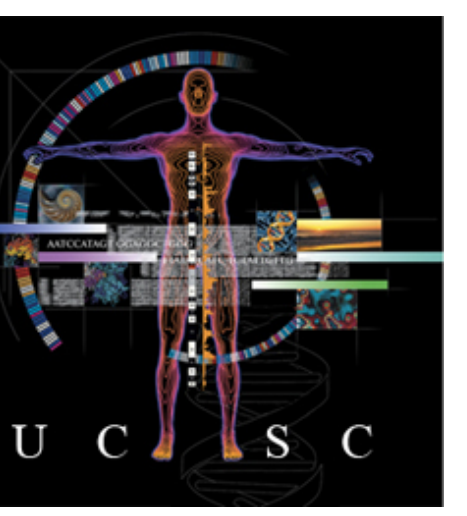


# Evaluation of off-target and on-target scoring algorithms: Integration into crispor.org and the UCSC Genome Browser



Maximilian Haeussler<sup>1</sup>, Kai Schoenig<sup>2</sup>, Helene Eckert<sup>3</sup>, Alexis Eschstruth<sup>4</sup>, Joeffrey Mianne<sup>5</sup>, Jean-Baptiste Renaud<sup>6</sup>, Sylvie Schneider-Manouory<sup>4</sup>, Alena Shkumatava<sup>3</sup>, Lydia Teboul<sup>5</sup>, Jim Kent<sup>1</sup>, Jean-Stephane Joly<sup>6</sup>, Jean-Paul Concordet<sup>7</sup>  
<sup>1</sup>University of California Santa Cruz Genomics Institute <sup>2</sup>Medical Faculty Mannheim, Heidelberg <sup>3</sup>Institut Curie, Paris <sup>4</sup>Sorbonne University, Paris <sup>5</sup>MRC Harwell, UK <sup>6</sup>CNRS Gif-sur-Yvette, France <sup>7</sup>Museum d'Histoire Naturelle, Paris

## The optimal guide?

We all want Crispr/SpCas9 guides with:

- High specificity = few off-targets in genome
- High efficiency = high on-target cleavage rate

Many studies have identified off-targets genome-wide and quantified on-target cleavage. Some have presented algorithms that predict the "best" guides.

## Published off-target data

- Hsu et al. Nat Biot 2013 targeted PCR
- Cho et al. Gen Res 2014 targeted PCR
- Frock et al. Nat Biot 2015 Translocation sequencing
- Tsai et al. Nat Biot 2015 GuideSeq
- Kim et al. Nat Meth 2015 DiGenome-Seq + PCR
- Wang et al. Nat Biot 2015 Lentiviral integration site seq.
- Ran et al. Nat 2015 BLESS
- Kim et al. Gen Res. 2016 DiGenome-Seq2 + PCR
- > 650 off-targets and cleavage frequencies for 31 guides

