Quantitative characterization of on-target and off-target variation induced by CRISPR+Cas9 systems at the single-cell resolution

Isai Salas-Gonzalez[1], Victor Weigman[1]

[1]BioSkryb Genomics, Durham, NC, United States

Introduction

Genome editing using CRISPR-Cas9 has revolutionized biomedical research by enabling the modification of DNA at targeted locations. The capacity of these systems to modify specific loci is dependent on the ability of Cas9 to induce DNA cleavage at genomic sites that exhibit precise base pairing with designed single guide RNAs (sgRNA). CRISPR/Cas systems are, conceptually, targeted editing tools, but the prevalence DNA edits that occur beyond the targeted gene (off-target effects) are increasingly observed. *In silico* tools have been developed to identify sgRNA guide-specific genome-wide potential off-target sites. Existing limitations of single cell genome enrichment have limited ability to confirm these in vitro, urging the need of unbiased methods of detection and validation. The goal of our study was to develop a bioinformatics workflow that could identify and prioritize on- and off-target gene editing events from single-cell whole genome sequencing data.

Methods

We leveraged previously published data in which CD34+ cord blood cells, U2OS sarcoma cells and the embryonic stem cell H9 were transfected with by two previously described sgRNAs (EMX1 and VEGFA) and Cas9. Mock treated and Cas9-only transfected cells served as a control. Single cells were isolated and primary template-directed amplification (ResolveDNA) was used to amplify genomes from single cells. Libraries were created and deep sequencing (~20X coverage) was performed. Secondary bioinformatics analysis was performed using the WGS pipeline available in BaseJumper (BioSkryb Genomics). We called CNV events using Ginkgo, translocations were called using an ensemble approach overlapping events called by Manta and GRIDSS. In silico prediction of off-target events was done with Cas-OFFinder allowing up to 7 mismatches and 1 indel.

Results

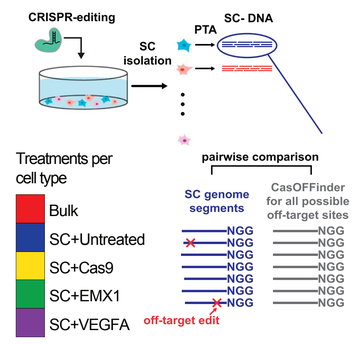

Figure 1: Experimental design. Overview of experimental and computational strategy where single edited cells are sequenced and indel calling is intersected to *in-silico* predcted sites with up to seven mismatches and 2 indels with the protospacer. Note that per each cell line (CD34+, U20S and H9) we have 5 distinct treatments: Bulk, SC+Untreated, SC+Cas9, SC+EXM1 and SC+VEGFA (colors in tiles).

