

Group 18 HDR-GPT: A Large Language Model Agent for Optimizing Homology-Directed Repair Efficiency in CRISPR Editing



Jingwen (Steven) Hui¹, Duy Nguyen¹, Freja Kjellaug Amalia Ekman², Hana Yousef Ghanim³, Sridhar Selvaraj⁴, Matthew Porteus⁴, Le Cong⁵, Xiaohua Huang¹

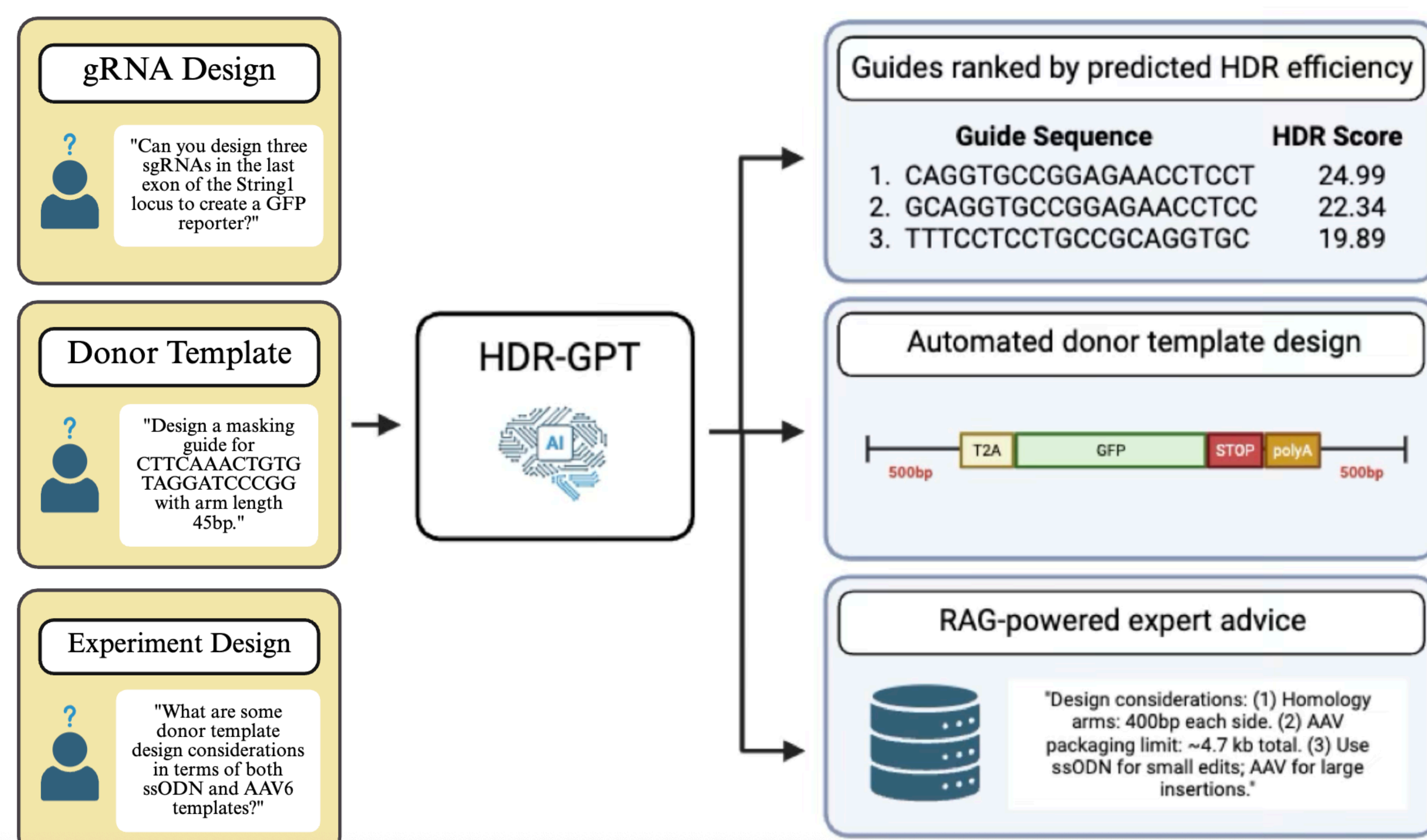
¹Department of Bioengineering, UC San Diego. ²Department of Genetics, Stanford University, ³Stem Cell Biology and Regenerative Medicine Program, Stanford University. ⁴Department of Pediatrics, Stanford University. ⁵Department of Pathology, Stanford University.

