**A new paradigm for mRNA display selection by sequence space analysis**

Tommy Y Lu1,2, John Chen3, Colin Jackson2,3, Richard Payne1,2. School of Chemistry, University of Sydney1, Sydney, NSW, Australia; Australian Research Council Centre of Excellence for Innovations in Peptide and Protein Science2; Research School of Chemistry, The Australian National University3, Canberra, ACT, Australia

A graph showing a number of columns

Description automatically generated with medium confidence**Introduction.** mRNA display is a screening technique that can select peptides against a drug target through iterative binding and amplification cycles. This method is used in drug discovery where sequences that enrich for the target can be optimised for therapeutic use. Taking the most enriched peptides as the top hits is the current conventional approach. Yet this method is fallible, as false or non-specific hits can arise due to insufficient screening, random distribution and inherent biases in the initial library. We propose that investigating sequence space can assist in verifying high affinity ligand selection across selection rounds.

**Aims**. Establish benchmarks for deep sequencing data quality and mRNA display selections across rounds. Assess these benchmarks across multiple studies to ascertain the quality of hits identified under the current standard of display selection.

**Methods**. A pipeline was developed to extract, clean, and unify deep sequencing data. A sequence space model was designed by evaluating the complementarity of embedding and dimensionality reduction methods. Models of theoretical distribution and random selection were employed as a baseline to quantify and characterise round selection.

**Results.** Sequence space benchmarks effectively delineated between datasets by metrics of quality. This highlighted key areas of improvement for studies that obtained low quality or unexpected results. Correlation between quality of round selection and biological properties of tested hits affirm the significance of the established benchmarks.

**Discussion.** The inadequacy of selection in several studies demonstrates the necessity of improved rigour in mRNA display deep sequencing screens. Sequencing across more rounds with deeper reads would ensure higher quality and preserve the consistency of enrichment. Observed biological correlations posit the potential for improved peptide hits with greater affinity for their drug targets.