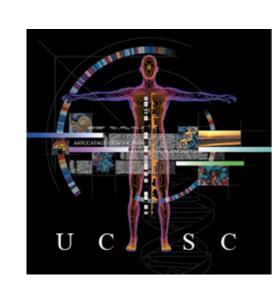
Evaluation of off-target and on-target scoring algorithms: Integration into crispor.org and the UCSC Genome Browser



Maximilian Haeussler¹, Kai Schoenig², Helene Eckert³, Alexis Eschstruth⁴, Joeffrey Mianne⁵, Jean-Baptiste Renaud⁶, Sylvie Schneider-Manoury⁴, Alena Shkumatava³, Lydia Teboul⁵, Jim Kent¹, Jean-Stephane Joly⁶, Jean-Paul Concordet⁷ ¹University of California Santa Cruz Genomics Institute ²Medical Faculty Mannheim, Heidelberg ³Institut Curie, Paris ⁵ MRC Harwell, UK ⁶ CNRS Gif-sur-Yvette, France ⁷ 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 cleaveage. 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 cleaveage frequencies for 31 guides

Published on-target data

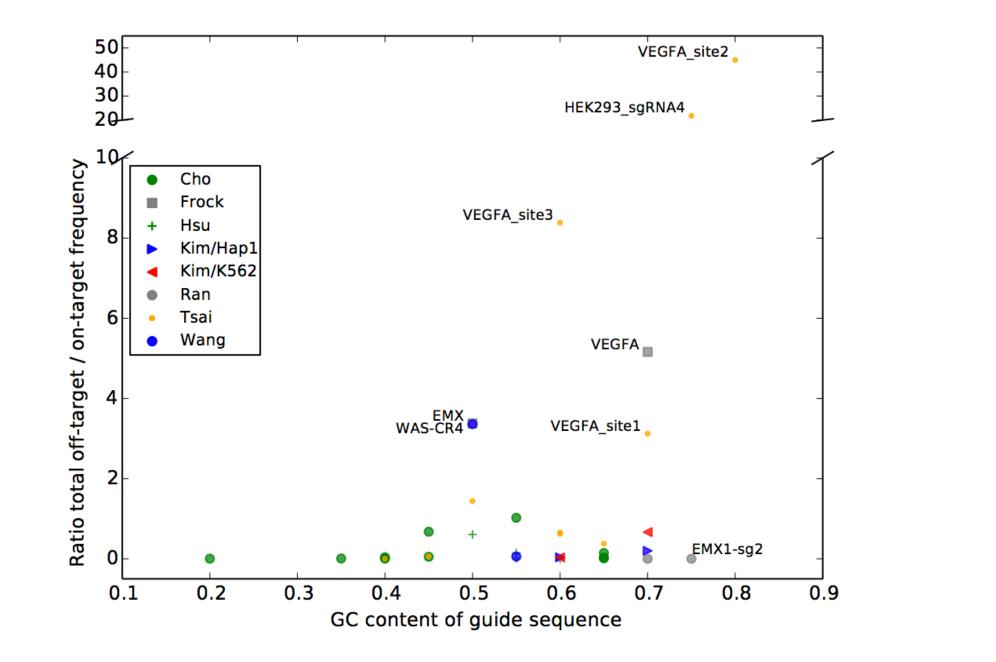
- Wang et al, Science 2014, 2077 guides, human cells
- Gagnon et al, Plos One 2014, 118 guides, zebrafish
- Doench et al, Nat Biot. 2014, 881 guides, human cells,
- Ren et al, Cell Rep 2014, 39 guides, fruitfly embryos
- Chari et al, Nat Meth 2015, 1235 guides, human cells
- Farboud et al, Genetics 2015, 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. intestinali* >19.000 on-target cleavage frequencies