## Published on-target data

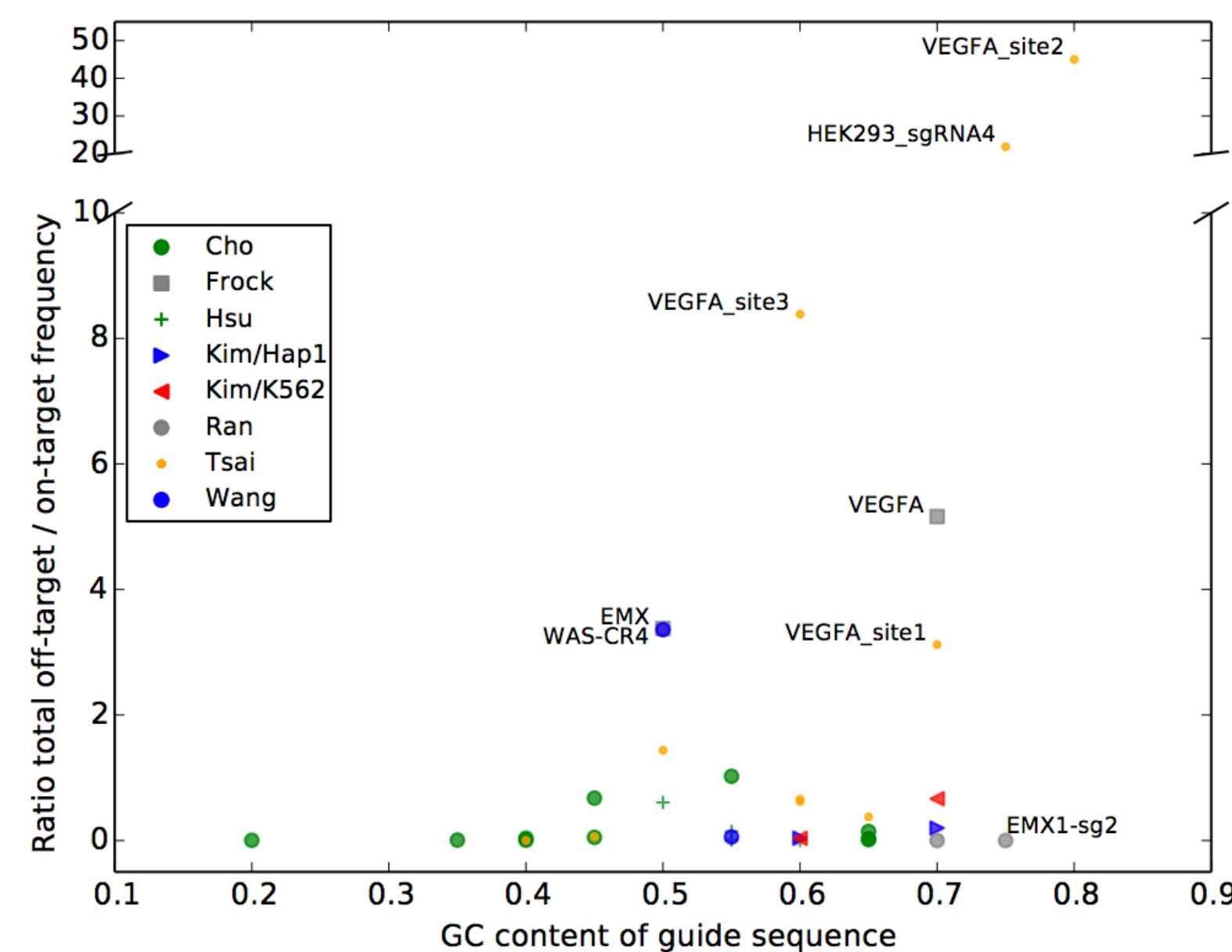
- Wang et al, Science 2014, 2077 guides, human cells
- Gagnon et al, Plos One 2014, 118 guides, zebrafish
- Doench et al, Nat Biot. 2014, 881 guides, human cells,
- Ren et al, Cell Rep 2014, 39 guides, fruitfly embryos
- Chari et al, Nat Meth 2015, 1235 guides, human cells
- Farboud et al, Genetics 2015, 50 guides, *C. elegans*
- Moreno-Mateos et al, Nat Meth 2015, 1021 guides, zebrafish
- Hart et al, Cell 2015, 8276 guides, human cells
- Gandhi et al, BioRxiv 2016, 72 guides, *C. intestinalis*
- >19,000 on-target cleavage frequencies

