Figure 4.3**, tRNA end-processing and repair enzymes consistently enriched with both AMP-Rlig1 and K57A-Rlig1 baits compared to controls across all 18 samples. This interaction remained stable under both stressed and unstressed conditions. Because the enzymes clustered together by abundance, they appear to function as a organized group rather than isolated hits.

STRING clustering independently recovered known sub-complexes (RNase P, RTCB complex, ELAC2–TRNT1), confirming the biological validity of the co-enriched dataset. Because Rlig1 (C12orf29) and ANGEL2 lacked documented interactions within the STRING network, we performed a targeted query in IntAct to map known direct interactors of Rlig1. Robustness checks show the co-enrichment holds under stress (~19/22 in both states) and is partially retained by the inactive K57A mutant (13/22).

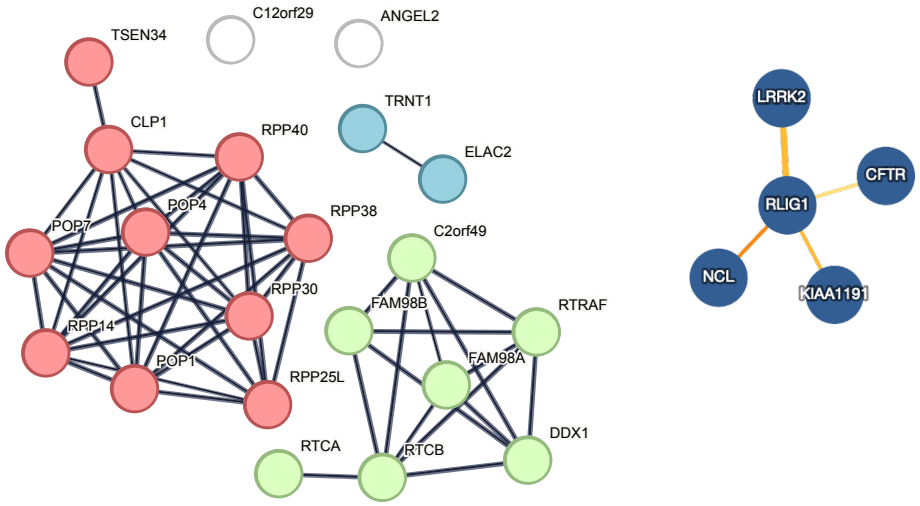


Figure 4.4. Functional and physical networks of RLIG1. RLIG1 (C12orf29) co-enriched proteins have a clear functional relationship as evidenced by the STRING network (min. score: 0.7), such proteins are co-enriched independently from direct interactions with RLIG1, whose interactions are shown on the right-hand side (data taken from IntACT)

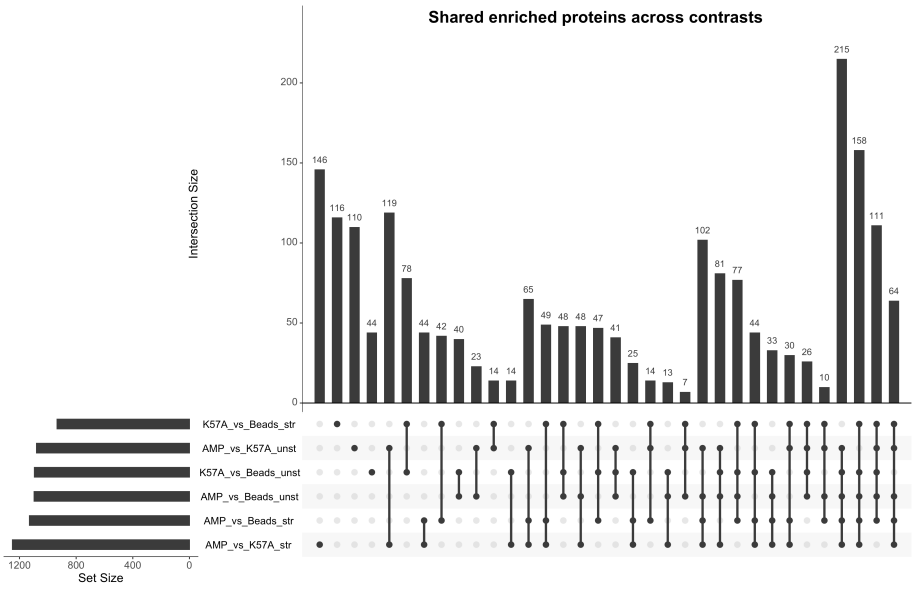


Figure 4.5. Intersection analysis of enriched proteins across experimental conditions. UpSet plot displaying the shared and unique significantly enriched proteins (FDR<0.05) across six pairwise contrasts. Bottom-left horizontal bars indicate the total set size (number of enriched proteins) for each contrast. The dot matrix illustrates specific combinations of overlapping contrasts, with vertical bars showing the corresponding intersection sizes (number of shared proteins). The broad variation in intersection sizes highlights both baseline pathway-wide target sharing and condition-dependent protein shifts across bait types (AMP-Rlig1, K57A-Rlig1, Bead control) and stress states (stressed vs. unstressed).