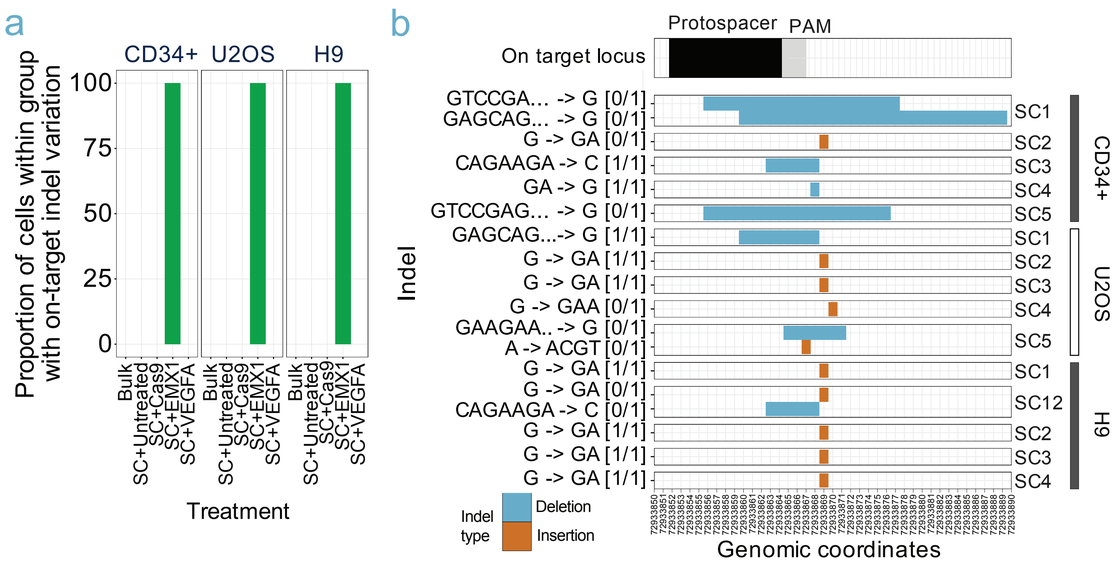

Figure 2: Detected indels within the EMX1 target locus. **a:** Bargraphs denoting the total proportion of cells (y-axis, n=5 per treatment), for each treatment (x-axis) across cells (x-axis facets), that exhibited indel variation in the targeted EMX1 locus. Note the 100% effectiveness of the SC+EMX1 treatment in inducing indel variation across the three cell types. **b:** Plot showing the specific indel calls (y-axis) for each of the cells (y-axis facets) with indel variation quantified in 2a across the EMX1 locus (x-axis). Fill colors represent the type of indel. The location of the EMX1 guide site is displayed in the top panel of the plot. Note the heterogeneity of indel calls across the dataset.

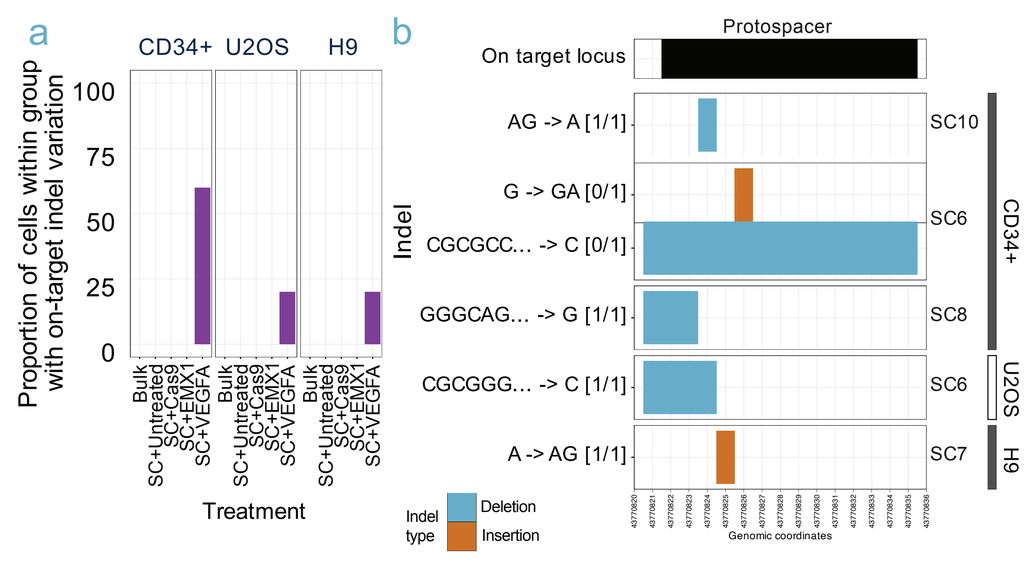

Figure 3: Detected indels within the VEGFA target locus. **a**: Bar graphs denoting the total proportion of cells (y-axis, n=5 per treatment), for each treatment (x-axis) across cells (x-axis facets), that exhibited indel variation in the targeted VEGFA locus. Note the variable effectiveness of the SC+VEGFA treatment in inducing indel variation across the three cell types. **b:** Plot showing the specific indel calls (y-axis) for each of the cells (y-axis facets) with indel variation quantified in 3a across the VEGFA locus (x-axis). Fill colors represent the type of indel. The location of the VEGFA guide site is displayed in the top panel of the plot. Note the heterogeneity of indel calls across the dataset.