? **What are the best off-target prediction algorithms?**

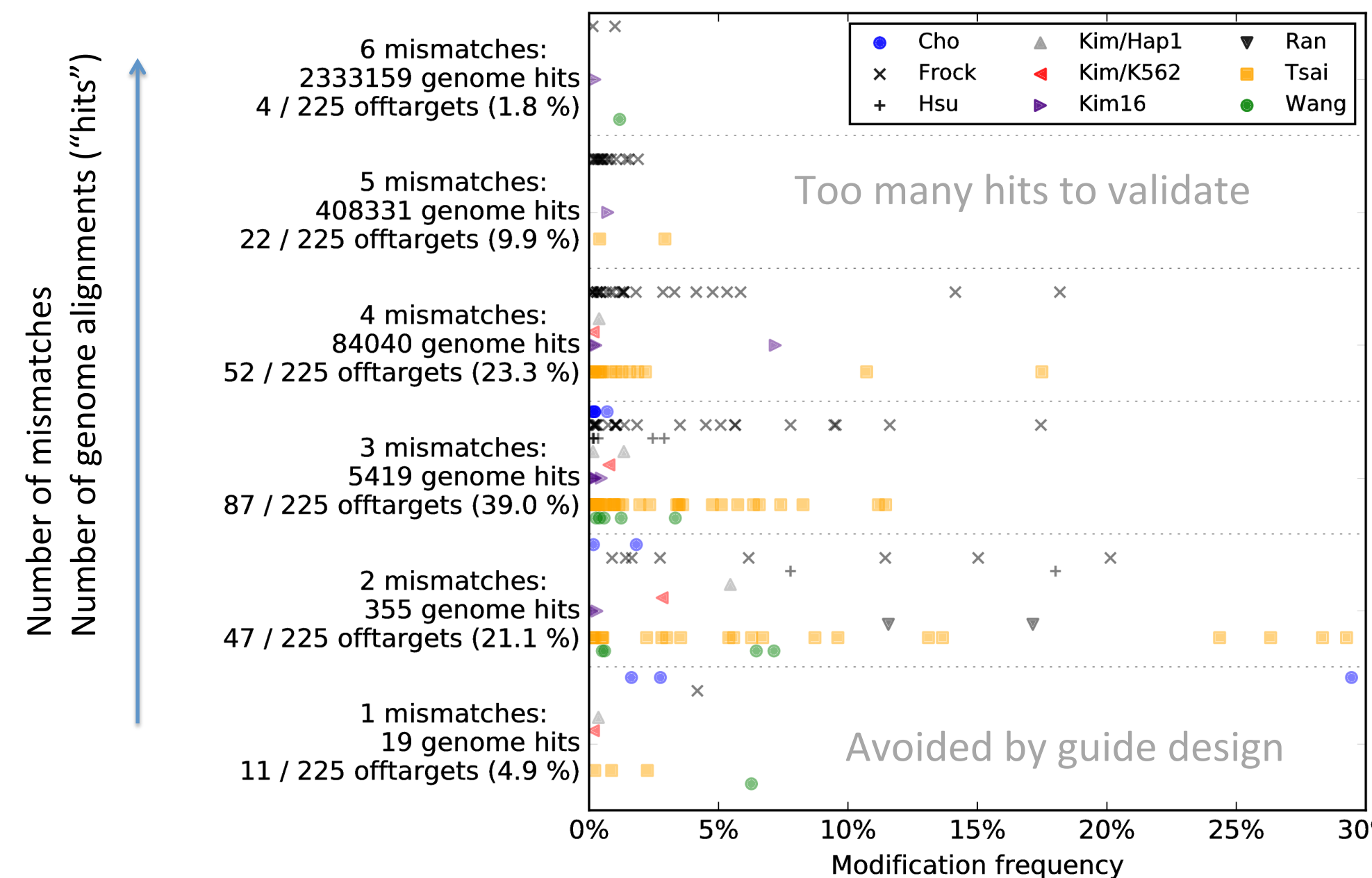
**What are the most appropriate on-target algorithms, do they save time in practice?