Long-read Individual-molecule Sequencing Reveals CRISPR-induced Genetic Heterogeneity in Human ESCs

Chongwei Bi, Lin Wang, Baolei Yuan, Xuan Zhou and Mo Li

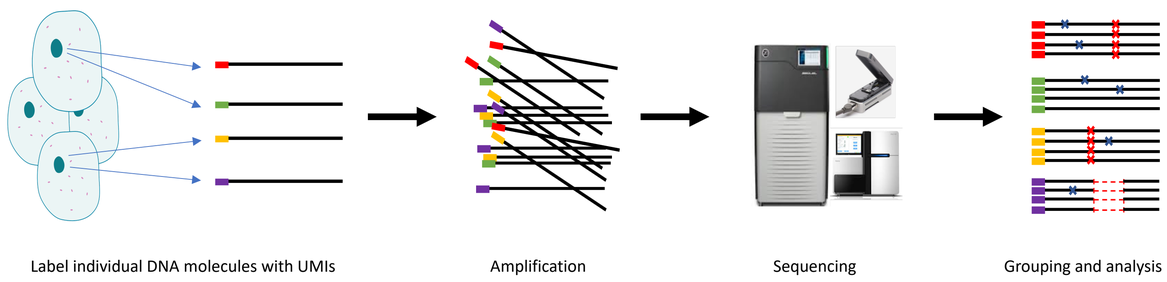
Biological and Environmental Science and Engineering Division, King Abdullah University of Science and Technology, Thuwal, KSA

King Abdullah University of Science and Technology

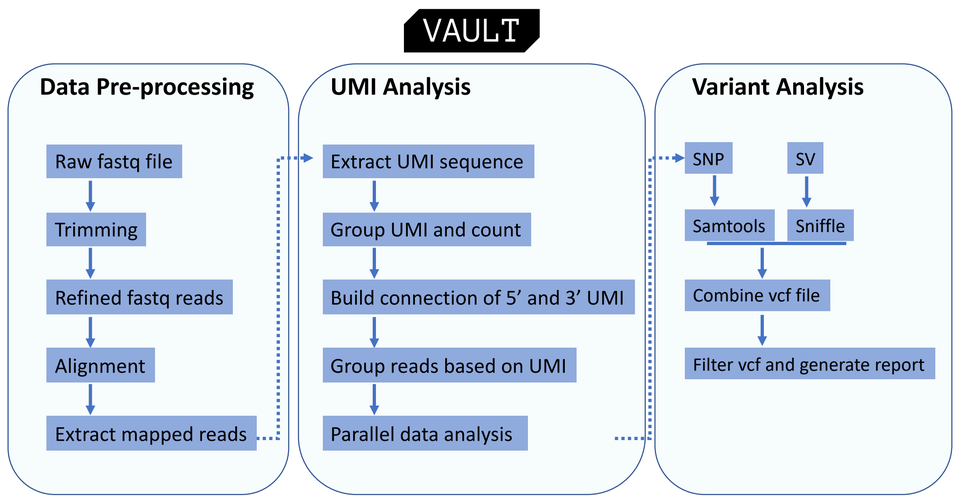
Motivation

Understanding the genetic heterogeneity of a cell population is extremely important for studies of biological systems. However, it remains challenging to analyze various types of genetic variants, because current methods are inadequate for detecting rare and/or complex variants. To address these issues, we develop a universal method to label individual DNA molecules for analyzing diverse types of rare genetic variants, with frequency as low as $4x10^{-5}$, using short- or long-read sequencing. It enables base-resolution haplotype-resolved quantitative characterization of rare variants. We showed that Cas9 cleavage induced SVs in ~4% of edited hESCs. Surprisingly, a significant number of SVs (up to 87%) were reoccurring deletions and insertions. These data provide the first quantitative evidence of nonrandom repair outcomes of Cas9 cutting and hotspots for Cas9-induced large indels.