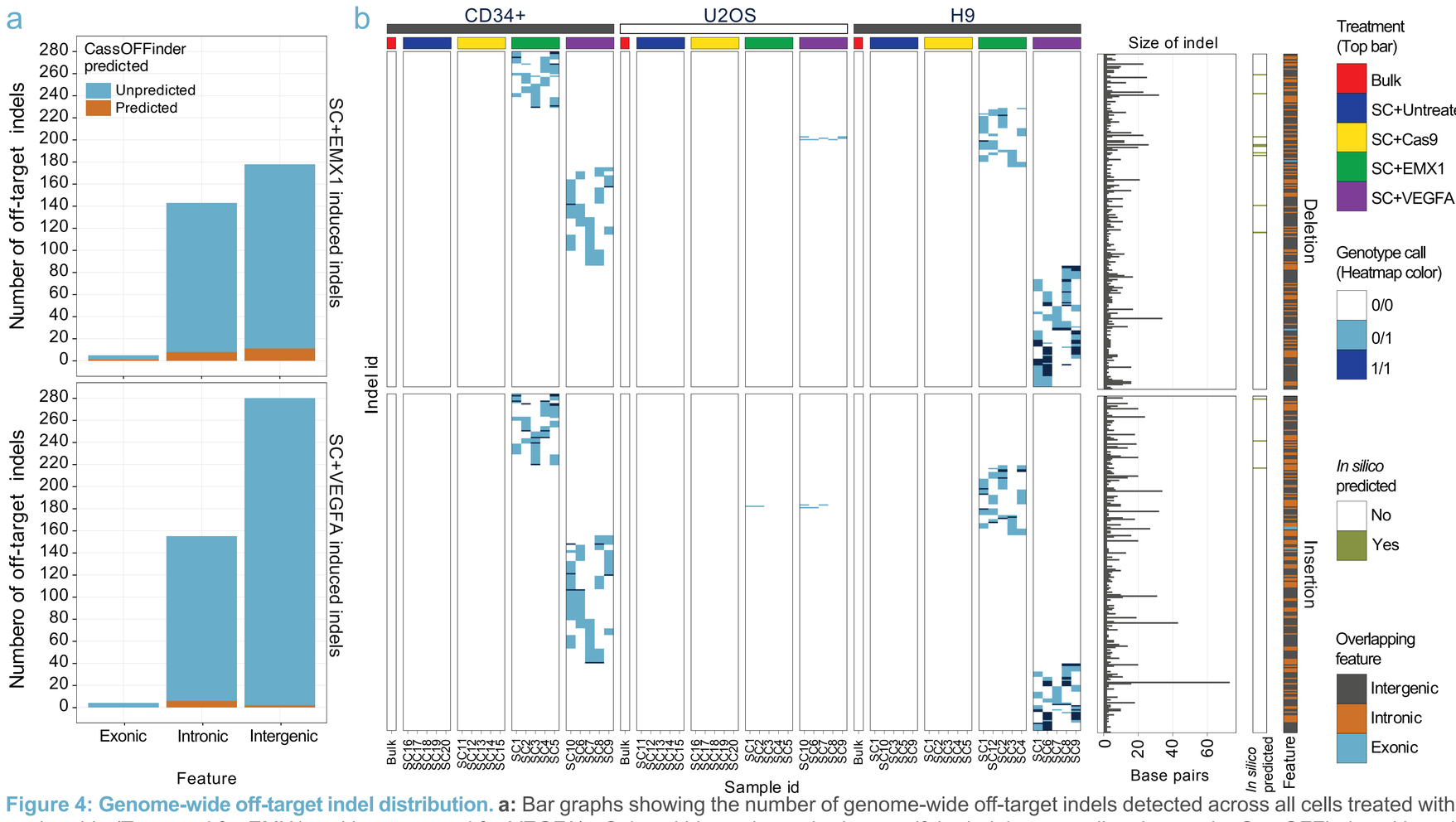

Figure 4: Genome-wide off-target indel distribution. **a:** Bar graphs showing the number of genome-wide off-target indels detected across all cells treated with each guide (Top panel for EMX1 and bottom panel for VEGFA) . Color within each graph denotes if the indel was predicted or not by Cas-OFFinder with and overlapping range of +/- 50 bps. Note that majority of (>85%) of off-target indels detected are unpredicted by the *in silico* approach. **b:** Heatmap showing the prevalence of the off-target indels (y-axis) exclusively induced by each of the two guides across the three cell types (x-axis) . Color of the tiles represent called genotype. Cells and treatments across the heatmap are denoted by the colored top bars. First panel in the right shows the size of the indels. Second panel in the right represents if the indel was predicted by Cas-OFFinder, finally the last bar represents the genomic feature that overlaps with each indel.

