# Correlation between edit length and Onyx™ Edit Fraction for substitution and insertion edit

The Onyx platform enables rapid generation of thousands of unique user-specified edits. The resulting Onyx Cell Library can be characterized using assays and metrics to determine the success of the genome engineering experiment. Onyx Edit Fraction is the fraction of the edited cells carrying a complete and intended genomic edit, as determined by the Onyx Edit Identification Assay [1]. Higher values for this metric indicate a greater fraction of cells in the population contain one of the desired edits in the library.

Several factors can influence the efficiency of the Onyx editing process, including the fitness effects of the conferred edits, genomic context, guide RNA and nuclease cutting efficiency, and repair efficiency. One of the key features that affect the homology directed repair (HDR) efficiency - and consequently the edit fraction - is the length of the edit design type, such as insertion or substitution. Here we show that length correlates inversely with edit efficiency for edits between 24 and 60 nucleotides.

To measure the correlation between the edit length and the Onyx Edit Fraction, we generated Onyx Design DNA libraries composed of substitutions and insertions, where genomic context and guide RNA were held constant and the edit lengths were varied. We constructed four substitution and four insertion libraries in the E. coli MG1655-like strain (INSC 1003). Each library contained 500 edits of the same length, targeting nonessential gene loci distributed over the genome, as well as two positive edit controls. The designs targeted the same locations in the genome so that only the length of the insertion and substitution edits varied between each design pair. We tested edit lengths of 24, 30, 45, and 60 nucleotides (Table 1). Similarly, we generated libraries with the same edit lengths for the S. cerevisiae S288C-like (INSC 1019) and S. cerevisiae CEN.PK-like (INSC 1020) strains, with 447 or 436 edit designs in each of the insertion or substitution libraries, respectively.

| Strain | Edit Type | Edit length | | | |
|---|---|---|---|---|---|
| | | 24 nt | 30 nt | 45 nt | 60 nt |
| E. coli MG1655 (INSC 1003) | Insertion | 500 | 500 | 500 | 500 |
| | Substitution | 500 | 500 | 500 | 500 |
| S. cerevisiae S288C (INSC 1019) | Insertion | 447 | 447 | 447 | 447 |
| | Substitution | 436 | 436 | 436 | 436 |
| S. cerevisiae CEN.PK (INSC 1020) | Insertion | 447 | 447 | 447 | 447 |
| | Substitution | 436 | 436 | 436 | 436 |

**Table 1:** Number of designs in each of the tested libraries. All E. coli libraries consisted of 500 designs; S. cerevisiae libraries contained 447 and 336 designs for insertions and substitution libraries, respectively.

When configuring the edit sequences, we aimed to sample ~80% of edits within the coding DNA sequence (CDS) regions and ~20% of edits upstream of CDS start positions and no more than 200 nucleotides upstream of the translation start site. For all edits that targeted regions within each CDS, the entire edit length was also contained within the CDS (no edits started upstream of the CDS and entered into the coding region and no edits started within the coding region and extended outside of the coding region). Positions within the CDS tend to favor the region near the 5' end. All substitution and insertion edit sequences within CDS regions started with the sequence "TAATTAATTAA" to knock-out the targeted gene. To achieve the targeted substitution or insertion edit length, the stop codon sequences were concatenated with a random nucleotide sequence that was constructed to mimic the GC content of the targeted microbe (~50% for E. coli and ~40% for S. cerevisiae). For substitution edits, random nucleotides were selected to ensure that every nucleotide in the target region would be modified (so as to prevent microhomology regions within the reference to edit sequence alignment). For regions upstream of CDSs, a random nucleotide sequence was constructed using the method described above.

The cell populations were generated on the Inscripta™ Onyx platform and Onyx Edit Fractions were determined using the Edit Identification Assay workflow. The effect of edit type and edit length on edit fraction was first assessed by fitting an ANOVA model for each strain separately, treating edit length and edit type as categorical factors. For models with a significant main effect (overall p-value $<0.05$), all pairwise comparisons between edit lengths and edit types were assessed using Tukey Contrasts, with an adjusted p-value of $<0.05$ considered strong evidence against the null hypothesis (no change). Linear regression was then used to assess the continuous effect of edit length on edit fraction by fitting separate models for each edit type and strain. In this model the intercept could be interpreted as the theoretical maximum edit rate and the slope as the edit fraction lost when increasing the edit length by one base.

Processing the edits by edit length and edit type showed that the Onyx Edit Fraction is strongly and negatively correlated with the edit length in a nearly linear trend (Figure 1). For S. cerevisiae S288C, the Onyx Edit Fraction decreases almost 2 times faster as a function of edit length than for E. coli MG1655. For S. cerevisiae CEN.PK, the Onyx Edit Fraction decreases with edit length even more substantially than for S288C. The steeper decrease in Edit Fraction with increasing edit length indicates the underlying biology of repair mixed with the physical limitations of joining DNAs.

We also observed that the impact of edit length varies between insertion and substitution edit types (Figure 2). In E. coli, insertions are generally less efficient than substitutions (insertion < substitution), while in S. cerevisiae substitutions are generally less efficient than substitutions (substitution < insertion). However, in all cases the edit length contributes more to editing efficiency than the edit type.

This data highlights the trade-offs between the length of the desired edit and expected Onyx Edit Fraction. This consideration is important when selecting the most biologically relevant edits: for example, choosing between a genome-wide insertion library of complete regulatory sequences (promoters and ribosome binding sites) versus insertion of promoter sequences alone. The included dataset is intended to help guide you in the design process to increase the probability of successful editing.

