9

Large alterations in the genomes of pigs edited using CRISPR-Cas9

Giselle Blanco[1†], Christine Deaver[1], Karyn Howard[1], Cong Li[1], Carlo Mercado[2], Ashley L. Cook[2], Dhanansayan Shanmuganayagam[3], C. Dustin Rubinstein[4], Adam Moyer[2], Alexis Norris[2], Mayumi Miller[1*]

[1]Office of Applied Science (OAS) and [2]Office of New Animal Drug Evaluation (ONADE), Center for Veterinary Medicine (CVM), FDA
[3]Center for Biomedical Swine Research and Innovation and [4]Biotechnology Center, University of Wisconsin-Madison, Madison, WI
*presenting; †current affiliation: Center for Devices and Radiological Health (CDRH), FDA

Introduction

CRISPR-Cas9 genome editing is being used to introduce intentional genomic alterations (IGAs) into pigs to improve both animal health (e.g., disease resistance) and human health (e.g., xenotransplantation). Unintended large alterations (>50bp) are known to occur during genome editing and include deletions, duplications, inversions, and insertions (collectively, structural variants [SVs]). Their putative effect on gene function may present hazards to the health of the animal and the safety and efficacy of the human product, in the cases of 'biopharm' animals. Compared to humans, there is relatively little data on the large genome-wide effects of CRISPR-Cas9 editing in pigs. This study sought to characterize SVs in pig genomes following *in vitro* CRISPR-Cas9 editing of three pig genes, singly or in combination (simultaneous or sequential).

Materials and Methods

Primary pig cells from newborn ear fibroblasts were generously provided by our collaborators at the University of Wisconsin. We designed guide RNAs (gRNA) using CHOPCHOP[1] targeting *CMAH* exon 3 (chr7), *GGTA1P* exon 7 (chr1), and *β4GALNT2* exon 7 (chr12). The resulting gRNAs were delivered singly or in combination (**Fig. 1**). Lonza's P3 Primary nucleofection kit and pulse code EH-100 were used to deliver 10μM of single gRNA (IDT) and 10μM of TrueCut Cas9 V2 (ThermoFisher) to the cells. One day post nucleofection, the cells were plated at a limiting dilution for clonal growth. Clones with Sanger-confirmed on-target edits were sent to Psomagen for whole genome sequencing (WGS), along with an unedited control. Psomagen sequenced Illumina TruSeq Nano DNA kit (550bp) libraries to 30X coverage on Illumina NovaSeq 6000 instrument.

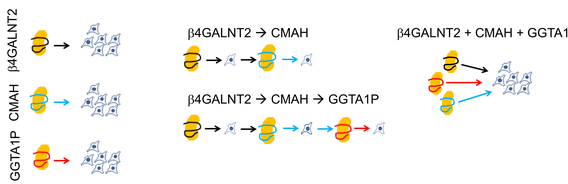

Figure 1. Editing schematic. Samples were edited with either single sgRNAs, multiple sequential sgRNAs, or multiple simultaneous sgRNAs

WGS fastq data were transferred to precisionFDA[2] for bioinformatics analysis by CVM (**Fig. 2**). Trimmomatic[3] quality-trimmed reads were aligned to susScrofa 11.1 (Ensembl) using bwa-mem[4] and SVs called using lumpy[5]. SnpSift[6] was used to filter out SVs that were low quality or present in an unedited control sample. SVs were visually confirmed in IGV[7]. Cas-OFFinder[8] (v3.0.0b3) was used to identify predicted target sites allowing for up to eight sequence differences (up to six mismatches [MM] and two bulges in either gRNA or target DNA), and NRG protospacer adjacent motif (PAM) sequence (canonical NGG or alternate NAG). RepeatMasker[9] was used to annotate overlapping repetitive elements.

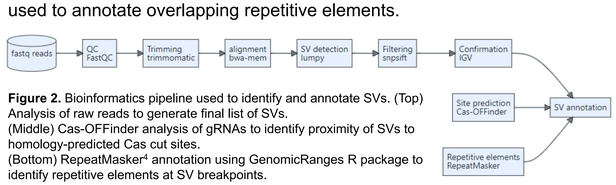

Figure 2. Bioinformatics pipeline used to identify and annotate SVs. (Top) Analysis of raw reads to generate final list of SVs.
(Middle) Cas-OFFinder analysis of gRNAs to identify proximity of SVs to homology-predicted Cas cut sites.
(Bottom) RepeatMasker[4] annotation using GenomicRanges R package to identify repetitive elements at SV breakpoints.

