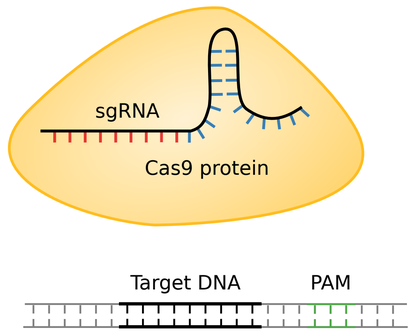
CrispRVariants: visualising gene editing experiments with R and shiny

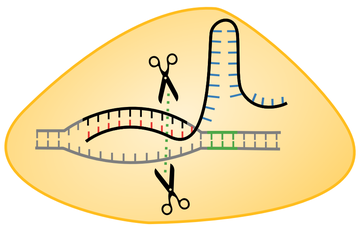
Helen Lindsay[1,2] and Mark D. Robinson[1,2]

1. Institute of Molecular Life Sciences, University of Zurich 2. Swiss Institute of Bioinformatics

CRISPR gene editing technologies enable DNA to be altered at precise locations

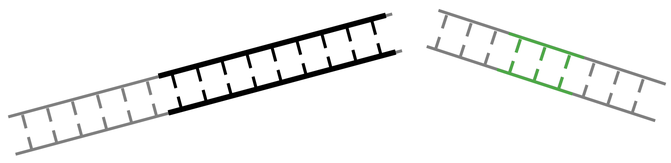
A complex of a single guide RNA (sgRNA) and a nuclease (here Cas9) is introduced into a cell. The sgRNA contains a region of ~20 nucleotides which binds to the complementary DNA region.





Figures adapted from www.clontech.com

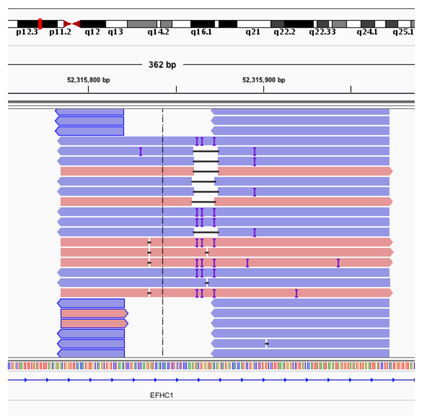
The target DNA must contain a protospacer adjacent motif (PAM). CRISPR nucleases differ in their PAM requirements and cutting location. Cas9 recognises PAMs with the sequence "NGG" downstream of the target region and cuts both strands of DNA ~3 bases upstream of the PAM.



Typically the broken pieces of DNA are repaired by an error-prone repair system known as non-homologous end joining. Nucleotides can be deleted…

