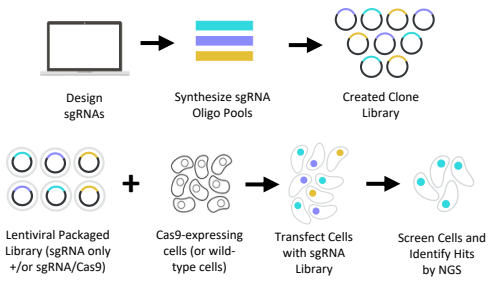
A Platform to Create High-Throughput Uniform Cloned Oligo Pools for CRISPR Screens

Elian Lee, Siyuan Chen, Rebecca L. Nugent

Abstract

CRISPR systems and their associated nucleases are endonucleases that target DNA cut sites through small guide RNAs (gRNAs). gRNAs contain a variable 20-nucleotide sequence for targeting specific DNA cut sites (crRNA), and a longer scaffold sequence is required for nuclease binding (tracrRNA). For convenience, single-guide RNAs (sgRNAs) link the crRNA and tracrRNA into a chimeric molecule. CRISPR screens use libraries of sgRNA sequences delivered in vectors as DNA for *in vivo* transcription to gRNAs. Often the most challenging part of a screen is not its characterization, but the cloning of the oligo pool into a vector under an inducible promoter. The problems are two-fold: first is the lack of high-quality oligo pools and the second is bias in oligo sequence representation after amplification, cloning, and growth. We highlight here the ability of Twist Bioscience's multiplexed oligo pool libraries to maintain uniformity throughout amplification and cloning. We measure cloning method efficiency by the number of colony-forming units (CFUs) and measure sequence distributions and uniformity after amplification and cloning using next-generation sequencing (NGS). We demonstrate the correction of method-specific cloning biases with longer inserts for increased sequence uniformity in the final cloned pool.

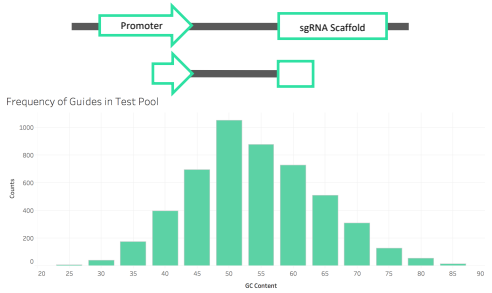
Pooled Lentiviral CRISPR Libraries



Preparation and screening of CRISPR libraries
The sgRNAs in a CRISPR library are designed and optimized *in silico* for high specificity and low off-target binding before being synthesized as oligonucleotides and cloned as a pool into lentiviral vectors. CRISPR libraries typically contain 4-6 unique sgRNAs per gene to ensure adequate modification of target genes. CRISPR libraries are also amplified and verified by NGS to ensure uniform representation before being packaged into lentivirus. Finally, the cells are transduced with the lentiviral vector and positive or negative selection is applied to identify relevant genes (hits) with NGS.

Results

Design

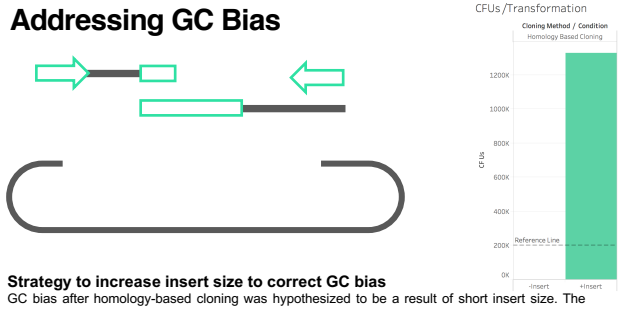


Experimental pool of sgRNA protospacer sequences
A pool of 5,000 randomly generated 20-bp protospacer sequences with a variety of GC content was selected for cloning into a lentiviral vector between a promoter and sgRNA scaffold. Designs were made to test the cloning method options:
- Restriction digest (*Bsm*BI) and ligation
- Homology-based cloning and ligation-independent cloning (LIC; homologous ends)