Background/Motivation

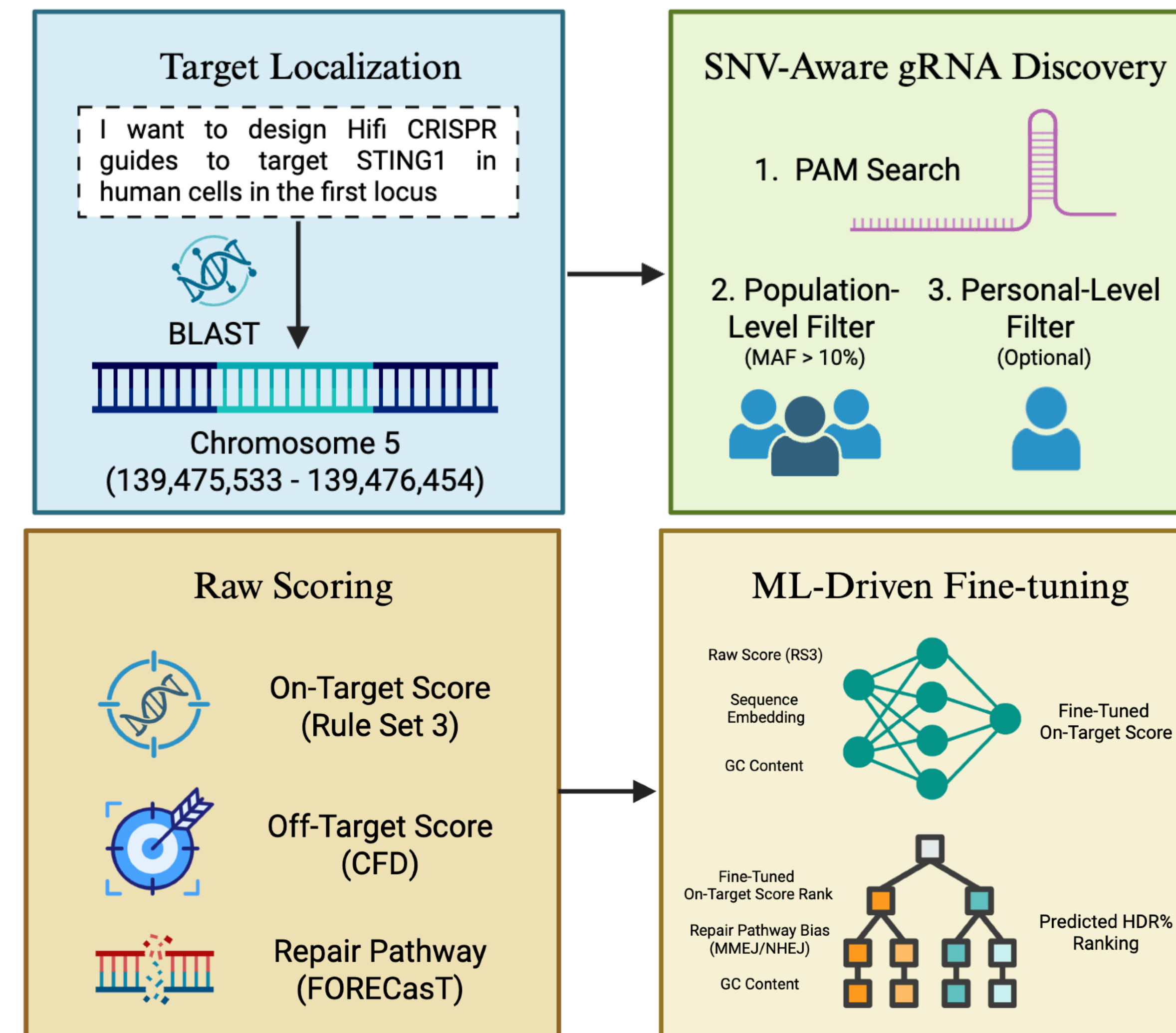
CRISPR-Cas9 has emerged as a highly efficient and programmable tool for genome engineering. Beyond simple gene knockouts, Cas9 enables precise gene editing and transgene insertion through homology-directed repair (HDR), wherein an exogenous donor template guides repair of the double-strand break. HDR is critical for applications in tool development, disease modeling, and gene therapy. However, achieving efficient HDR remains challenging and often requires extensive field expertise, limiting broader adoption. Compounding this issue, existing computational design tools are trained using wild-type Cas9 primarily on immortalized cell line data, which poorly reflects editing outcomes in therapeutically relevant primary cells. To address these gaps, we developed HDR-GPT, an agentic AI platform that guides users through HDR-optimized sgRNA selection and donor template design tailored for primary cell editing with p.R691A high-fidelity Cas9. Additionally, HDR-GPT incorporates a retrieval-augmented generation (RAG) system trained on curated CRISPR and HDR literature, enabling users to query the agent for evidence-based guidance on experimental design.

