Optimizing CRISPR-based lineage tracing tools to study metastasis

Harish Bayana[1], Kamen Simeonov[2], Christopher Lengner[2] - [1]CAS 2024, University of Pennsylvania [2]Department of Cell & Molecular Biology, University of Pennsylvania

INTRODUCTION

Cancer metastasis is the spread of cancer cells from the original tumor site to a distant site in the body, resulting in a disease that is significantly more complicated and difficult to treat. Despite most localized cancers seeing significant improvements in survival rates over the past few decades, metastasis remains a largely incurable disease, presenting with five year survival rates of well below 20%. Despite accounting for such a massive proportion of cancer related deaths, metastasis remains an extremely poorly understood phenomenon. Tracing the lineage of metastatic cells has the potential to provide a robust means of studying metastasis, allowing for the identification of specific cells that metastasize along with providing the transcriptional identity of those cells.

macsGESTALT: CRISPR/Cas9 Lineage Tracing Tool

Tracing the lineage of metastasis requires an extremely robust tool that allows for cells to be uniquely marked and identified at the single cell level. Most established methods currently used to study cancer, such as tracking somatic mutations across the genome, fail to mark cells at a sufficient level of resolution. CRISPR/Cas technologies provide the framework for a system that overcomes these limitations. By using an version of the Genome Editing of Synthetic Target Arrays for Lineage Tracing (GESTALT) method adapted to mammalian systems, known as multiplexed, activatable, clonal and subclonal GESTALT (macsGESTALT), we are able to efficiently track metastatic processes in vivo. This system consists of three components: an inducible Cas9 complex, an array consisting of a series of gRNAs, and a genetic barcode. The barcode is a 250 bp sequence composed of a series of five gRNA target sites each separated by 3 bp protospacer adjacent motif (PAM) sites. Upon induction of the system via doxycycline, the Cas9 is successfully transcribed and can bind with the constitutively expressed gRNAs. This complex is then able to recognize and bind to the target barcode, inducing either insertions or deletions. As cells divide, this modified version of the barcode is passed on to their progeny, in which this mutagenesis can continue. The barcode is expressed in the 3' UTR of a puromycin N-acetyl-transferase (pac) transcript, allowing it to be sequenced via single cell RNA sequencing.

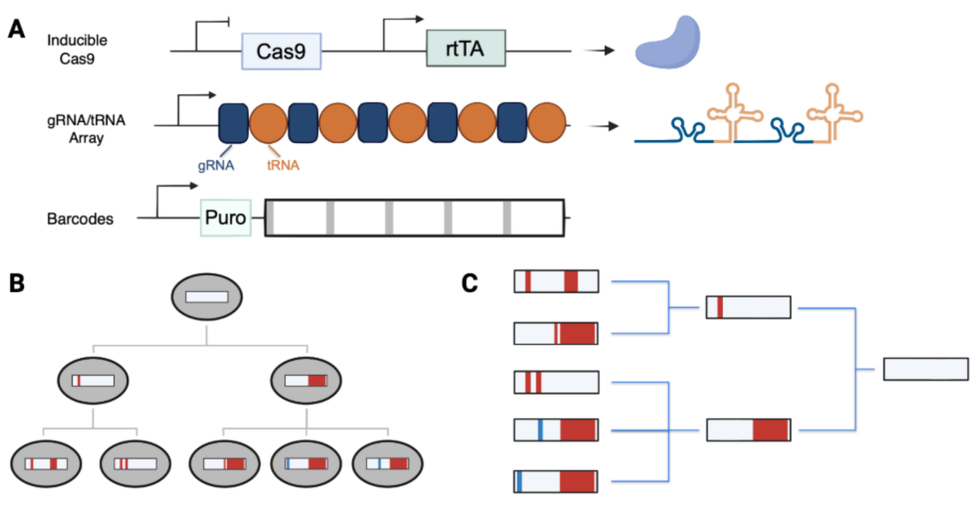

Figure 1: macsGESTALT overview
(A) Three genetic components macsGESTALT: inducible Cas9, gRNA/tRNA array, and genetic barcode
(B) Example of barcode editing over three generations, starting from a single cell with a single barcode integration. Deletions are represented in red and insertions are represented in blue
(C) Phylogenetic reconstruction of example barcode data from (B)