**

**Are the different studies consistent?**

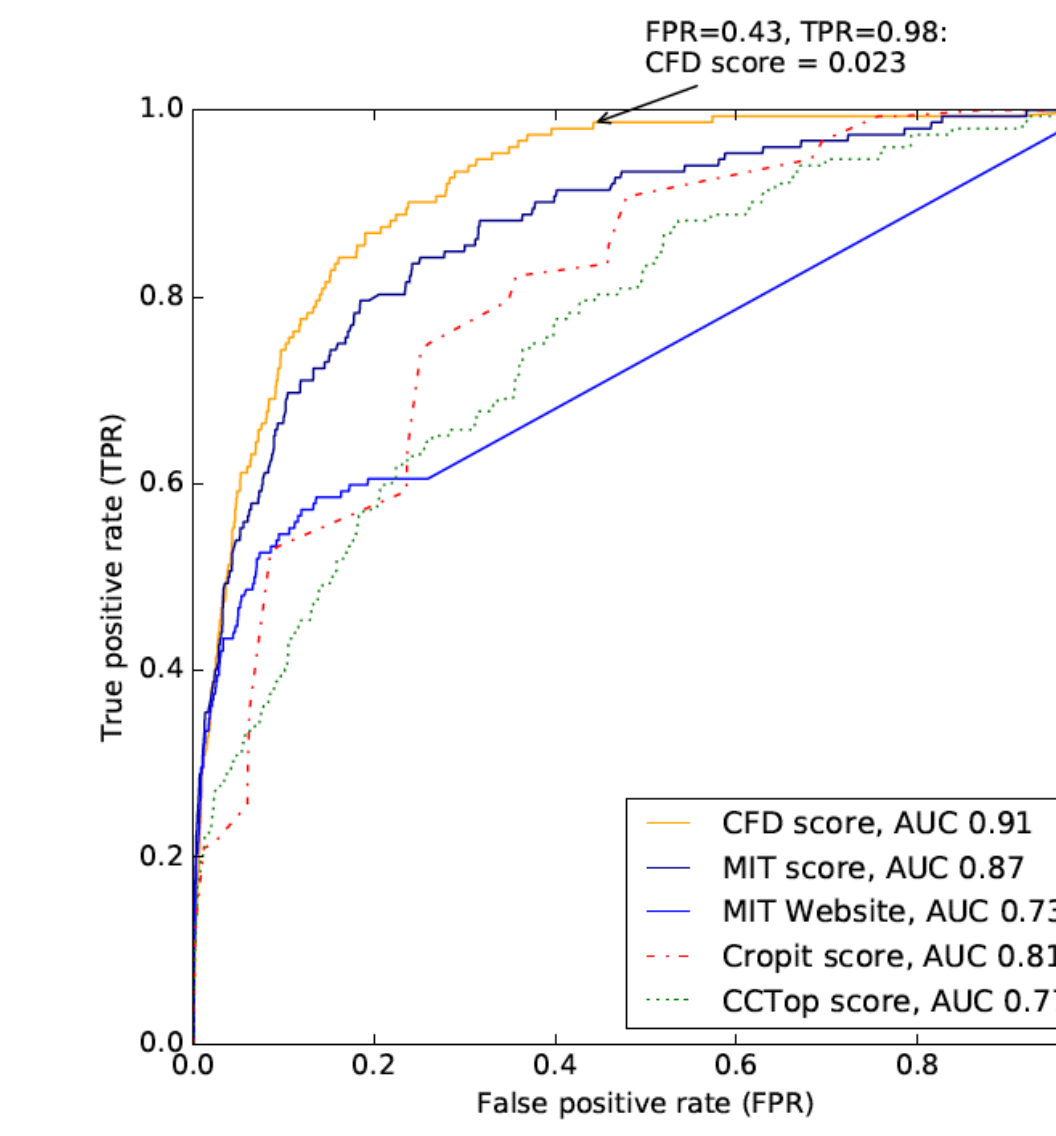
## Off-target counts relatively consistent, except two Tsai guides



