A063 · BaseEvolve: AI-guided directed evolution of large serine recombinases for therapeutic gene insertion

Presenter: Aaron Kollasch — Basecamp Research

Authors: Aaron Kollasch, Jenna Hoersten, Matthew Bakalar, Tanggis Bohnuud, Jonathan Finn

Session: Day 2 · Fri Oct 2, 2026 · 2:15–4:15 PM

Large serine recombinases (LSRs) catalyze unidirectional, site-specific integration of multi-kilobase DNA cargoes, making them attractive tools for therapeutic gene insertion. However, their native attB target sites are absent from the human genome, requiring LSRs to be re-targeted to endogenous, therapeutically relevant loci. To enable this, we built BaseEvolve, an active learning platform for iterative, AI-guided optimization of LSR activity.

BaseEvolve operates through closed-loop cycles of computational design, synthesis, and experimental testing. In each round, a protein language model is aligned to LSR fitness data using a learn-to-rank objective, and multi-mutant variants are then generated by Gibbs sampling. Variants are expressed by in vitro transcription-translation (IVTT), assayed for recombination activity, and the results are fed back into the next training round.

We demonstrate BaseEvolve in two campaigns. In the first, starting from a naturally occurring LSR variant identified from our genomic database, BaseData, we achieved a 14-fold activity gain over the best round-1 sequence with only five amino acid substitutions from wild type, demonstrating that active learning drives substantial round-over-round improvement. In the second, we applied BaseEvolve to Bxb1 and reached activity levels exceeding previously described LSR evolution methods in just two rounds. Activity gains were confirmed in both IVTT and mammalian cell lysate assays.

BaseEvolve: AI-guided directed evolution of large serine recombinases for therapeutic gene insertion.

Aaron Kollasch*, Jenna Hoersten, Matthew Bakalar, Tanggis Bohnuud, Jonathan Finn
Basecamp Research

Large serine recombinases (LSRs) catalyze unidirectional, site-specific integration of multi-kilobase DNA cargoes, making them attractive tools for therapeutic gene insertion. However, their native attB target sites are absent from the human genome, requiring LSRs to be re-targeted to endogenous, therapeutically relevant loci. To enable this, we built BaseEvolve, an active learning platform for iterative, AI-guided optimization of LSR activity.

BaseEvolve operates through closed-loop cycles of computational design, synthesis, and experimental testing. In each round, a protein language model is aligned to LSR fitness data using a learn-to-rank objective, and multi-mutant variants are then generated by Gibbs sampling. Variants are expressed by in vitro transcription-translation (IVTT), assayed for recombination activity, and the results are fed back into the next training round.

We demonstrate BaseEvolve in two campaigns. In the first, starting from a naturally occurring LSR variant identified from our genomic database, BaseData, we achieved a 14-fold activity gain over the best round-1 sequence with only five amino acid substitutions from wild type, demonstrating that active learning drives substantial round-over-round improvement. In the second, we applied BaseEvolve to Bxb1 and reached activity levels exceeding previously described LSR evolution methods in just two rounds. Activity gains were confirmed in both IVTT and mammalian cell lysate assays.