gRNA/tRNA array expression limited by premature termination of RNA polymerase

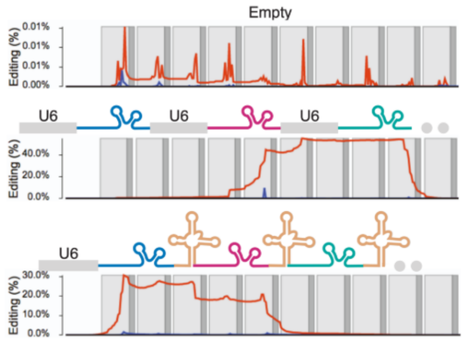

Figure 2: Expression rates of various gRNA constructs

The optimization of gRNA expression is one of the forefront concerns for the application of this system, as the diversity of the barcodes produced is directly dependent on the levels of gRNAs expressed. In a previous experiment, various structures for the gRNA construct were tested, with the highest expression resulting from a structure in which the gRNAs are alternating in an array with tRNAs, as shown in the figure to the left. While this testing was done in a robust 293T cell line via transient transfection, the cancer model relies on lentiviral infection into PDAC cells, which occurs at a significantly lower rate. Hence, gRNA expression cannot be elevated to the same levels by simply increasing the copy number. One of the major limiting factors of gRNA expression in this construct is premature termination of the RNA polymerase during transcription. A previous study found that the termination efficiency of RNA polymerase is dependent on the number of thymine residues in series, with a 6T residue being the standard terminator region. Each of the gRNA sequences has a 4T sequence, corresponding to a termination efficiency of around 75%. Since the gRNAs are originally produced in a single transcript, this effect is compounded, resulting in the fifth gRNA only being produced 0.098% of the time, assuming exponential compounding.

Optimizing gRNA expression via A-T flip

This effect would explain the limited editing rates that have been measured thus far. Hence, the focus of this project has been to optimize the gRNA structure to avoid this unwanted termination. Switching the A and T residues was previously shown to have no effect on the function of the gRNA. By introducing this switch into the gRNA structure, the premature termination efficiency would drop from 75% (4T) to 0% (3T), which would result in a rather significant increase in expression levels. Three different cell lines have been produced to test this hypothesis. In all three, the gRNAs target GFP as a proxy for their expression levels. The first cell line has two gRNAs with the original gRNA structure, the first with a non-relevant target site and the second one targeting GFP, separated by tRNAs. The second cell line has the same structure, but both of the gRNAs now have this A-T switch introduced. The third cell line, which serves as a positive control for expression levels, is a single gRNA under its own promoter targeting GFP. Data has not yet been collected, but an example dataset is shown below. Comparing the differences in GFP expression between the original and the modified gRNA structures will demonstrate the efficacy of this solution.

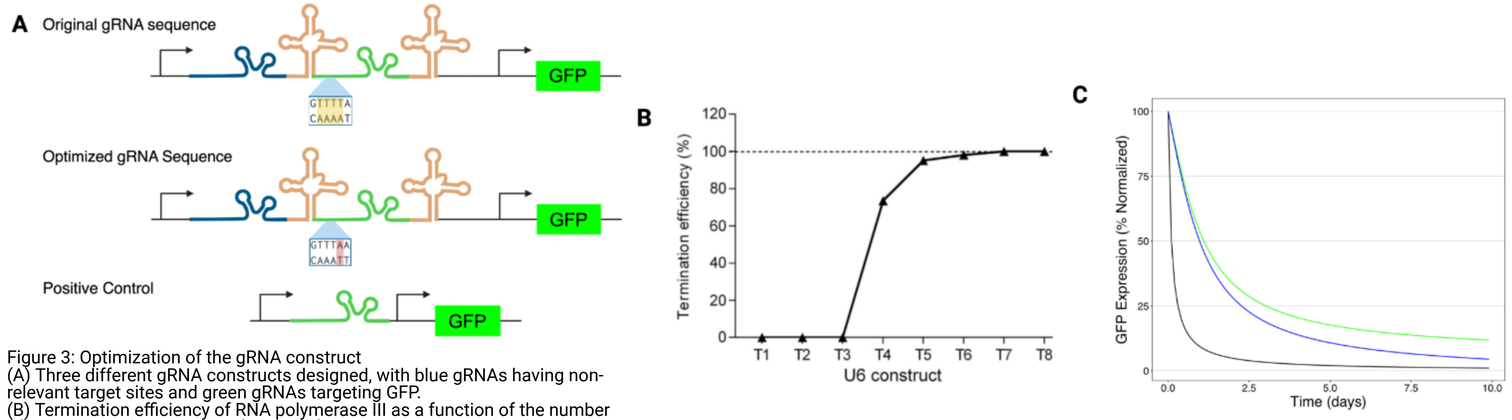

Figure 3: Optimization of the gRNA construct
(A) Three different gRNA constructs designed, with blue gRNAs having non-relevant target sites and green gRNAs targeting GFP.
(B) Termination efficiency of RNA polymerase III as a function of the number of tyrosine residues in series (adapted from Gao, Z., et al., Molecular Therapy Nucleic Acids (2018)).
(C) Example data for GFP expression rates for the three different constructs (original gRNA: green, optimized gRNA: blue, positive control: black)

ACKNOWLEDGEMENTS

Roy and Diana Vagelos Scholars Program in the Molecular Life Sciences, University of Pennsylvania Center for Undergraduate Research and Fellowships: College Alumni Society Undergraduate Research Grant