Generation of large genomic deletions to remove zebrafish *rca2.1*

Chon-Hwa Tsai-Morris[1], G. Margolin[2], R. Dale[2], D. Chauss[3], C. Kemper[4], B. Afzali[3] and B. Feldman[1],
[1]NICHD Zebrafish Core; [2]NICHD Bioinformatic and Scientific Programming core, [3]NIDDK, [4]NHLBI .
Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD), NIH, Bethesda MD.

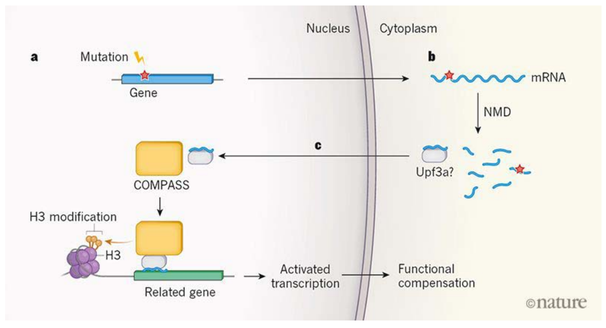
Abstract

CD-46 plays a crucial role in the human immune system. We wish to illuminate the role of zebrafish *rca2.1*, a CD-46 ortholog, in development and health. The sequence and expression of *rca2.1* suggests it is a better orthologue to human CD-46 than murine candidate genes. Located on chromosome 23, *rca2.1* spans 25 kb and the latest genome assembly predicts three alternative splice forms. Our initial goal is to generate *rca2.1* null zebrafish using CRISPR/Cas9. Based on concerns that novel translational start sites and/or decay of otherwise non-functional RNAs can elicit genetic compensation, we sought to disrupt all *rca2.1* transcription. To identify the transcriptional start site (TSS) and to determine which predicted alternate transcripts are expressed during early development, we established a private UCSC track featuring valuable CAGE and RNAseq data from public sources. We used this track to select gRNA combinations aimed at cooperatively deleting the TSS target and downstream spans of *rca2.1*. Each candidate gRNA was pre-tested for cutting efficiency. Fertilized embryos were then co-injected with either (#1) three gRNAs targeting the TSS, start codon and termination codon, or (#2) two gRNAs targeting the TSS and termination codon. With the intention of stimulating precise joining of the deletion gap, a 100 nt oligonucleotide was included in the second co-injection, with hybrid sequence from upstream of the 5' cut site and downstream of the 3' cut site. Fragment analysis indicated the presence of diagnostic peaks arising from the desired 20 Kb deletion in close to 20% of F_0 embryos injected by either approach. Germ-line transmission from 7 out of 8 F_0 adults representing both approaches has subsequently been achieved. Thus, using multiple gRNAs with or without a facilitating oligo, we have efficiently recovered large genomic *rca2.1* deletion alleles. Unlike classic CRISPR/Cas9 in-del alleles, interpretation of phenotypes observed should be uncomplicated by concerns of extant or novel alternative transcripts or genetic compensation linked to nonsense-mediated decay.

Rationale

Create transcript-less allele to avoid potential mutation-induced genetic compensation.

Genetic compensation triggered by mutant mRNA degradation



★:Premature termination codon

NMD: Nonsense-mediated RNA decay

Upf3a: Regulatory factors

COMPASS: Protein complex [Setd1a/b (or MII1/2)-wdr5-Ash2-Rbbp5]

Wilkinson, M.F., Nature, 2019,568:179
Summary model based on
1.Ma et al, Nature, 2019, 568: 259
2.El-Brolosy et al, Nature, 2019, 568:193

Goal

Assess our ability to create large genomic deletions with CRISPR-Cas9, which would be useful to avoid genetic compensation and for other applications