## Off-targets: Allow up to four mismatches when searching



## Use the CFD Off-target score to reduce the screen, avoid crispr.mit.edu

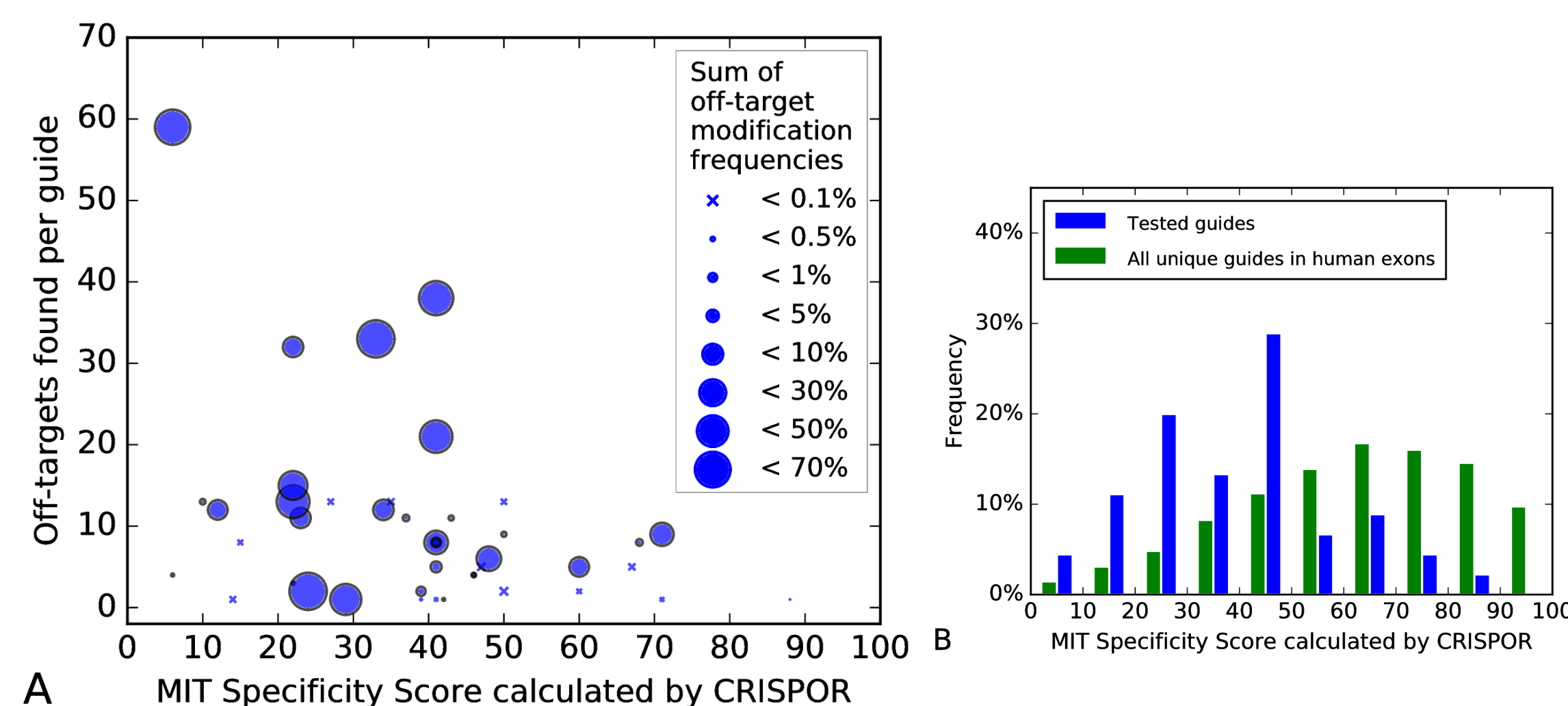


ROC plot

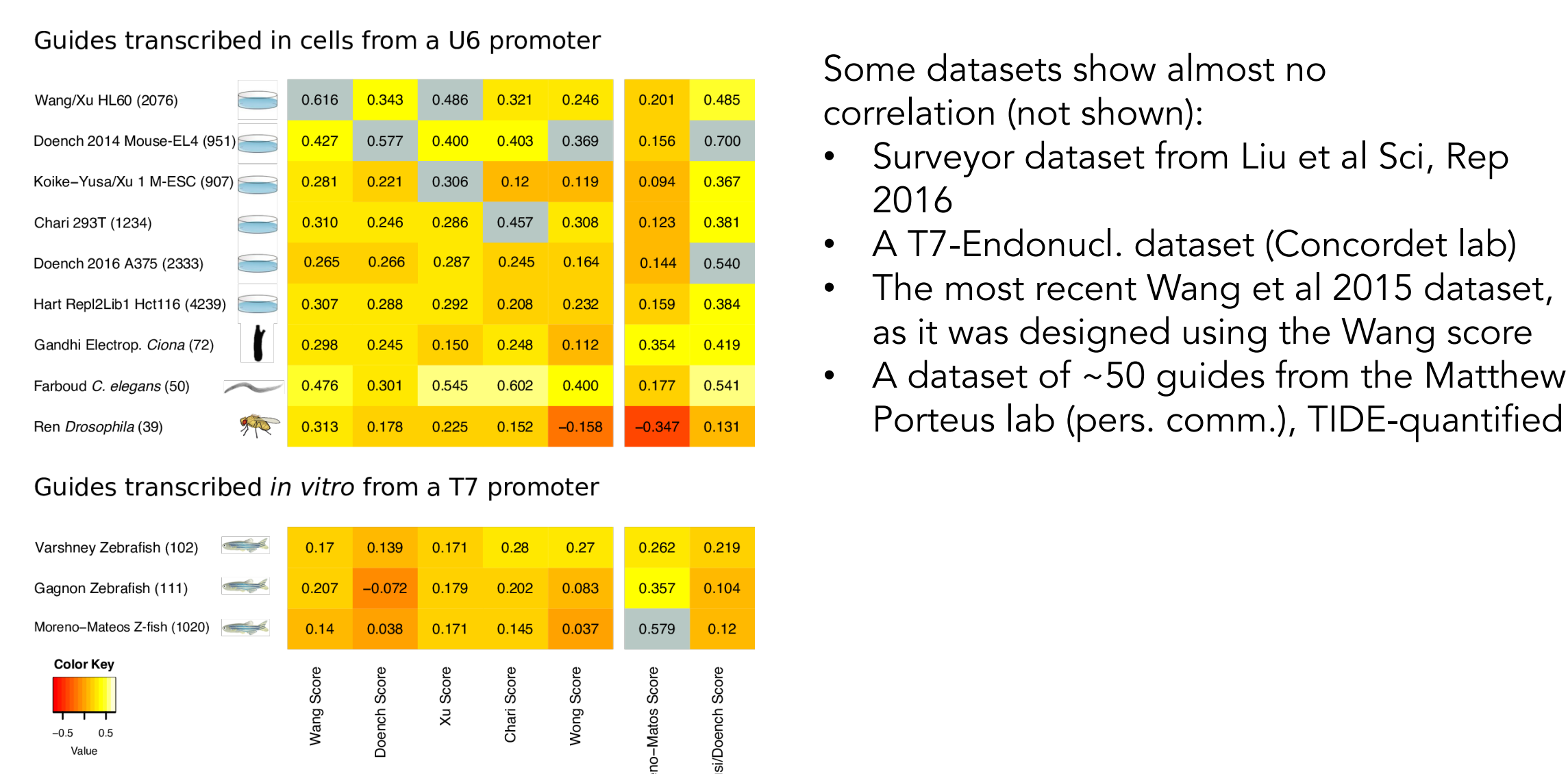
- Instances: 25,332 potential off-targets for 26 guides with  $\leq 4$  mismatches, Crispr or CasOffFinder
- True positives: 152 off-targets with a modification frequency  $> 0.1\%$
- True negatives: all other off-targets