What are the best off-target

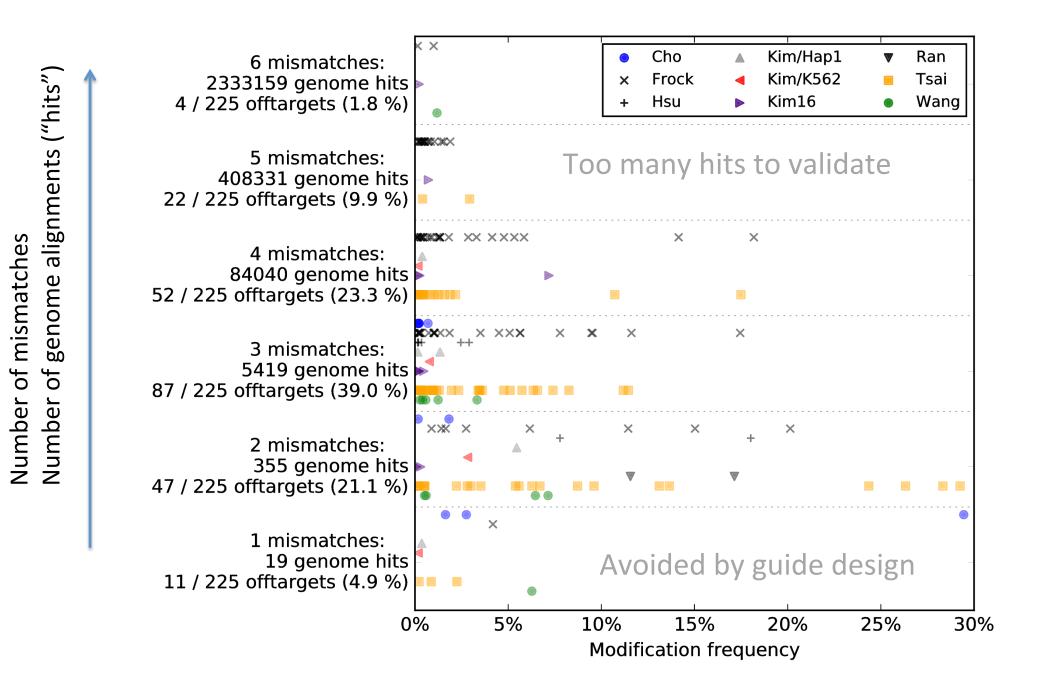
What are the most appropriate on-target algorithms, do they save time in practice?

Are the different studies consistent?

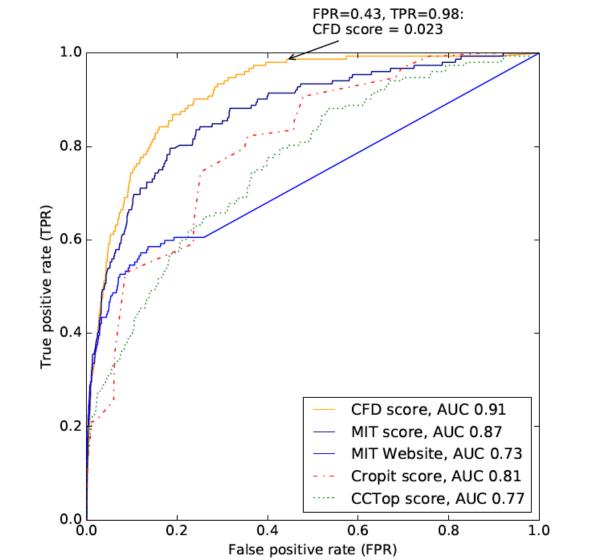
Off-target counts relatively consistent, except two Tsai guides