…or inserted during the repair

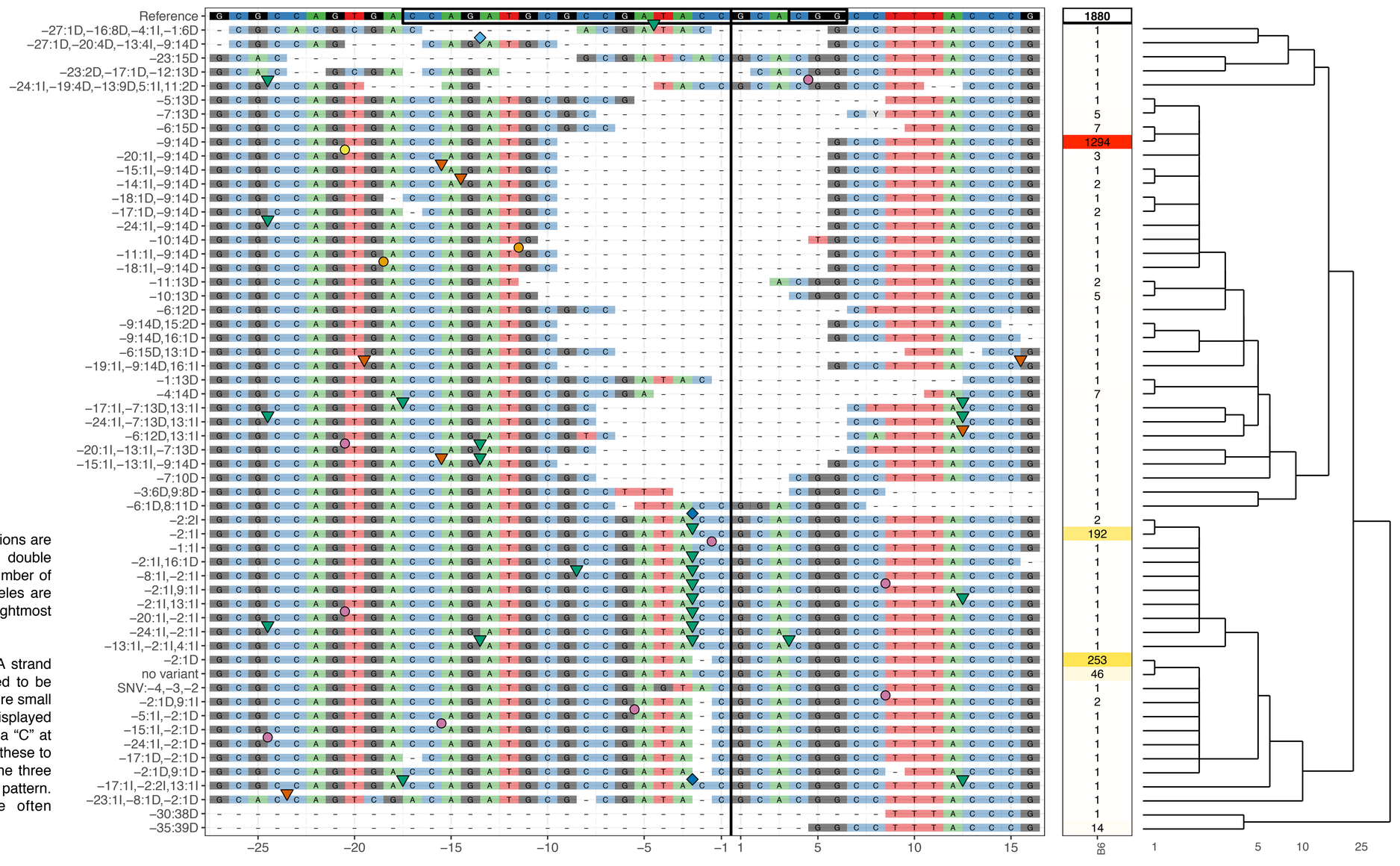
One way researchers assess the outcome of a CRISPR gene editing experiment is by sequencing the targeted DNA region.



Above: sequenced reads from an edited CRISPR target region aligned to the original genome sequence

Our R/Bioconductor package **CrispRVariants** allows researchers to evaluate and visualise sequences resulting from gene editing experiments. We have also developed a Shiny web application **CrispRVariantsLite** which implements a complete analysis pipeline for small, standard data sets.

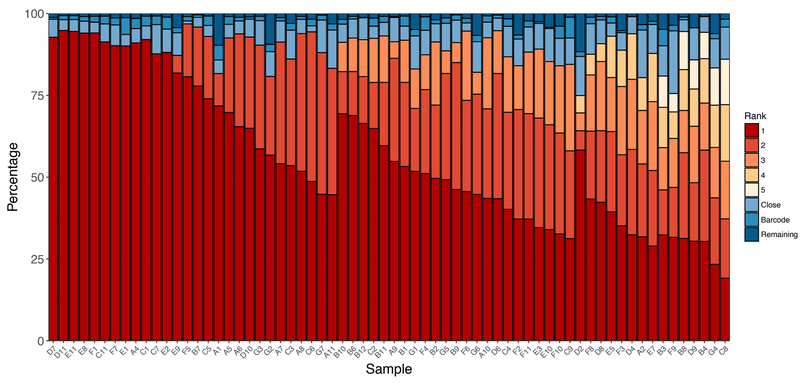
CRISPR technologies are evolving rapidly. Our ongoing work focuses on detecting and removing sources of error in CRISPR sequencing analyses as well as allowing users increased flexibility to specify custom variant counting and filtering methods.



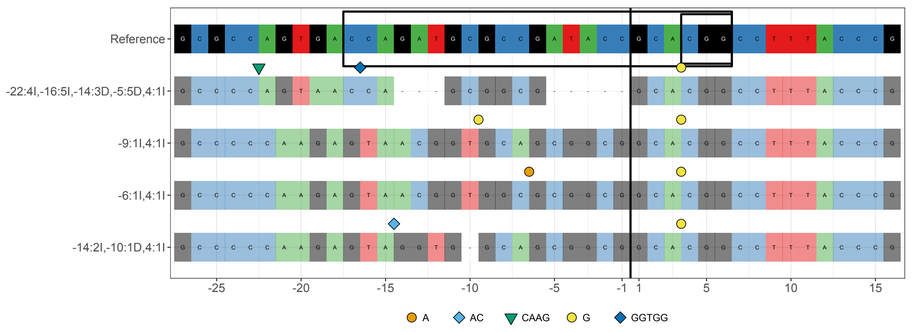
CrispRVariants allele plot. Insertions and deletions are numbered with respect to the site of the double stranded cut. The middle panel shows the number of reads matching each allele configuration. Alleles are clustered by Levenshtein distance. The rightmost figure shows the cluster dendrogram.