Cloning

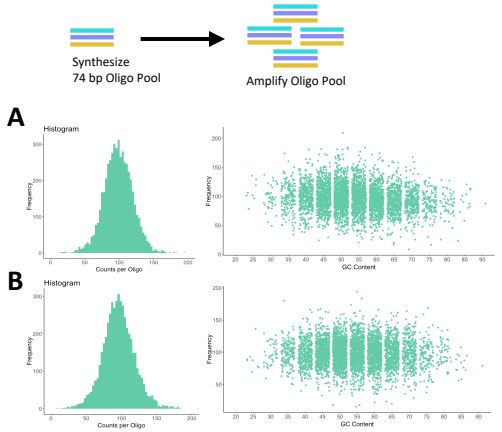
Homology-based cloning yields high CFUs with low background
E. coli colony-forming units (CFUs) from three different cloning methods with and without insert. CFUs ensure there's appropriate sgRNA protospacer representation in the pool. 20x coverage ensures proper representation of a cloned pool. Restriction enzyme digestion and ligation produced high CFUs but also yielded ~7% background colonies.

Addressing GC Bias



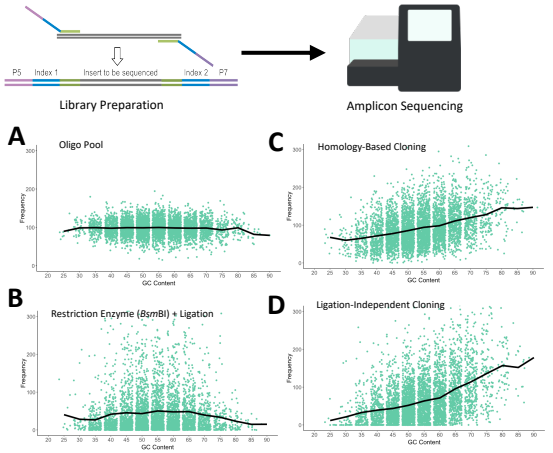
Strategy to increase insert size to correct GC bias
GC bias after homology-based cloning was hypothesized to be a result of short insert size. The cloning strategy was adjusted to incorporate a longer insert into the vector. A part of the vector backbone was amplified and stitched to the oligo pool via overlap-extension PCR. Re-cloning the extended oligo pool also significantly increased CFUs/transformation as compared with previous cloning attempts.

Amplification



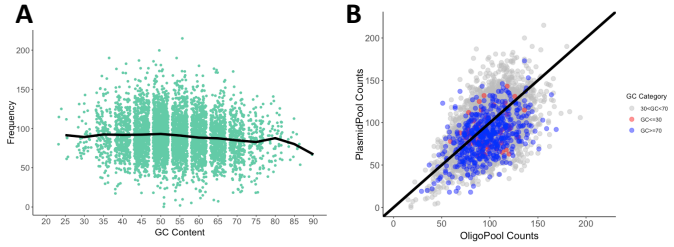
Oligo Pool is well-represented and uniform after amplification
74-bp Twist Oligo Pools were amplified for six PCR cycles and prepared for NGS. FASTQ reads were trimmed by Cutadapt and aligned with BWA. Graph shows the distribution of the sgRNA protospacer read counts and read counts per GC bin. **(A)** Restriction digest and ligation strategy; **(B)** homology-based and ligation-independent cloning strategy.

Sequencing



Pool representation is GC biased after cloning
Cloned oligo pools were plated onto QTrays for amplification for 13 hours, then scraped off to harvest plasmids. Plasmids underwent NGS library preparation for amplicon sequencing. FASTQ reads were trimmed by Cutadapt and aligned with BWA. Graphs show observed sgRNA protospacer frequency normalized to 100,000 reads against GC content.

Sequencing After Cloning Extended Insert



Plasmid pool is GC uniform with adjusted cloning strategy
Extended Insert was cloned and sequenced by NGS. sgRNA protospacer read counts were compared back to original oligo pool. **(A)** sgRNA protospacer counts per GC bin after cloning **(B)** sgRNA protospacer counts scatter plot before and after cloning.

Summary

Twist Oligo Pools exhibit tight sequence uniformity after DNA synthesis. This uniformity is maintained when pools are amplified by PCR with a high-fidelity DNA polymerase and minimal amplification cycles. Twist Bioscience developed a high-throughput platform to clone oligo pools. This platform lacks sequence or restriction enzyme restrictions, and uniformity is retained from synthesis to post cloning.

CRISPR screens enable whole genome-scale interrogation of gene functions. A uniform CRISPR pool improves the extent to which every sgRNA in a pooled library is present in a screen, thereby reducing bias and increasing reproducibility across replicates.