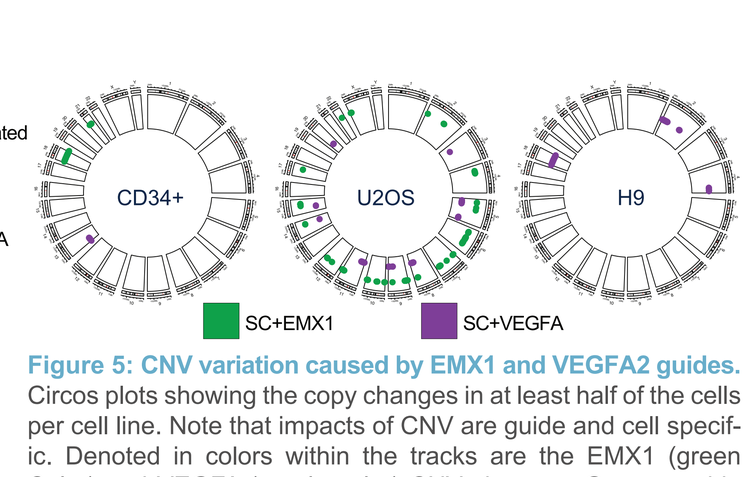

Figure 5: CNV variation caused by EMX1 and VEGFA2 guides. Circos plots showing the copy changes in at least half of the cells per cell line. Note that impacts of CNV are guide and cell specific. Denoted in colors within the tracks are the EMX1 (green Color) and VEGFA (purple color) CNV changes. Genome-wide we detected ~78 Mbp of changes in CD34+ cell, ~285 Mbp in U2OS and ~300 Mbp in H9.

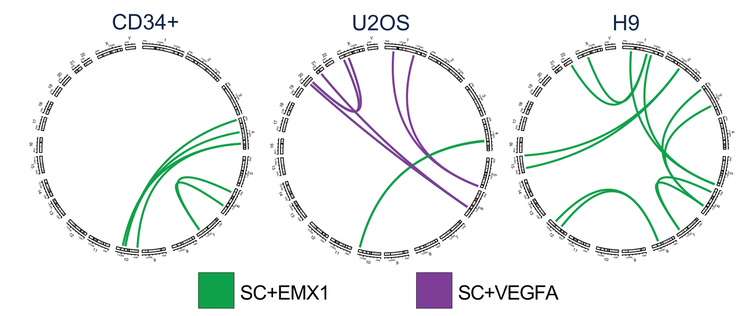

Figure 6: Induced translocations caused by EMX1 and VEGFA2 guides. Circos plots showing the reproducible (n>1) translocations events induced exclusively by guide treatments (SC+EMX1 and SC+VEGFA). Note that translocations events are mainly cell-specific but some breakpoint events are shared across cells. Colored lines within the circos plots denote the specific guide causative of the event.

Conclusions

- ResolveDNA (PTA amplification) coupled with our bioinformatics workflow permits to characterize unseen variation within sgRNAs targeted loci (Figures 2 and 3).
- In addition, we identified genome-wide off-target indels with potential physiological consequences induced by the sgRNAs treatments (Figure 4). The distribution of off-target indels within our dataset was cell and guide specific.
- Finally, we identified genome-wide CNV and translocations events exclusively induced by the guides treatments.

References

- Gonzalez-Pena, V., Natarajan, S., Xia, Y., Klein, D., Carter, R., Pang, Y., ... & Gawad, C. (2021). Accurate genomic variant detection in single cells with primary template-directed amplification. Proceedings of the National Academy of Sciences, 118(24), e2024176118.
- Bae, S., Park, J., & Kim, J. S. (2014). Cas-OFFinder: a fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases. Bioinformatics, 30(10), 1473-1475.

Acknowledgments

We are grateful to Professor Charles Gawad for granting access to the data used in this poster. In addition, we are thankful to the whole BioSkryb's Genomics team for their continuous effort and work.

If you are interested in learning more about our technology and this poster, scan the QR code or contact info@bioskryb.com