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

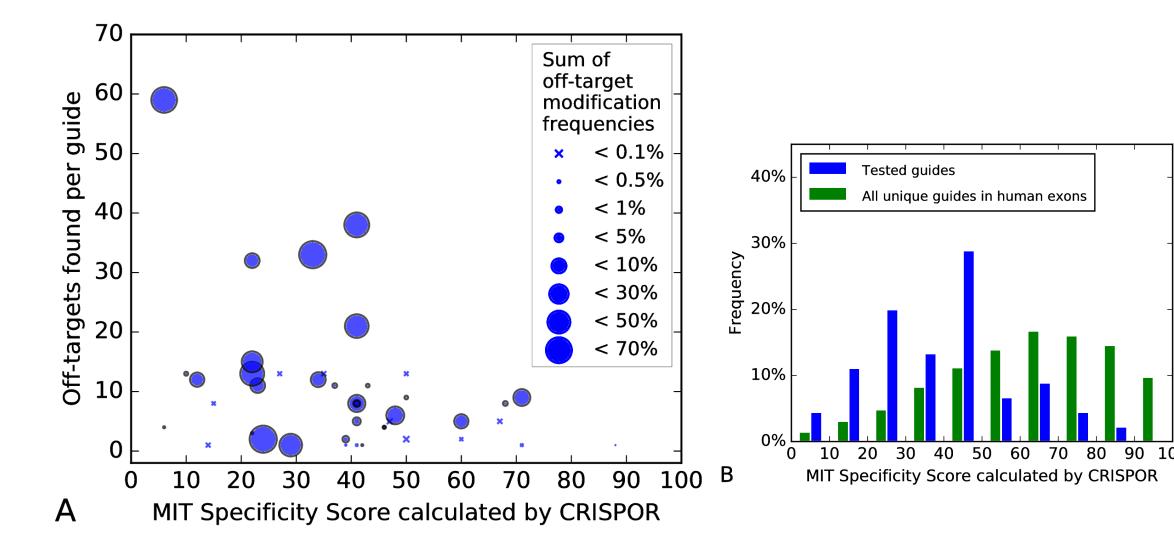
- Instances: 25,332 potential offtargets for 26 guides with <= 4 mismatches, Crispor or CasOffFinder
- True positives: 152 off-targets with a modification frequency > 0.1%
- True negatives: all other off-targets

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

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

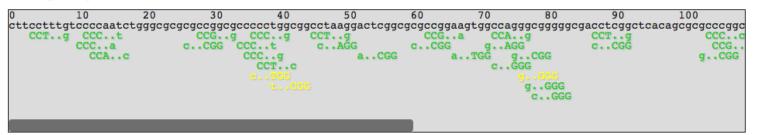
On-target rates mostly consistent But *in-vitro* guides are different.

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



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

Homo sapiens (hg19) sequence found at chr7:5569219-556941



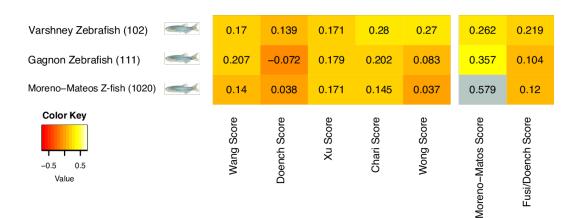
Predicted guide sequences for PAMs

Position/ Strand 🤨	Guide Sequence + PAM Restriction Enzymes 9	Specificity Score 🧕	Pre	etcie Chari	d Eff	icien Doench	cy 🙎	MorMateos	Housden	Prox GC	Out- of- Frame	Off-targets for 0-1-2-3-4 mismatches + next to PAM <u>U</u>	Genome Browser links to matches sorted by off-target score 9 exons only chr7 only
46 / rev	TTCCGGCGCGCCGAGTCCTT AGG PCR primers	96	49	66	-0.0	3	19	57	5	-	67	0 - 0 - 0 - 0 - 15 0 - 0 - 0 - 0 - 1 15 off-targets	4:exon:C3orf58 4:intergenic:SYNJ2BP/SYNJ2BP- COX16-RP11-486O13.4 4:intergenic:WI2-81516E3.1- RPL35P8 show all
12 / rev	CGCCGGCGCGCGCCCAGATT GGG PCR primers A High GC content	95	21	23	-0.2	0	24	58	4	-	71	0 - 0 - 0 - 3 - 28 0 - 0 - 0 - 0 - 0 31 off-targets	4:intron:TMEM116 4:exon:LMF1 4:intron:TLK1 show all
30 / fw	TCCCCAATCTGGGGGGGGGGC CGG PCR primers Restr. Enzymes: BssHII	93	37	51	0.5	31	61	36	4	+	65	0 - 0 - 0 - 0 - 55 0 - 0 - 0 - 0 - 1 55 off-targets	4:intergenic:RP11-302L19.1-RP11- 302L19.3 4:exon:MBD3 4:exon:BTBD19 show all
64 / fw	GGCCTAAGGACTCGGCGCGC CGG PCR primers Restr. Enzymes: BssHII,AscI	92	54	55	0.1	4	78	45	4	+	61	0 - 0 - 0 - 1 - 50 0 - 0 - 0 - 0 - 0 51 off-targets	4:exon:AC007292.7 4:exon:CTD-2026D20.2 4:exon:RBM15B/VPRBP show all