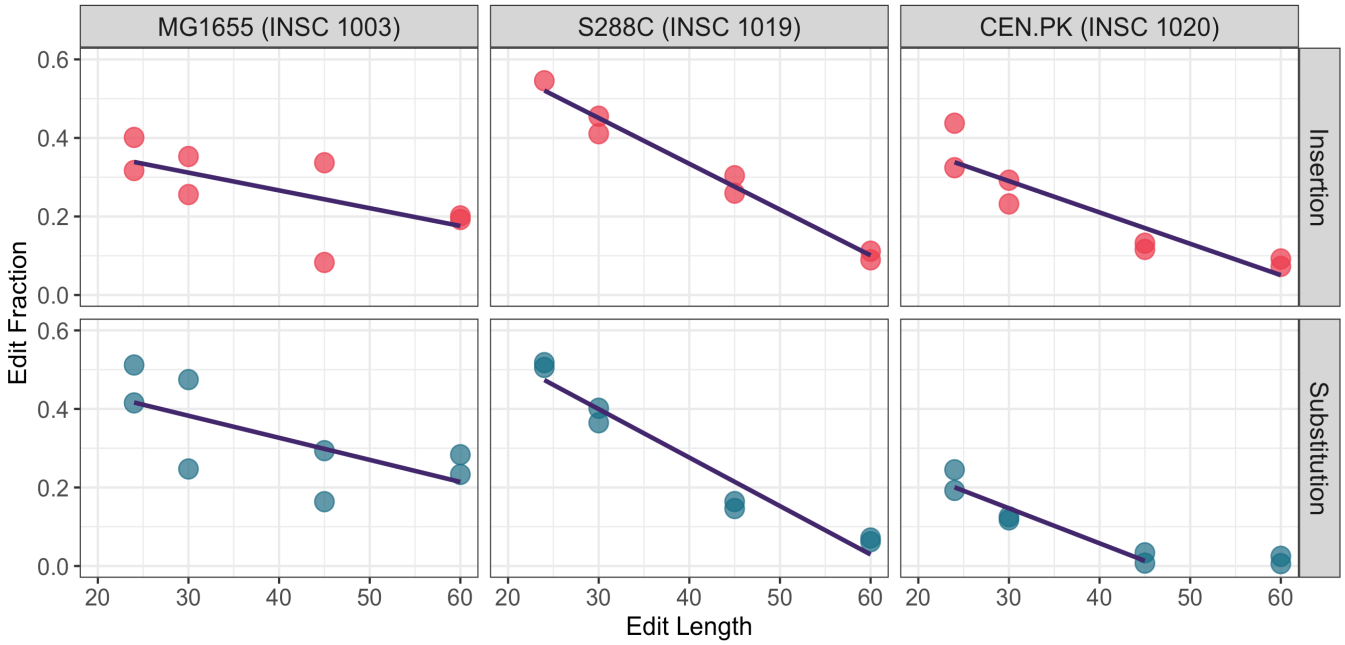


**Figure 1:** Onyx Edit Fraction correlates inversely with edit length. Edit fraction plotted as a function of edit length for E. coli MG1655 (INSC1003), S. cerevisiae S288C (INSC1019) and CEN.PK (INSC1020) strains. Each edit fraction estimate is derived from duplicate measurements of between 436 and 500 designs. Trend lines were calculated using simple linear regression.

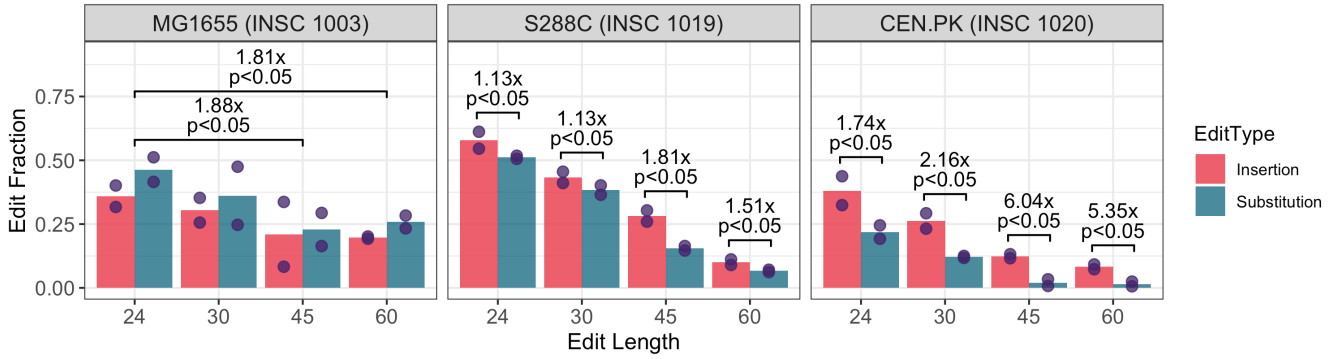


**Figure 2:** Onyx Edit Fraction is correlated with Edit Type. Bar charts showing the edit fraction as a function of edit type and edit length. Each edit fraction is derived from 436-500 designs. Significant differences are annotated by displaying the fold-change between groups and the calculated p-value.

## REFERENCES

1. Cawley, Simon, et al. "A framework for evaluating edited cell libraries created by massively parallel genome engineering." bioRxiv (2021).

Learn more at **INSCRIPTA.COM**