Architecture

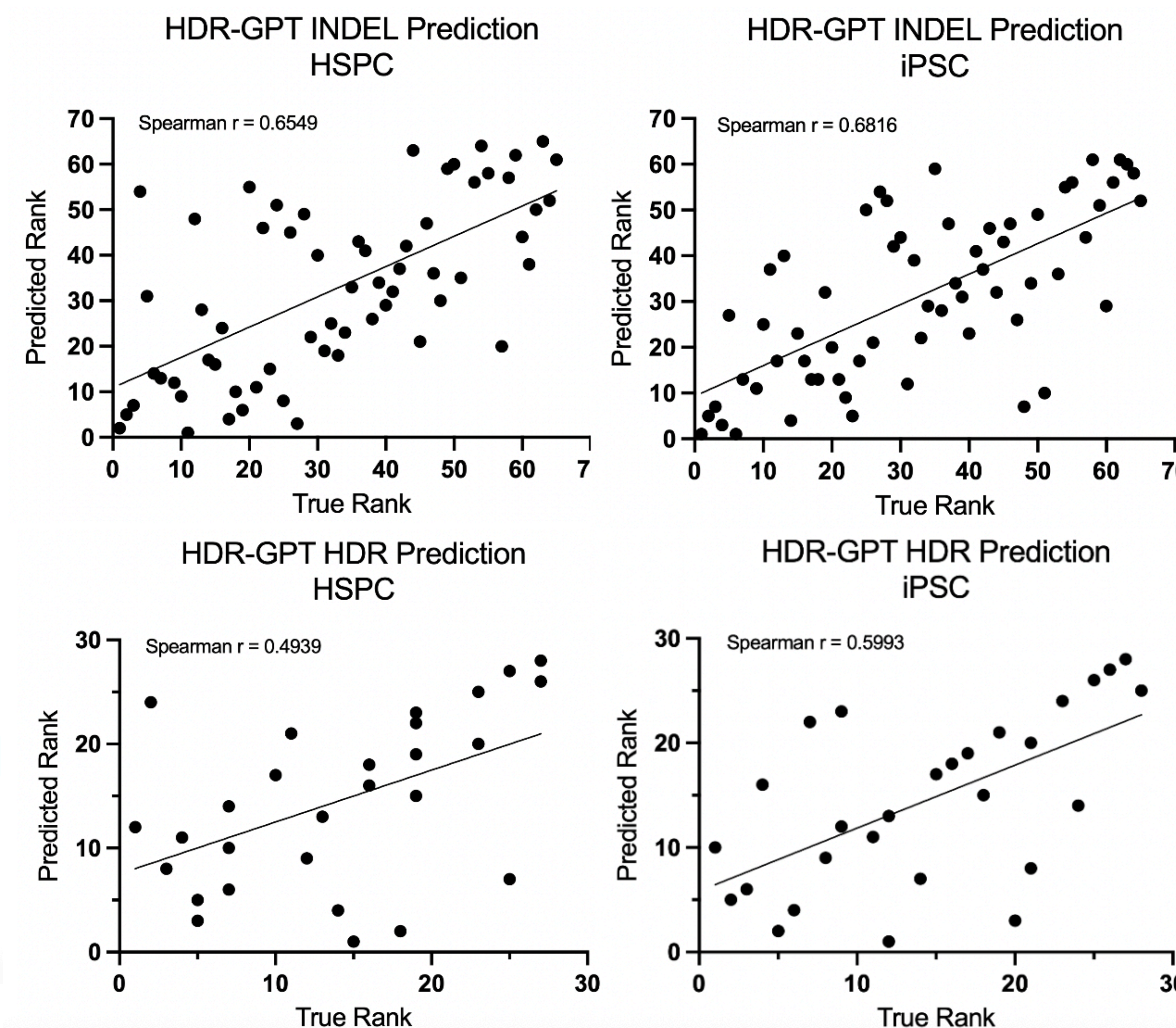
HDR-GPT enhances experimental design, gRNA selection, and donor template design