Guides transcribed in cells from a U6 promoter

Wang/Xu HL60 (2076)	0.616	0.343	0.486	0.321	0.246	0.201	0.485
Doench 2014 Mouse-EL4 (951)	0.427	0.577	0.400	0.403	0.369	0.156	0.700
Koike-Yusa/Xu 1 M-ESC (907)	0.281	0.221	0.306	0.12	0.119	0.094	0.367
Chari 293T (1234)	0.310	0.246	0.286	0.457	0.308	0.123	0.381
Doench 2016 A375 (2333)	0.265	0.266	0.287	0.245	0.164	0.144	0.540
Hart Repl2Lib1 Hct116 (4239)	0.307	0.288	0.292	0.208	0.232	0.159	0.384
Gandhi Electrop. <i>Ciona</i> (72)	0.298	0.245	0.150	0.248	0.112	0.354	0.419
Farboud <i>C. elegans</i> (50)	0.476	0.301	0.545	0.602	0.400	0.177	0.541
Ren Drosophila (39)	0.313	0.178	0.225	0.152	-0.158	-0.347	0.131

Guides transcribed in vitro from a T7 promoter



Some datasets show almost no correlation (not shown):

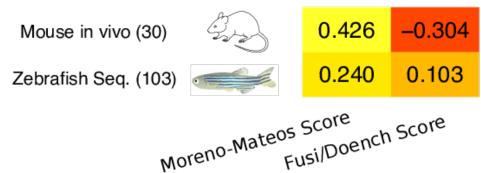
• Surveyor dataset from Liu et al Sci, Rep 2016

• A T7-Endonucl. 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

Guides transcribed in cells from a U6 promoter

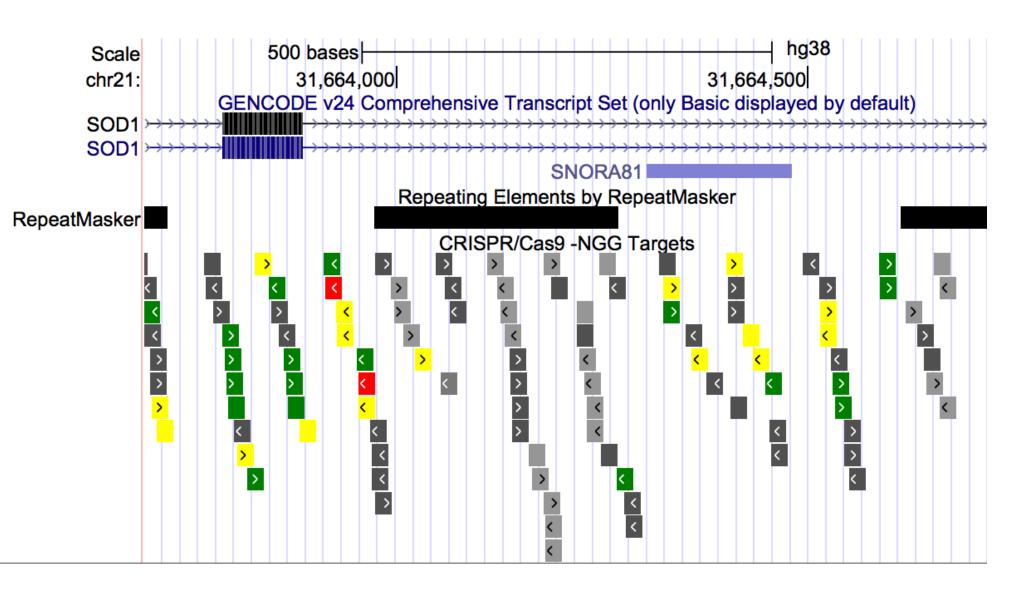
K562 LacZ Rank (24)	-0.037	0.320
U2OS/MEF/C6 T7Endo (52)	-0.178	0.415

Guides transcribed in vitro from a T7 promoter



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

Genome-test.soe.ucsc.edu: precalculated scores for six organisms



More information

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

Future work

- More coloring options on the browser track
- Other genomes?

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

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

• 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.