Background

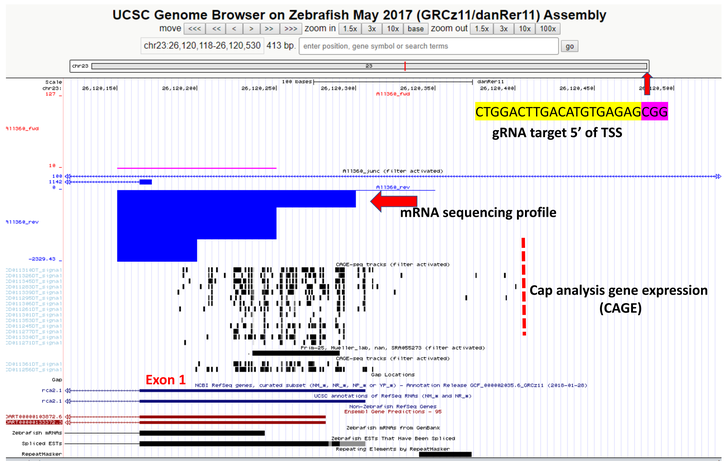
- Human CD-46 encodes a membrane-bound complement regulatory protein with multiple functions in both immune and non-immune systems.
- Zebrafish *rca2.1* is the closest zebrafish orthologue of human CD-46. It spans 25 kb (13 exons) on chromosome 23 and three splice forms are reported.

Strategy and Approach

- Used our standard CRISPR-Cas9 tools, including commercial Cas9 and gRNAs
- Designed gRNAs against targets upstream of the transcriptional start site (TSS), and downstream of the initiation codon (ATG) or termination codon (TAA).
- Established UCSC tracks featuring CAGE and RNAseq data from public sources to add confidence to our target selection
- Pre-screened gRNAs and selected only those with highest cutting efficiency in embryos, using fluorescent PCR fragment analysis
- Designed an oligonucleotide with 100 nts (50 up- and 50 downstream) flanking the predicted cut sites, as a potential contributor to precise deletion repair
- Injected embryos with all 3 gRNAs or 2 gRNAs plus oligonucleotide template.
- F_0 and F_1 screening by fluorescent PCR fragment analysis and further validation by sequence analysis.

Results

UCSC tracks featuring CAGE and RNA seq data to aid design of gRNAs targeting 5' of the TSS



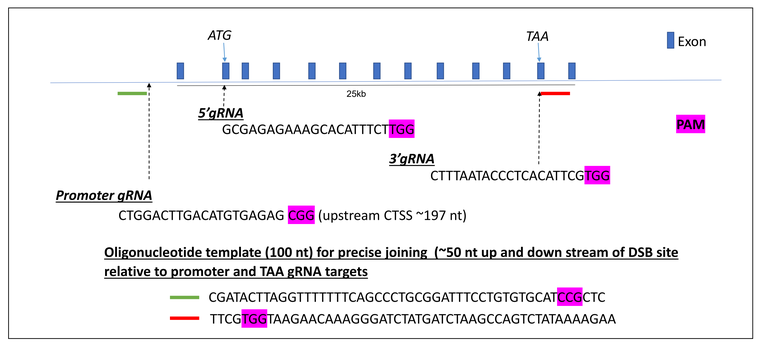
Summary of # embryos (F_0/F_1) with expected amplicons

Fo screening	Expt Design				F_0 screening for 5' flanking oligo joins to 3' flanking oligo					
	gRNA target site			Template	287+/- 2 bp (expected)					
	5' of TSS	3' of ATG	3' of TAA	ss oligo	286	287	289	286/289	287/289	Freq.
Expt 1 (F_0: 32 embryos)	+	+	+	-	2	-	14	-	-	0.5
	+	-	+	+	5	1	7	2	1	0.5
Expt 2 (F_0:16 embryos)	+	+	+	-	5		2	2		0.56
	+	-	+	+	1	1	5			0.44