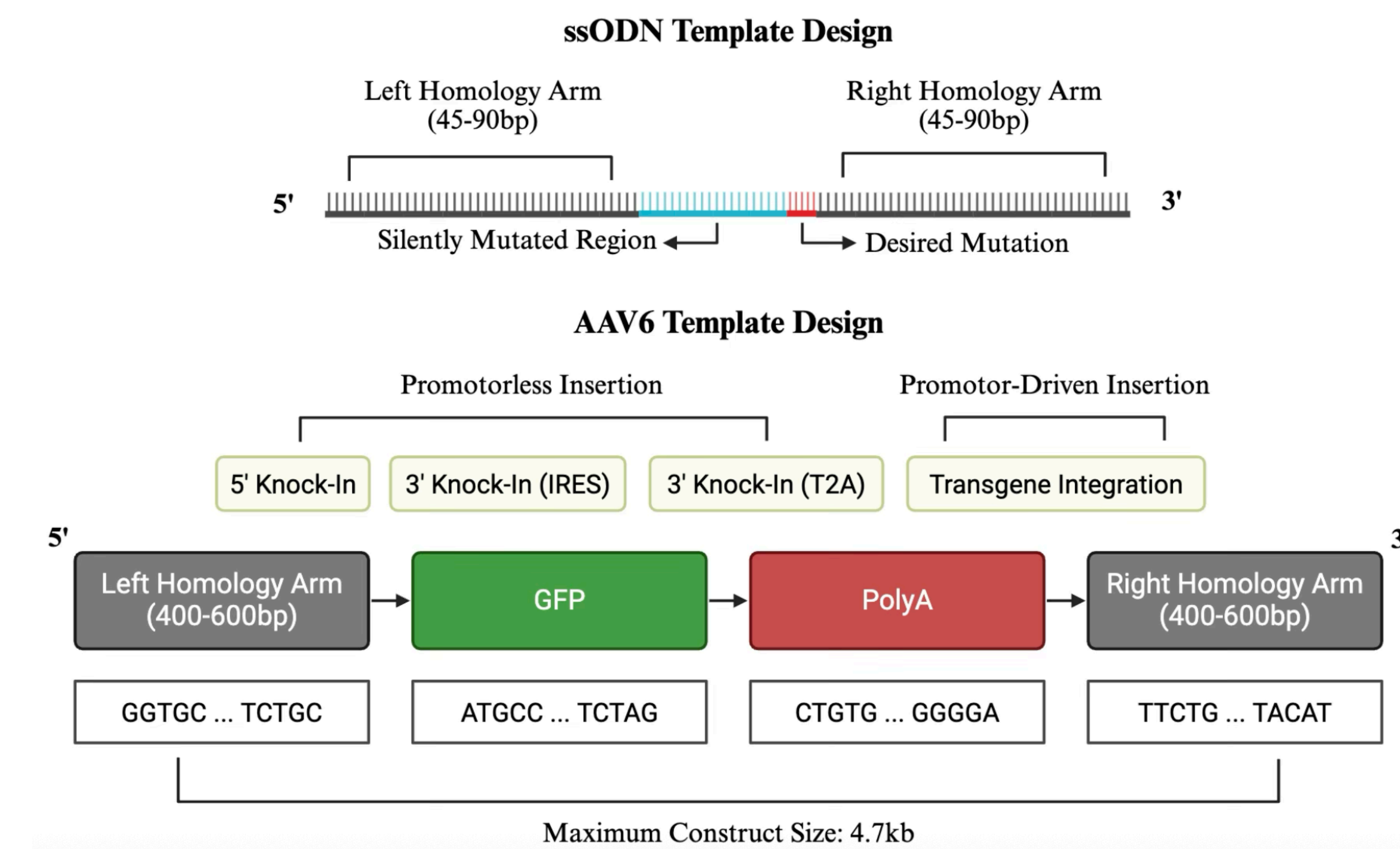
HDR-GPT Pipeline



On-Target Score and HDR Prediction



Automated Donor Template Design



Demo

| GUIDE RNA FINAL SCORES | | | | | | | | | | INDEL | | HDR | |
|------------------------|----------------------|-----|-----|------|-----------------------|--------|-------|-----------|----------|--------|------|-------|--|
| # | Spacer | PAM | Str | Exon | Locus | OnTgt | CFD | MMEJ/NHEJ | FineTune | Final | dATG | dSTOP | |
| 1 | GCAACCTCAACAGACACCA | TGG | - | 1 | chr11:5227828-5227840 | 1.272 | 0.045 | 2.003 | 89.715 | 45.859 | -2 | -1426 | |
| 2 | CCTGGATGAAGTTGGTGTG | AGG | - | 1 | chr11:5226942-5226962 | 0.168 | 0.022 | 8.095 | 75.828 | 41.962 | 76 | -1348 | |
| 3 | CTTGCCCAACAGGCGAGTAA | CGG | + | 1 | chr11:5226967-5226987 | -0.537 | 0.056 | 6.200 | 72.818 | 39.505 | 37 | -1387 | |
| 4 | AGTGTGCGCTTACTGCCCTG | TGG | - | 1 | chr11:5226976-5226996 | 0.969 | 0.081 | 3.788 | 61.864 | 32.826 | 42 | -1382 | |
| 5 | TCTGCCCTTACTGCCCTGTG | GGG | - | 1 | chr11:5226974-5226994 | -1.016 | 0.028 | 8.001 | 62.901 | 31.851 | 44 | -1380 | |
| 6 | CCTGTGCGGCAAGTGAACG | GGG | - | 1 | chr11:5226975-5226995 | -0.018 | 0.094 | 5.333 | 48.902 | 27.118 | 43 | -1381 | |
| 7 | CCTGTGCGGCAAGTGAACG | TGG | - | 1 | chr11:5226968-5226988 | 0.630 | 0.051 | 2.200 | 51.098 | 26.649 | 58 | -1366 | |
| 8 | CACGTTCACCTTGCCCAACA | GGG | + | 1 | chr11:5226958-5226978 | 0.127 | 0.079 | 2.355 | 49.872 | 26.114 | 46 | -1378 | |
| 9 | AAGTGAACGTGGATGAAGT | TGG | - | 1 | chr11:5226958-5226978 | 0.159 | 0.047 | 2.657 | 47.362 | 25.009 | 68 | -1356 | |
| 10 | GAAGTGGTGGTGAAGCCCT | GGG | - | 1 | chr11:5226935-5226955 | 0.579 | 0.020 | 2.835 | 44.620 | 23.328 | 83 | -1341 | |