5. Discussion

Building on prior work positioning tRNA as a substrate of Rlig1 (Hu *et al*., 2024), Stumpf *et al*. (2026) study initially highlighted co-enrichment with ribosomal and exosome proteins, alongside only two specific repair factors: ANGEL2 and CLP1. Driven by the tRNA-substrate hypothesis, our targeted re-analysis of the broader tRNA processing and repair pathway revealed that 19 out of 22 pathway proteins co-enriched with active Rlig1 ($p<10^{-10}$). Rather than yielding scattered hits, Rlig1 co-purifies with most of the pathway apparatus. This broad co-enrichment strongly validates the tRNA-substrate model and expands the proposed CLP1–ANGEL2–Rlig1 module into a pathway-wide interaction network.

These findings are further refined by two critical nuances. First, the isolation of Rlig1 and ANGEL2 within STRING network clustering is something that is worth noting rather than a methodological limitation, because it might reflect their newly recognized roles that remain unannotated in curated databases. Second, the robust co-enrichment of the RTCB complex introduces an unexpected dynamic, given that RTCB operates through a distinct chemical mechanism and is classically localized

While these bioinformatic findings place Rlig1 within human tRNA repair pathway, computational co-enrichment alone cannot resolve binding architecture or direct physical dynamics. These insights emphasize the necessity for downstream wet-lab and structural studies to experimentally validate these interactions and uncover the precise nature of the crosstalk between Rlig1, RTCB, and the broader repair machinery.

6. Limitations and Future Work

While these findings provide somewhat of an evidence for the module, several limitations should be considered. First, co-enrichment demonstrates association rather than direct binding or enzymatic function. Second, the sample size (n=3) is small, and imputed values are inherently less reliable than measured ones, even though I used eBayes moderation and only used real values for core proteins to minimize this impact. Finally, the statistical testing was confirmatory on a single pre-defined gene set and limited to one dataset.

To build on these findings, further work is needed in a few key areas. Experimentally, validating direct interactions using co-immunoprecipitation or proximity labeling using BioID, along with in vitro functional assays, will be essential to confirm biological activity.

Additionally, regarding the complex overlaps shown in the UpSet plot, I originally intended to apply more advanced multi-factorial models to better capture conditional dependencies and reciprocal protein shifts across stress states, however, time constraints prevented this from being included in the current project, making it a promising direction for future analysis.

7. Conclusion

This reanalysis positions Rlig1 as a primary candidate ligase in the human tRNA-repair landscape, showing that almost the entire repair machinery co-enriches with the active protein.

This is an association, not a demonstrated mechanism; defining Rlig1's substrates, its catalytic role, and how oxidative stress regulates it are the essential next steps.

By turning a narrow observation into testable, pathway-wide hypotheses, this reanalysis reframes Rlig1 from an isolated enzyme into a candidate component of a repair system we are only beginning to understand.

References

Stumpf FM, Glauner M, Jansen J, Marchand V, Motorin Y, Stengel F, Marx A. Human RNA ligase 1 as a novel regulator of ribosome function and translation under oxidative stress. *Nucleic Acids Research.* 2026;54(11):gkag528. doi:10.1093/nar/gkag528

Orellana EA, Siegal E, Gregory RI. tRNA dysregulation and disease. *Nature Reviews Genetics.* 2022;23(11):651–664. doi:10.1038/s41576-022-00501-9

Hu Y, Lopez VA, Xu H, Pfister JP, Song B, Servage KA, Sakurai M, Jones BT, Mendell JT, Wang T, Wu J, Lambowitz AM, Tomchick DR, Pawłowski K, Tagliabracci VS. Biochemical and structural insights into a 5′ to 3′ RNA ligase reveal a potential role in tRNA ligation. *Proceedings of the National Academy of Sciences.* 2024;121(42):e2408249121. doi:10.1073/pnas.2408249121

Yuan Y, Stumpf FM, Schlor LA, Schmidt OP, Saumer P, Huber LB, Frese M, Höllmüller E, Scheffner M, Stengel F, Diederichs K, Marx A. Chemoproteomic discovery of a human RNA ligase. *Nature Communications.* 2023;14:842. doi:10.1038/s41467-023-36451-x