A Novel Method For Generating Unprecedented Ultra-High-Sequence Diversity Libraries Of Plasmids And Recombinant Viruses (With Validation) For High-Throughput Screening, drug discovery And Targeted Therapy.

Praveen Singh Hajeri (phajeri@umn.edu), University of Minnesota, **Praveensingh Hajeri**, University of Minnesota, **Masato Yamamoto**, University of Minnesota

Background: The probability of finding a good hit during drug discovery/screening primarily depends on the quantity (diversity) and quality (distribution) of libraries. Libraries include peptides, single-chain antibodies, nanobodies, CRISPRs, shRNAs, phage display, yeast display, ribosome display, BiKEs, TriKEs, CAR-T cells, and recombinant viruses (including adenoviruses, lentiviruses, AAVs, etc.).

Generating high-sequence diversity libraries is very challenging. Several published studies and industries claim to have them. However, most of those libraries lack stringent experimental validation and depend on unreasonable statistical extrapolations. Therefore , there is an imminent need for a method to generate genuine high-sequence-diversity libraries of high quality.

Methods and results: We have developed a novel method to generate libraries with high sequence diversity (>109) and validated it with rigorous NGS-based methods. Compared to others, our libraries are at least 100 to 1000 times better in diversity (quantity) and have a more uniform and unbiased distribution (quality). Benchmarking studies validated diversity, duplication rate, sequence bias, ability to scale up , and screening.

Conclusions: Our unprecedented high-quality, unbiased libraries and serial screening results can serve as a dataset for machine training to aid AI-based design and high-throughput screening. We expect our libraries may help in avoiding AI-based design altogether in some cases.