| 11 | GTGAACGTGGATGAAGTTGG | TGG | - | 1 | chr11:5226947-5226967 | 0.844 | 0.004 | 2.785 | 39.169 | 28.977 | 71 | -1353 | |
| 12 | CSTTACTGCGCTTGGGCCA | AGG | - | 1 | chr11:5226969-5226989 | -0.644 | 0.057 | 1.819 | 36.796 | 19.308 | 49 | -1375 | |
| 13 | CCACGTTACCTTGCACCAC | AGG | + | 1 | chr11:5226957-5226977 | -0.699 | 0.056 | 2.679 | 35.235 | 18.957 | 47 | -1377 | |
| 14 | TGCACCATGGTGTCTTTTG | AGG | + | 1 | chr11:5227813-5227833 | -0.471 | 0.044 | 1.931 | 25.052 | 13.492 | -9 | -1433 | |
| 15 | TGAAGTTGGTGGTGAAGCC | TGG | - | 1 | chr11:5226936-5226956 | -0.476 | 0.027 | 3.283 | 16.752 | 18.018 | 82 | -1342 | |

Conclusion, Future Work, and Acknowledgements

HDR-GPT provides an accessible, end-to-end solution for HDR-based genome editing optimized for primary human cells and high-fidelity Cas9. By combining literature-informed experimental guidance with pathway-aware sgRNA selection and automated donor template design, the platform lowers technical barriers for non-expert users and reduces experimental iteration, thereby accelerating applications in disease modeling, mechanistic studies, and gene therapy development. Importantly, its performance is consistent across clinically relevant cell contexts (e.g., HSPCs and iPSCs), enabling more reliable guide prioritization without extensive empirical screening. Future work involves continuing to fine-tune the model by adding more experimental datapoints of various cell lines to ensure generalizability. A website will be created for ease of access as well. We would like to thank the senior design project instructional team for their continued guidance in the development process for our project.