Very low CFD scores: unlikely to be detectable off-targets

## MIT Guide Specificity score should be $> 50$ , but studies favor $< 50$



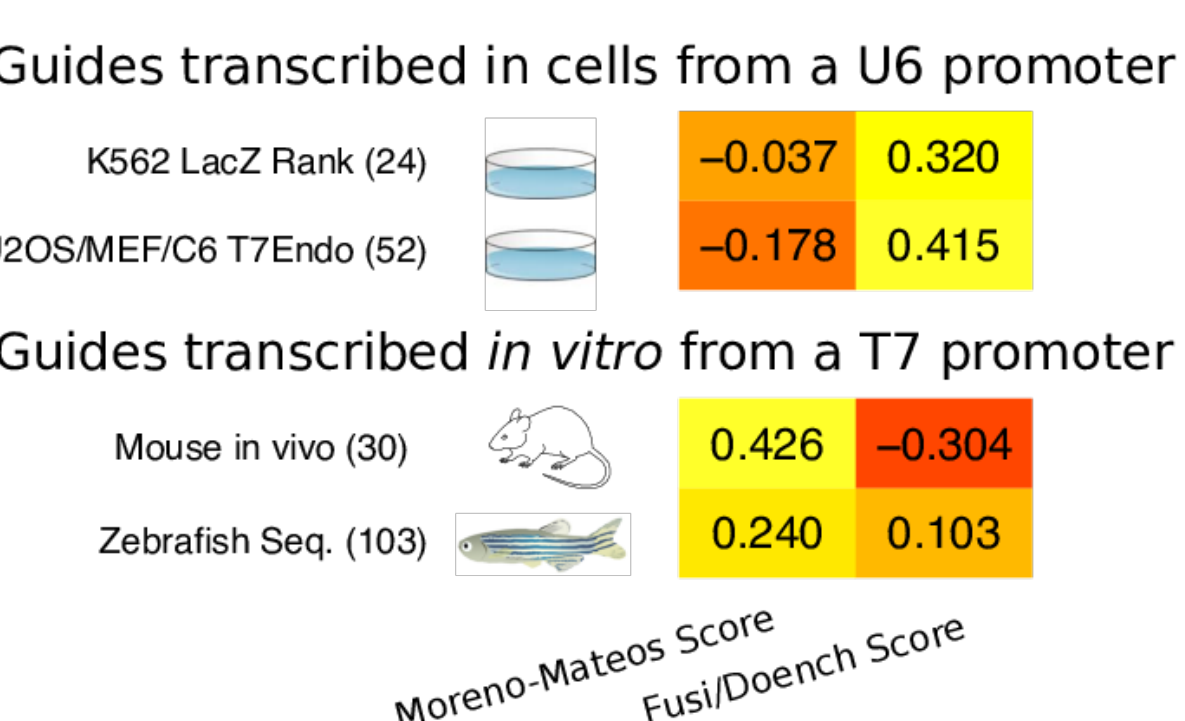
## On-target rates mostly consistent But *in-vitro* guides are different.



Some datasets show almost no correlation (not shown):

- Surveyor dataset from Liu et al Sci, Rep 2016
- A T7-Endonuc. dataset (Concordet lab)
- The most recent Wang et al 2015 dataset, as it was designed using the Wang score
- A dataset of ~50 guides from the Matthew Porteus lab (pers. comm.), TIDE-quantified

