A110 · PPI-seq: A Massively Parallel System to Decode Genetic Variant Impacts on Protein Interactions

Presenter: Justin Delano — Harvard Medical School · Massachusetts General Hospital

Authors: Justin Delano, Patrick Cann, Nicolas Rey, Tian Yu, Luca Pinello, Richard Sherwood

Session: Day 2 · Fri Oct 2, 2026 · 1:15–2:15 PM

Protein-protein interactions organize cellular function, yet quantitative measurement of how they change across partners, conditions, and coding variants remains scarce. No existing approach combines native mammalian cellular context, variant-level resolution, and multiplexed throughput — a gap that particularly limits variant interpretation. We developed PPI-seq, a pooled two-hybrid assay that couples protein-protein interaction to a sequence-readable signal inside mammalian cells. Each bait protein is fused to an adenine base editor, and each prey protein is fused to an RNA binding protein; interaction between the two proteins triggers editing of a barcoded RNA reporter, and sequencing quantifies such edits in parallel. Extracting reliable interaction scores from these count data requires accounting for library size, transfection efficiency, and sample-to-sample variation. We therefore developed PyroPPI, a Bayesian hierarchical model that isolates the interaction-specific signal from confounders and reports calibrated confidence intervals. We applied PPI-seq to the BCDX2 complex, a DNA repair complex whose pairwise contacts are well established from biochemical studies. PyroPPI recovered the known binding pairs that were obscured in the unprocessed data. We then performed deep mutational scanning of RAD51D and XRCC2, testing ~3200 variants. Stop-gain variants showed the greatest reduction in binding, synonymous variants remained near baseline, and missense variants exhibited graded intermediate effects, with positions in the conserved Walker A and B ATPase motifs among the most sensitive. PPI-seq and PyroPPI together provide a scalable platform for quantifying how coding variation reshapes protein interaction networks in their native setting.

PPI-seq: A Massively Parallel System to Decode Genetic Variant Impacts on Protein Interactions

Justin Delano*, Patrick Cann, Nicolas Rey, Tian Yu, Luca Pinello, Richard Sherwood

Introduction

Missense substitutions are the most common protein-coding variation, yet whether one alters binding to a given partner in living cells is largely unmeasured, leaving many clinical alleles of uncertain significance. Yeast two-hybrid is heterologous, co-immunoprecipitation and proximity labeling lack throughput, and predictors such as ESM and AlphaMissense score single proteins rather than interactions. We developed PPI-seq, a pooled two-hybrid assay that records interaction as sequence inside mammalian cells, with PyroPPI, a Bayesian model returning calibrated interaction scores per pair.

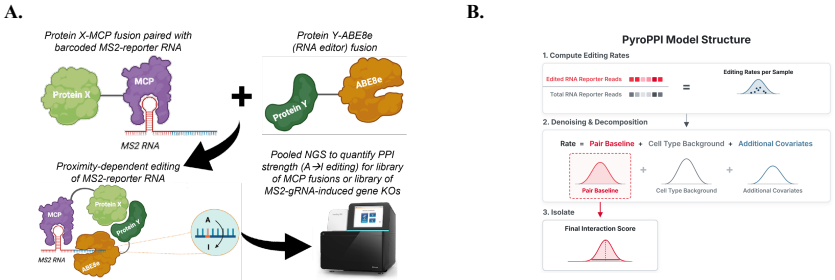

Fig 1. PPI-seq and PyroPPI. A, the two-hybrid assay and its sequencing readout. B, PyroPPI denoises and decomposes PPI-seq counts into latent interaction propensities per pair.

Methods

Each plasmid encodes an adenine base editor (ABE8e) fused to one protein, MCP fused to a second, and a barcoded RNA reporter. Interaction brings the editor to the reporter, converting adenines to inosines; sequencing reads out editing per barcoded pair. Thousands of constructs are pooled and transfected into human cell lines across replicates. PyroPPI models edited counts as binomial, decomposing editing rate into additive pair, cell-type, and condition terms fit by Hamiltonian Monte Carlo in NumPyro, with calibration verified by posterior predictive checks.

Results

PyroPPI profiled all pairwise combinations of a nine-protein panel around the BCDX2 RAD51-paralog complex in HCT116 and U2OS cells, recovering the established pairs (RAD51B-RAD51C, RAD51D-XRCC2, RAD51C-XRCC3) with significant separation from GFP and mCherry controls. We then tested 3180 variants tiling every codon of RAD51D (306 positions) and XRCC2 (260), five missense per position plus synonymous and stop-gain controls. Stop-gains lost the most interaction, synonymous variants stayed near baseline, and missense spanned a graded range, with the Walker B motif of RAD51D and Walker A motif of XRCC2 most sensitive. Of 2,892 missense variants, 90 showed strong loss yet were missed by at least one predictor (80 called benign by AlphaMissense, 54 unflagged by ESM), enriched near the binding interface.

Discussion

PPI-seq measures interaction in native mammalian context at single-amino-acid resolution and thousand-variant scale, recovering known BCDX2 pairs and grading variant effects that predictors miss. Software and data will be released openly. As coverage grows, such maps supply the training data predictors lack: not whether a substitution is deleterious, but which partnership it disrupts.