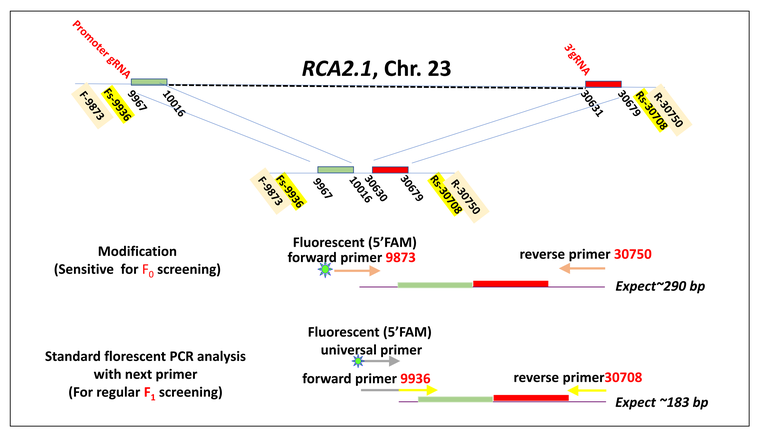
F1 screening	gRNA target site			Template	# embryos germ line transmitted			
n=16 embryos	5' of TSS	3' of ATG	3' of TAA	ss oligo	180+/- bp	Freq.	195+/- bp	Freq.
F_0-1	+	+	+	-	5	0.31	5	0.31
F_0-2	+	+	+	-	6	0.38	2	0.13
F_0-3	+	+	+	-	8	0.5	2	0.13
F_0-4	+	+	+	-	6	0.38	5	0.31
F_0-5	+	+	+	-	6	0.38	7	0.44
F_0-6	+	-	+	+	6	0.38	0	0.00
F_0-7	+	-	+	+	7	0.38	2	0.13
F_0-8	+	-	+	+	11	0.69	5	0.31

embryo > 1k fluorescent units by PCR
Frequency (Freq.) = sum of # embryos with expected PCR size/ total # embryos analyzed

gRNA targets and oligo template relative to r*ca2.1*

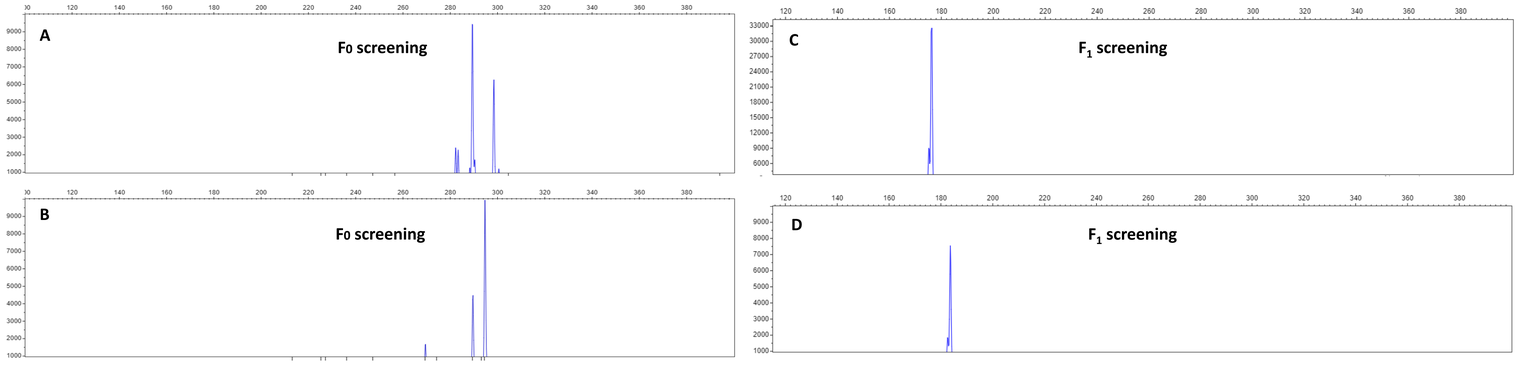


Screening strategy: fluorescent PCR fragment analysis



Representative somatic deletion profiles (A &B)
50% of F_0 embryos display predicted deletion amplicon ~290+/-bp

Representative germline transmitted deletion profile (C & D)



Sequence validation

Deletion of 20kb coding *rca2.1* from promoter to termination codon established in germline transmitted F_1s

- WT repair ss oligo template: 50 nt 5' upstream (YELLOW) and 3' downstream(red) of DSB of Cas9.
- PAM in pink relative to gRNA target site. PCR primer F (GREEN) and R (blue). F was also used for sequencing.