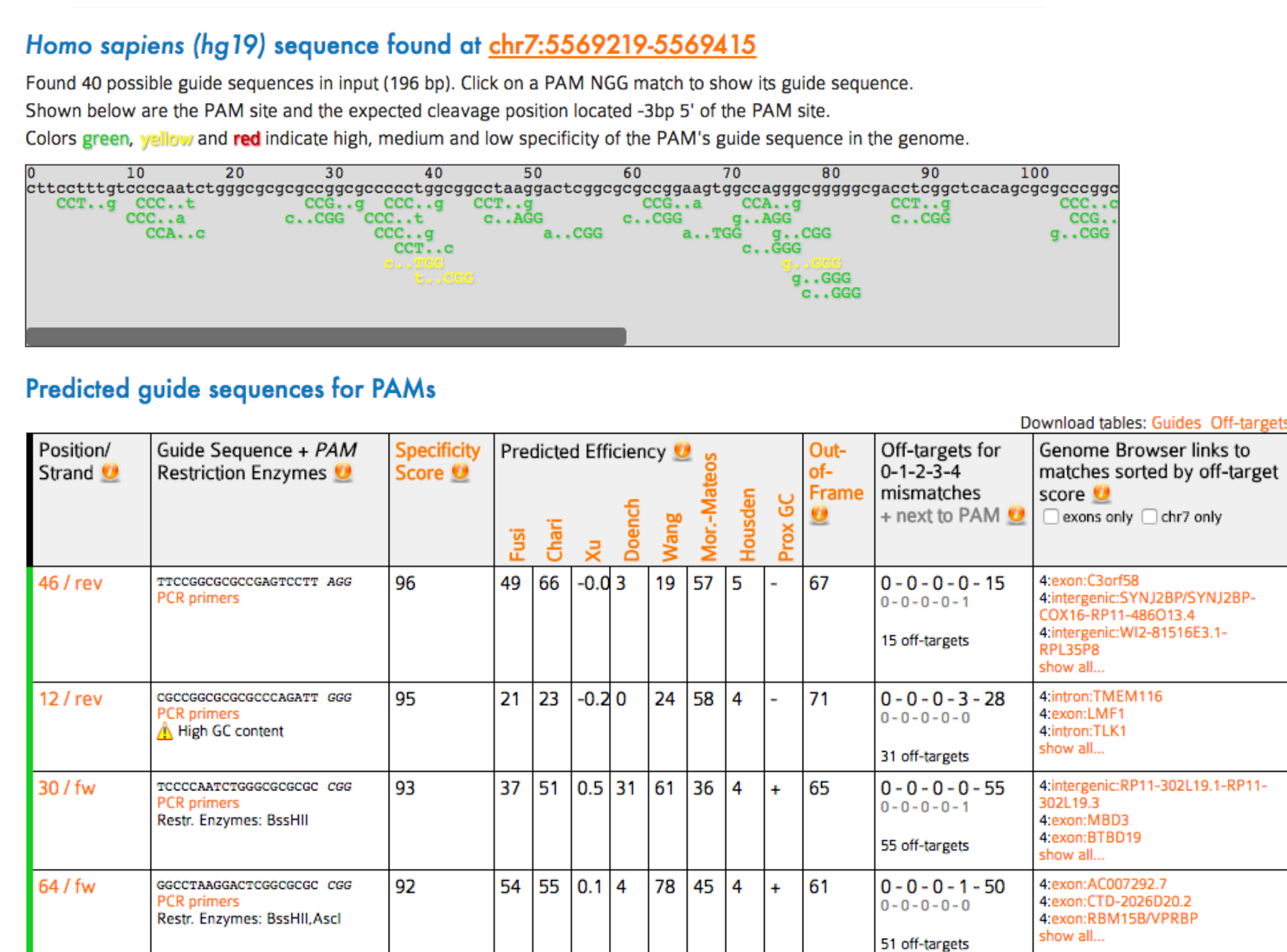
## Algorithms can reduce mouse/zebrafish screening work by ~1/3