Results

No significant differences between editing strategies

We identified at least one SV in 90% of samples (27/30) (**Fig. 3B-C**). The samples in which B4GALNT2 ("B") and CMAH ("C") gRNA were introduced sequentially ("B->C") had a median of six SVs identified, which is greater than the sum observed when both gRNA were introduced singly (**Fig. 3A**). A similar effect was not seen with the addition of GGTA1P ("G") gRNA ("B->C->G"). The simultaneous addition of the three gRNAs ("B+C+G") trended towards less SVs, but this did not reach significance (p = 0.18, unpaired Wilcoxon test) (**Fig. 3B**). Individual samples harbored up to eight SVs, including deletions (DEL), duplications (DUP), inversions (INV), and insertions (INS) (**Fig. 3C**). Note: the lumpy software reports the insertions as translocations because of the junctions between the two chromosomes.

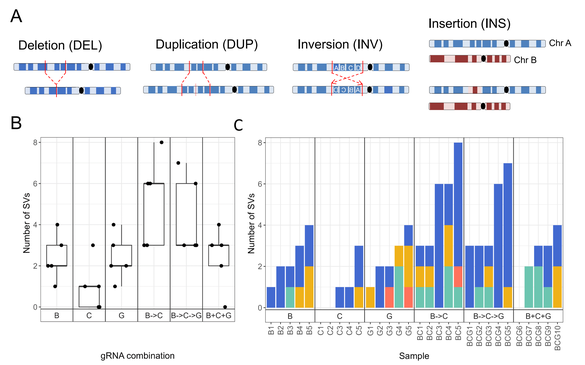

Figure 3. Frequency and form of SVs identified in edited cell lines but not present in unedited cell line. (A) Schematic of SV types. The insertions we identified were due to a region of one chromosome ("Chr B") inserted in a second chromosome ("Chr A"), likely the result of the cell using "Chr B" as a template when repairing DNA break on "Chr A". (B) Total number of SVs identified in each sample by gRNA combination; box indicates 25-75% (Q25-Q75); horizontal line indicates median; whiskers extend to 1.5 x interquartile range. (C) Number of SVs identified in each sample by SV type.

Majority of SVs were deletions; sometimes very large

Most observed SVs were deletions (58.7%), followed by duplications (25.4%), insertions (11.1%), and inversions (4.8%) (**Fig. 4A**). Surprisingly, most deletions were large (57%, >50kb), and some were very large (16%, >250kb) (**Fig. 4B; Table 1**).

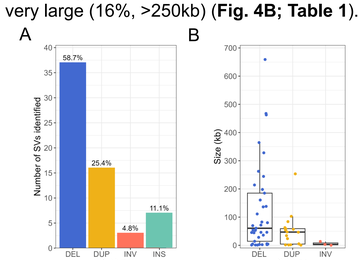

Table 1. Statistical summary of SV sizes by type (in bp).

	SV Type		
	DEL	DUP	INV
N	37	16	3
Min	890	471	110
Max	658,987	253,489	13,648
Mean	126,749	52,178	5,678
SD	156,345	62,190	7,081
Median	60,564	47,021	3,277
Q25	14,323	4,452	1,694
Q75	185,101	59,181	8,463

Figure 4. Frequency and form of SVs identified in edited cell lines but not present in unedited cell line. (A) Total number of SVs identified by type. (B) SV sizes (in kb); box indicates 25-75% (Q35-Q75); horizontal line indicates median; whiskers extend to 1.5 x interquartile range.

Proximity to *in silico* off-target site predictions

Cas-OFFinder was used to predict Cas9 cut sites based on the susScrofa11.1 reference genome, Cas PAM, and gRNA homology thresholds (**Fig. 5A**). This analysis generated a list of nearly three million potential Cas9 cut sites to assess overlap of identified SVs in CRISPR-Cas9 edited cells (**Fig. 5B**). All SVs were within 20kb of a predicted site using our permissive parameters (up to eight differences). However, when limiting to the more common threshold (four differences), only the SVs at the target sites would be identified (**Fig. 5C**).

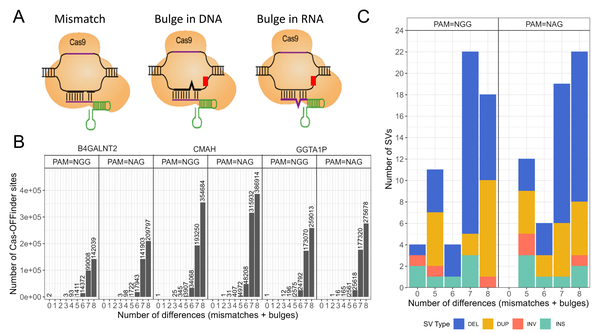

Figure 5. SV proximity to Cas-OFFinder predicted sites. (A) Cas-OFFinder's homology-based prediction allows for mismatches between target DNA (black) and gRNA (purple), bulges in target DNA (black) or gRNA (purple), and alternate PAM (red). (B) Number of sites identified by Cas-OFFinder, after collapsing adjacent sites (±10bp). (C) Identified SVs are grouped by number of differences for nearest Cas-OFFinder site (for both NGG and NAG PAMs).