2 gRNA /ss Oligo co-injection

20 kb deletion : promoter to termination codon

5' flanking 3' flanking

```
Predicted fusion:183BP  ACATTAACATCTGTATTAAGGAAATAGTGTGCGATACTTAGGTTTTTTTCAGCCCTGCGGATTTCCTGTGTGCATCCGCTCTTCGTGGTAAGAACAAAGGGATCTATGATCTAAGCCAGTCTATAAAAGAATAAGAAGCATCTCCTTCCTTTCTTGTTAC
#18:183BP               ----------------------------GTGCGATACTTAGGTTTTTTTCAGCCCTGCGGATTTCCTGTGTGCATCCGCTCTTCGTGGTAAGAACAAAGGGATCTATGATCTAAGCCAGTCTATAAAAGAATAAGAAGCATCTCCTTCCTTTCTTGTTA-
#19:176BP               ----------------------------GATACTTAGGTTTTTTTCNGCCCTGCGGATTTCCTGTGTGCATCCG--------GGAAGAACAAAGGGATCTATGATCTAAGCCAGTCTATAAAAGAATAAGAAGCATCTCCTTCCTTTCTTGG---
```

3 gRNA co-injection

```
Predicted fusion:183BP  ACATTAACATCTGTATTAAGGAAATAGTGTGCGATACTTAGGTTTTTTTCAGCCCTGCGGATTTCCTGTGTGCATCCGCTCTTCGTGGTAAGAACAAAGGGATCTATGATCTAAGCCAGTCTATAAAAGAATAAGAAGCATCTCCTTCCTTTCTTGTTAC-
#3:176BP                ----------------------------GTGCGATACTTAGGTTTTTTTCAGCCCTGCGGATTTCCTGTGTGCA--------TCGGGGTAAGAACAAAGGGATCTATGATCTAGGCCAGTCTAGAAAAAAATAAGAAGCATCTCCTTCCTTTCTTGTGAC--
#4:183BP                ----------------------------TGTGCGATACTTAGGTTTTTTTCAGCCCTGCGGATTTCCTGGGGGNATCCGCTCTTCGTGGTAAGAACAAAGGGATCTATGATCTAAGCCAGTCTATAAAAGAATAAGAAGCATCTCCTTCCTTTCTTGTTA---
#7:180BP                ----------------------------GCNATNCTTAGGTTTTTTTCAGCCCTGCGGATTTCCTGTGTGCATCCGC---TCGTGGTAAGAACAAAGGGATCTATGATCTAGGCCAGTCTATAAAAGAATAAGAAGCATCTCCTTCCTTTCTTGTTAGNA
#8:181BP                ----------TTGTGCGATACTTAGGTTTTTTTCA--CCCGCGGNTTTCCTGTGTGCATCCGCTCTTCGTGGTAAGAACAAAGGGATCTATGATCTAAGCCAGTCTATAAAAGAATAAGAAGCATCTCCTTCCTTTCTTGTT----
```

Conclusions

- Large genomic deletions spanning from 5' of the TSS to 3' of the TAA stop of *rca2.1* were generated and germ-line transmitted. The majority of recovered alleles had precise genomic deletions, and the remaining had minor deletions, regardless of whether 3 gRNAs alone or 2 gRNAs with repair template were used. This does not represent the entire range of outcomes, as F_0s with F_1 progeny displaying peaks of the wrong size were excluded.

- Unlike classic CRISPR for in/del mutations, gene requirements can be explored using these alleles presumably without the concern of residual or potential novel mRNAs bearing coding sequence or nonsense-mediated decay linked genetic compensation.