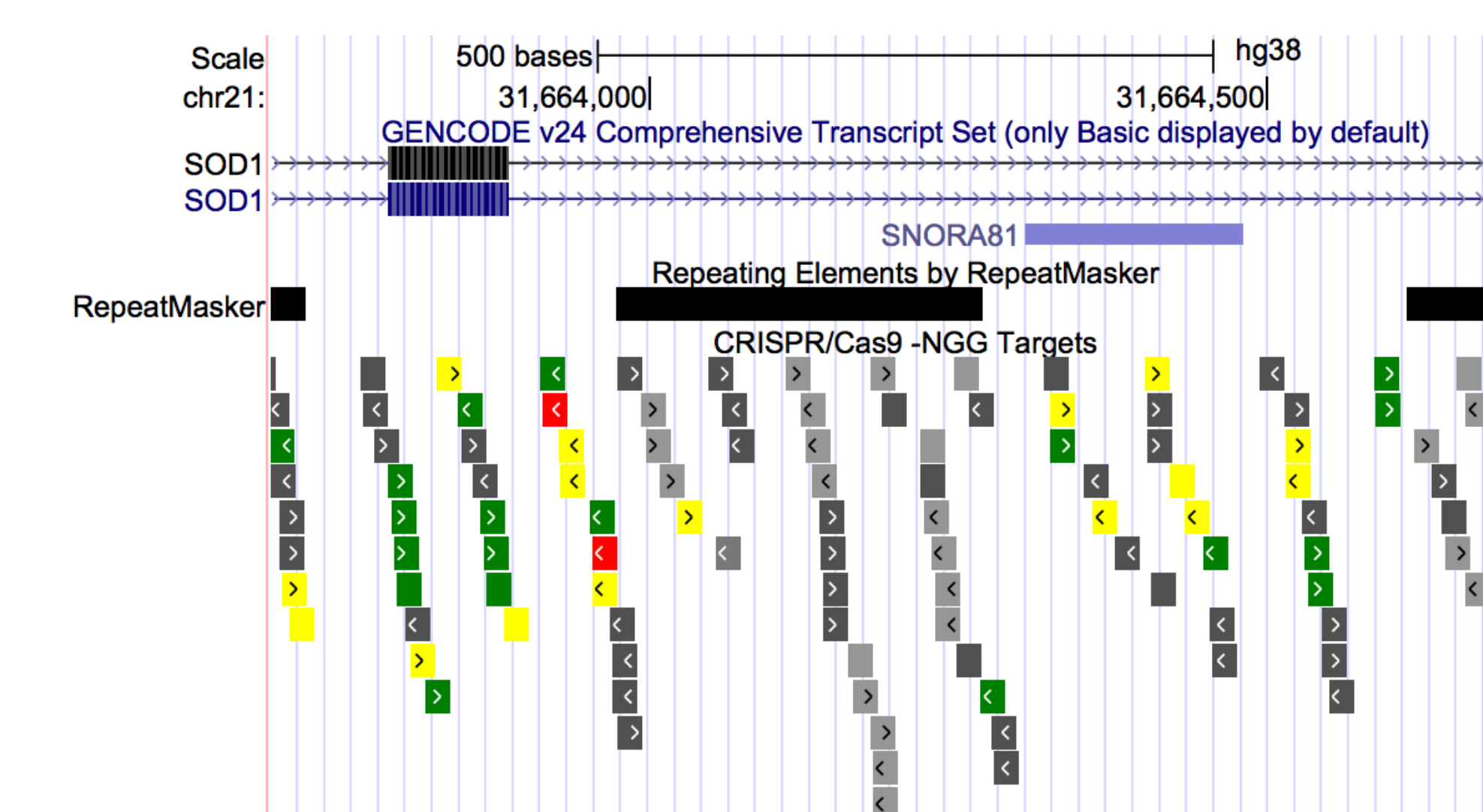
- New efficiency data from our labs:
- 24 guides in K562 cells
- 52 guides in U2OS cells
- 104 guides in Zebrafish embryos
- 31 guides in mouse embryos

Post-hoc: We could have saved one third of guide screens, hundreds of hours, by designing guides with the Moreno-Mateos score in Zebrafish/Mice or the Doench 2016 score in cells

## Crispor.org: 127 genomes, all scores, fast runtime



## Genome-test.soe.ucsc.edu: pre-calculated scores for six organisms



- Gray: specificity score  $< 50$
- Green/Yellow/Red: Doench 2016 score 100-55/55-30/30-0
- Moreno-Mateos scores currently only on mouse-over

## More information

- Webtool: [crispor.org](http://crispor.org)
- Article (Genome Biol. 2016) [bit.ly/crisporPaper](http://bit.ly/crisporPaper)
- Source code: [github.com/maximilianh/crisporWebsite](https://github.com/maximilianh/crisporWebsite)
- UCSC tracks: [genome-test.soe.ucsc.edu](http://genome-test.soe.ucsc.edu)

## Future work

- More coloring options on the browser track
- Other genomes?
- Cfp1 support - no need for scores?
- Faster off-target calculations - Bowtie vs. BWA

## Acknowledgements:

This work was funded by National Human Genome Research Institute (5 U41 HG002371-15 to UCSC Center for Genomic Science). We would like to acknowledge the work of the UCSC Genome Bioinformatics technical staff (<http://genome.ucsc.edu/staff.html>), our collaborators, and all authors that provided Crispr datasets for their support.