Repetitive elements at SV breakpoints

Given that all off-target SVs had poor gRNA homology (5-8 differences), we sought to assess non-CRISPR-Cas9 mechanisms for these events. Alternative mechanisms include somatic mutation (*de novo* events during cell culture) and mobile element insertions (MEIs). We assessed the repetitive elements (from RepeatMasker[9]) annotated at SV breakpoints. We found that the most common element at a breakpoint was Pre0_SS (**Fig. 6A**). Most SVs had a repetitive element overlapping at least one breakpoint (38/59; 64%), and four SVs had the same repeat class at both breakpoints (**Fig. 6B**).

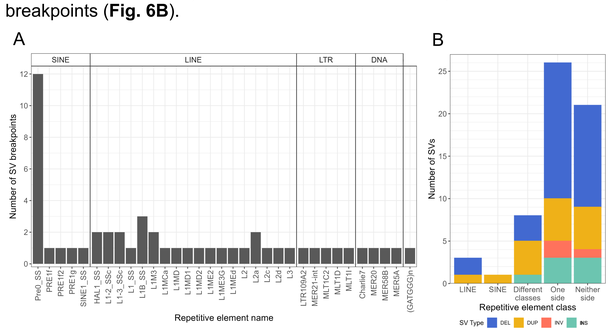

Figure 6. Repetitive elements at SV breakpoints. (A) Number of SV breakpoints overlapping the repeat element. (B) Number of SVs with repeat element class at both breakpoints.

Unintended large alterations at target site

We identified two SVs at *CMAH* target site that were insertions of a second chromosome, suggesting that chromosome was used as a template during DNA repair. The first was 642bp insertion (chr1:138,340,320-962) in place of 407bp deletion (chr7:19,903,384-791) in one sample (BC4) (**Fig. 7**). The second was 140bp insertion (chr13:15,879,019-159) in place of 13bp deletion (chr7:19,903,765-778) in two samples (BCG7, BCG8).

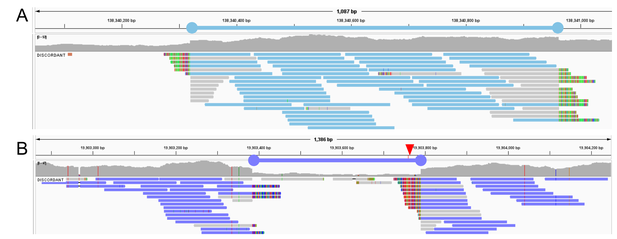

Figure 7. Unintended insertion at *CMAH* target site. Shown is read-level evidence for the 642 bp insertion of chr1 (A) in place of 140 bp deletion at *CMAH* locus on chr7 (B). The first level of evidence is read coverage (top panels): in A, there is an increased coverage for the region that was copied into *CMAH* locus (light blue line); in B, there is a decreased coverage for the deleted region (dark blue line). The second level of evidence is discordant read pairs and split reads (bottom panels). Red triangle indicates *CMAH* cut site (chr7:19,903,760).

Conclusions & Future Directions

Conclusions

- Unexpected SVs are commonly identified in pig cells after genome editing (90% of cells had ≥1 SV).
- SVs include complex genomic rearrangements at the target site.
- Alterations were identified in cells regardless of editing approach (single edit, sequential edits, or simultaneous edits).
- SVs are commonly deletions and duplications, which may be very large.
- All SVs at non-target sites had >4 differences. *In silico* prediction tools should include bulges and extend beyond four differences.
- It is unclear whether genome editing is causal mechanism for all SVs, suggesting the need for additional controls built in to test the contribution of Cas and/or nucleofection.

Future directions

- Use MEI tool to investigate MEI as an underlying mechanism for SVs.
- Additional Cas-only and mock transfection to determine their potential contribution to the instability/MEI activity potentially underlying SVs.
- *De novo* assembly of wildtype cell line to improve in silico predictions.
- Empirical site nomination using cell-based assay (i.e., GUIDE-Seq).
- Compare these *in vitro* results to *in vivo* editing of pig zygotes.

References

1. CHOPCHOP. https://chopchop.cbu.uib.no/
2. precisionFDA. https://precision.fda.gov/
3. Bolger, et al. Bioinformatics 2014.
4. Li. arXiv 2013 (1303.3997v2).
5. Layer, et al. Genome Biology 2014.
6. Cingolani, et al. Frontiers in Genetics 2012.
7. Robinson, et al. Nature Biotechnology 2011.
8. Bae, et al. Bioinformatics 2014. https://github.com/snugel/cas-offinder
9. Smit, et al. RepeatMasker Open-3.0. https://hgdownload.cse.ucsc.edu/goldenpath/susScr11/database/rmsk.txt.gz