Report Title

Final Project: Statistics C245F

Florica Constantine, Lingrong Jin, Andy Shen

1 Introduction

The discovery of the Cas9 clustered regularly interspaced short palindromic repeats (CRISPR) enzyme by Ran et al. (2013) has revolutionized the world of genomics and bioinformatics. The ability to perform genome editing and cleaving opens the door to an endless array of biological research that explores the effect of Cas9.

One method is the Model-based Analysis of Genome-wide CRISPR/Cas9 Knockout (MAGeCK) proposed by Li et al. (2014). This technique utilizes the unique ability of Cas9 by identifying both positively and negatively-regulated genes in a robust and powerful manner. Here we utilize the output of MAGeCK on a CRISPR assay for children with Severe combined immunodeficiency (SCID). This rare disorder directly inhibits T-cells from being created, significantly altering a child's immune system. Such a disorder is typically fatal[1].

However, the advent of CRISPR assay sequencing technologies enables domain experts to investigate the genomic origins of SCID and discover genes that play a role in immune system development. In this process, a guide RNA (sgRNA) enters the genome with the intention to target a gene. A dissipating guide indicates that the sgRNA, and hence its target gene, was important for the specific cell-line. These discoveries point out the potential usefulness of certain genes in T-cell differentiation.

2 Data Organization

Our data is presented as a series of counts per gene. These counts correspond to the number of cells observed per gene. These cells have markers which indicate the presence of T-cells. A double-negative count indicates the presence of a cell with no T-cell markers. A single-positive count indicates one T-cell marker, while a double-positive count indicates the presence of two T-cell markers, the clearest evidence of T-cell detection.

While each sgRNA has a fixed efficiency, we are not provided that information, so this analysis assumes that the efficiency is fixed across all sgRNAs.

References

Li, Wei, Han Xu, Tengfei Xiao, Le Cong, Michael I Love, Feng Zhang, Rafael A Irizarry, Jun S Liu, Myles Brown, and X Shirley Liu. 2014. "MAGeCK Enables Robust Identification of Essential Genes from Genome-Scale CRISPR/Cas9 Knockout Screens." *Genome Biology* 15 (12): 1–12.

Ran, FAFA, Patrick D Hsu, Jason Wright, Vineeta Agarwala, David A Scott, and Feng Zhang. 2013. "Genome Engineering Using the CRISPR-Cas9 System." *Nature Protocols* 8 (11): 2281–2308.

[1] https://www.niaid.nih.gov/diseases-conditions/severe-combined-immunodeficiency-scid