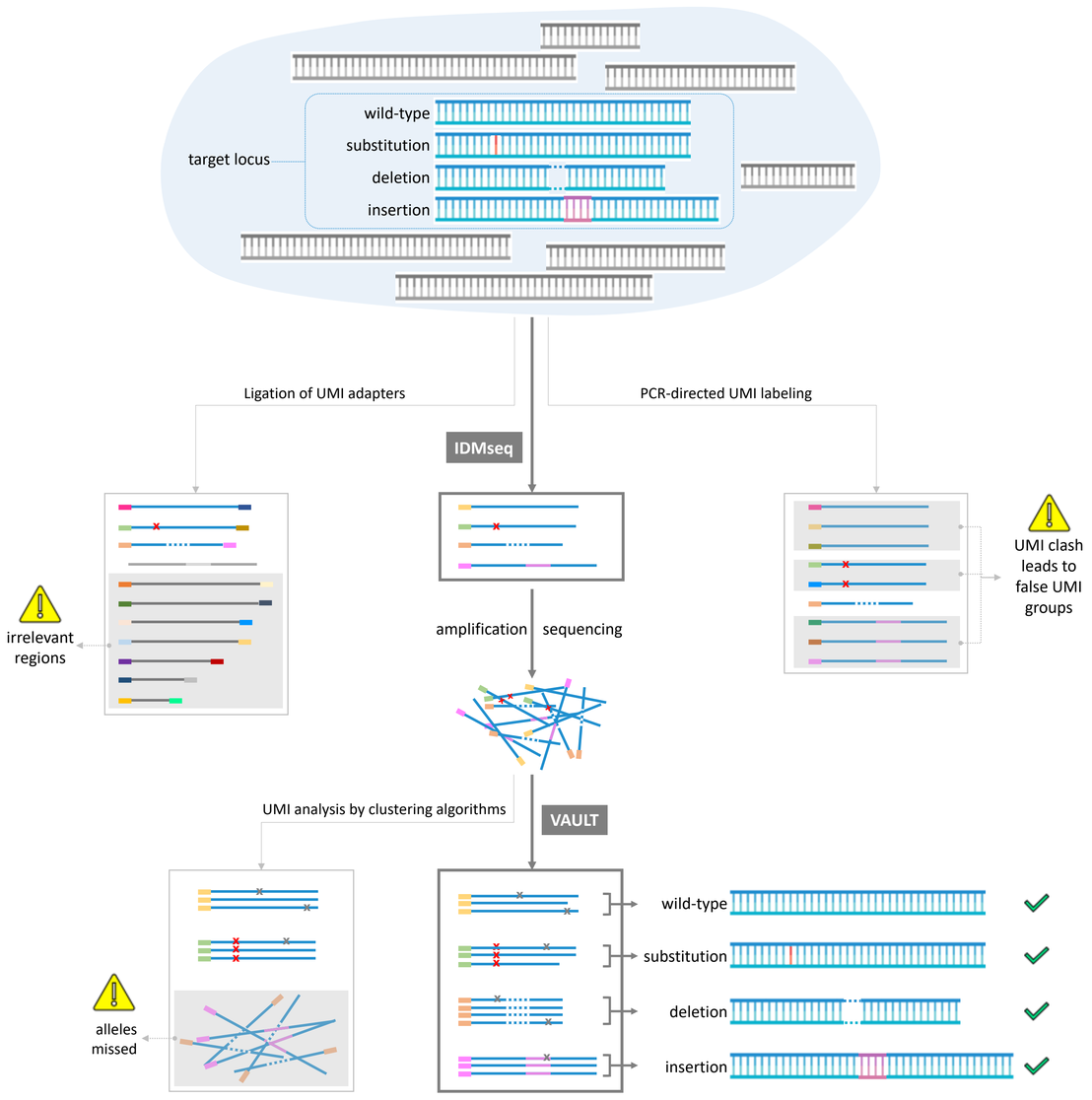
Targeted Individual DNA Molecule sequencing - IDMseq



