EVALUATING PREDICTORS OF SpCas9 GUIDE RNA PERFORMANCE

Mark Dunne[1], Neil Humphryes-Kirilov[1], Søren Hough[1], **Riley Doyle**[1]

[1]Desktop Genetics, London, UK || Contact: **crispr@deskgen.com**

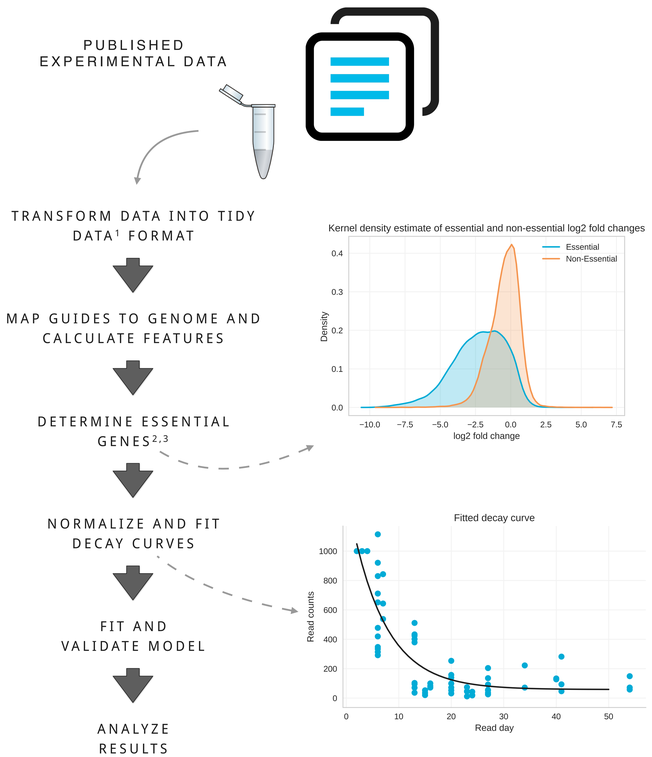
Learn more

WWW.DESKGEN.COM

1. INTRODUCTION

- Understanding the features which dictate CRISPR guide behavior remains an outstanding problem in the field of genome editing
- We combined data from multiple sources and built a novel CRISPR guide scoring function, **Dunne 2017**
- We compared **Dunne 2017** to others using Spearman's Correlation Coefficient to test the quality of the ranking, and Coefficient of Determination to test the accuracy of the predicted log2 fold change
- With a better understanding of what factors affect the performance of a guide, we can design smaller but more effective libraries, ultimately reducing costs

2. ANALYSIS OVERVIEW

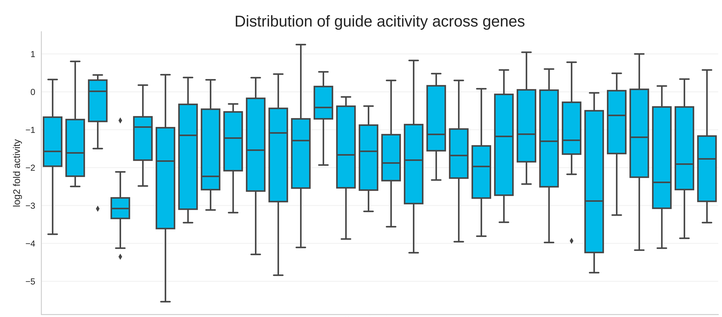


3. SEQUENCE FEATURES

- Our model found different weights for nucleotide identity and position as compared with older algorithms
- Guide performance is determined primarily by nucleotides in the seed region of the protospacer (position 15–20)
- G in the seed region is a positive predictor of guide performance
- T in the seed region is a negative predictor of guide performance

4. GENOME-WIDE TARGETING

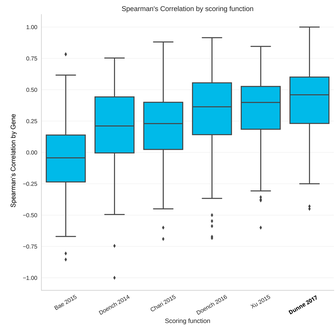
- Some genes require fewer guides to knock out, while others require more
- More guides are required when there are fewer high-quality guides to choose from
- Ranking guides within a gene is therefore insufficient; a small number of high performance guides can do the work of many low performers



5. COMPARING SCORING FUNCTIONS

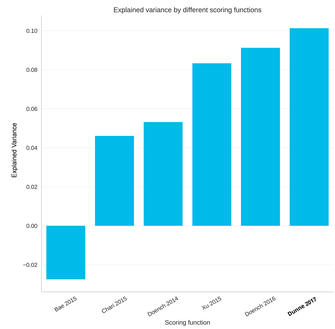
SPEARMAN'S CORRELATION COEFFICIENT

- The quality of guide ranking within genes is not the only important metric
- Spearman's measures how well the features predict guide performance based on *ranked* guide activity
- Predicting *actual* guide activity is key to designing more efficient libraries



COEFFICIENT OF DETERMINATION

- Coefficient of Determination explains how much of the variance can be explained by the model
- With better prediction of *actual* guide activity, we can reduce the minimum necessary number of guides for successful gene knockout



6. CONCLUSIONS

- By incorporating more features in our algorithm, we can build a better model (**Dunne 2017**) for predicting the behavior of guide RNAs
- Ranking guides within a gene is insufficient; a small number of high performance guides can do the work of many low performers
- It is important that scoring functions can predict the absolute activity of a guide
- **Dunne 2017** outperforms older algorithms measured both by Spearman's Correlation Coefficient and Coefficient of Determination

7. REFERENCES

1. Wickham, Hadley. "Tidy data." Journal of Statistical Software 59.10 (2014): 1-23.
2. Wang, Tim, et al. "Genetic screens in human cells using the CRISPR-Cas9 system." Science 343.6166 (2014): 80-84.
3. Hart, Traver, et al. "Measuring error rates in genomic perturbation screens: gold standards for human functional genomics." Molecular systems biology 10.7 (2014): 733.
4. Doench, John G., et al. "Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation." Nature biotechnology 32.12 (2014): 1262-1267.
5. Doench, John G., et al. "Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9." *Nature biotechnology* 34.2 (2016): 184.
6. Xu, Han, et al. "Sequence determinants of improved CRISPR sgRNA design." Genome research 25.8 (2015): 1147-1157.
7. Kim, Daesik, et al. "Digenome-seq: genome-wide profiling of CRISPR-Cas9 off-target effects in human cells." Nature methods 12.3 (2015): 237-243.
8. Chari, Raj, et al. "Unraveling CRISPR-Cas9 genome engineering parameters via a library-on-library approach." Nature methods 12.9 (2015): 823-826.