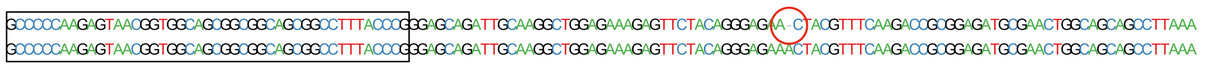
Mutations introduced during the repair of DNA strand breaks created by CRISPR-Cas9 are expected to be centred around the cut site; however in this figure small insertions and deletions are found across the displayed region. In particular single base insertions of a "C" at "GCC/CCG" motifs are common. We attribute these to sequencing errors, but also note that one of the three most common variant alleles (-2:1I) shares this pattern. True variants and sequencing errors are often ambiguous.

Reanalysis of data from Bell *et al,* BMC Genomics (2014) showing that the number of variants and their frequencies can vary substantially from theoretical expectations. Here, each column is derived from a single cell. Ideally, the variant proportions should reflect the number of DNA copies present when the mutation occurred. For example, a mutation that occurred in the original cell should be found in 50% of the sequenced reads, as a single (mammalian) cell contains two copies of DNA.



Recent developments in CRISPR technology enable repeated mutation at a single locus. In applications such as cell lineage tracking, researchers aim to determine the order of repeated mutational events. To date, studies have focused on the locations of insertions and deletions to reconstruct the order of events. However, sequencing errors and naturally occurring variation can affect how sequence aligners place insertions and deletions. The figure to the right shows four sequences which are identical within the region of interest despite differences in alignment. The figure below shows the complete sequences of the first two of these, which differ only in the presence of a single "A" outside the region of interest.





In the latest version of CrispRVariants, we added support for visualising subregions of interest from within a large target region. This figure shows results from a paired CRISPR experiment, where two separate sgRNAs were introduced simultaneously. Alignments that do not span the entire target region are indicated with ">".

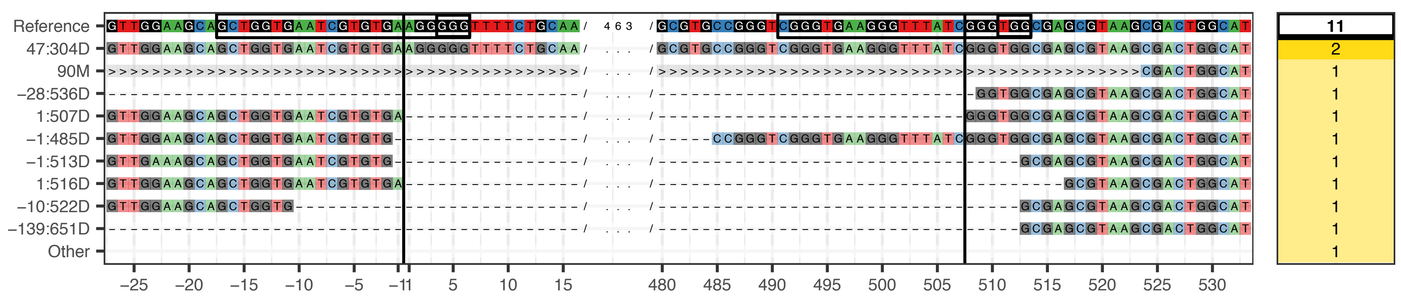


Figure from Mecariello *et al* (in press)

CrispRVariants
https://bioconductor.org/packages/release/bioc/html/CrispRVariants.html

CrispRVariantsLite
http://imlspenticton.uzh.ch:3838/CrispRVariantsLite/

Lindsay H, Burger A, Biyong B, Felker A, Hess C, Zaugg J, Chiavacci E, Anders C, Jinek M, Mosimann C and Robinson MD (2016). "CrispRVariants charts the mutation spectrum of genome engineering experiments." Nature Biotechnology, 34, pp. 701–702. doi: 10.1038/nbt.3628.

Contact: helen.lindsay@uzh.ch