Data analysis toolkit - VAULT

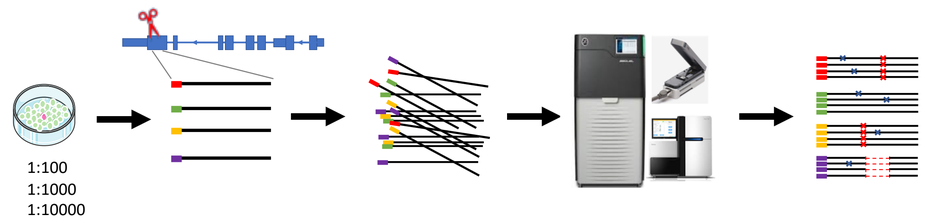


Advantages of IDMseq and VAULT as compared to current methods



IDMseq for detecting rare mutation

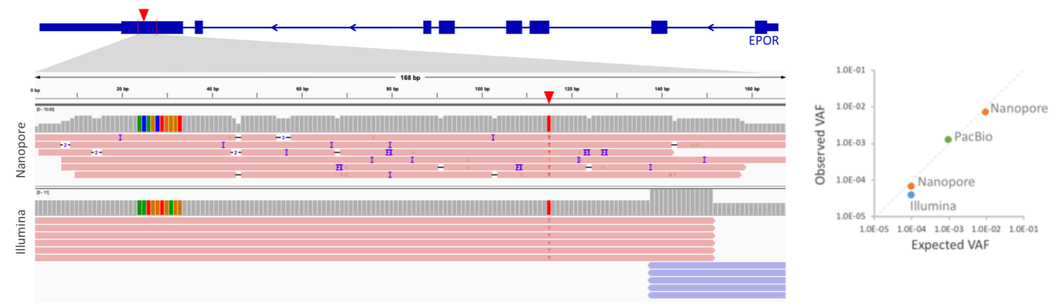
Schematic of using IDMseq to detect introduced rare mutation



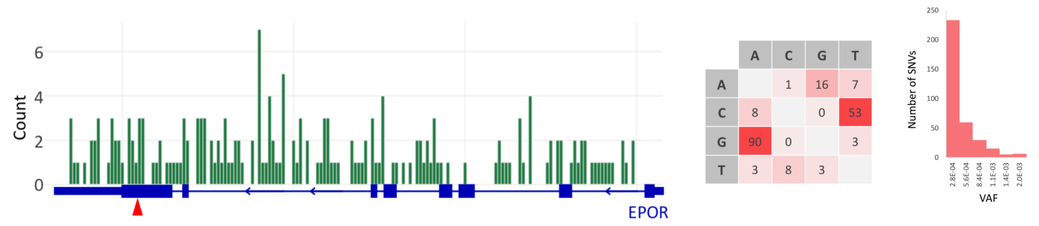
Sequencing summary

Gene	Mutant allele frequency (%)	Amplicon size	Sequencing platform	Read count	Reads with UMI	UMI groups for variant calling (>= 5 reads)	UMI groups with introduced mutation	Somatic SNV count	Somatic SNV load Per megabase	SV groups
EPOR	1:100 (1%)	168 bp	Nanopore	17,634	6,444	284	2 (0.7%)	0	N/A	N/A
EPOR	1:1,000 (0.1%)	6,789 bp	PacBio	227,206	136,399	3,184	4 (0.126%)	192	8.9	3
EPOR	1:10,000 (0.01%)	168 bp	Nanopore	1,093,683	494,009	15,598	1 (0.006%)	10	7.1	N/A
EPOR	1:10,000 (0.01%)	168 bp	Illumina	7,488,257	7,236,007	132,341	5 (0.004%)	85	7.1	N/A

Introduced SNVs detected in 1:10k sequencing

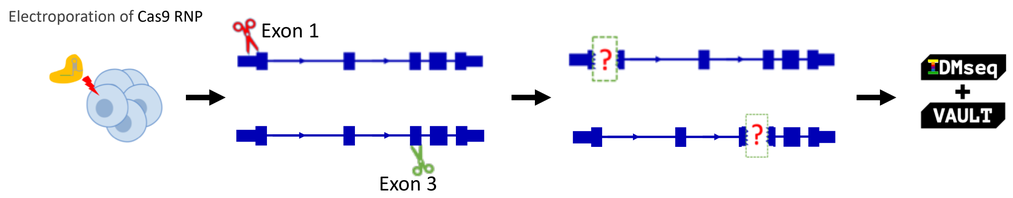


Somatic SNVs detected in PacBio sequencing



IDMseq for characterizing the outcomes after CRISPR genome editing

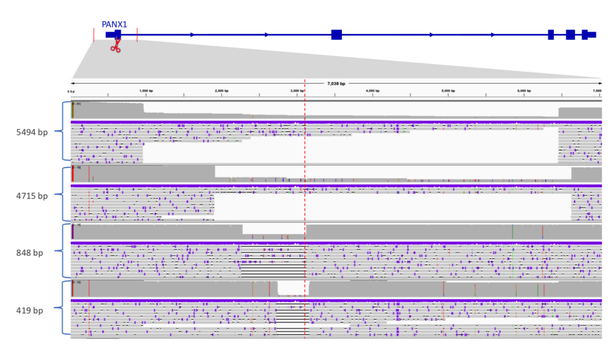
Experimental design

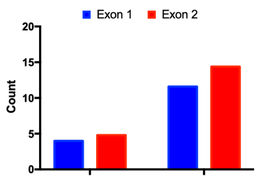


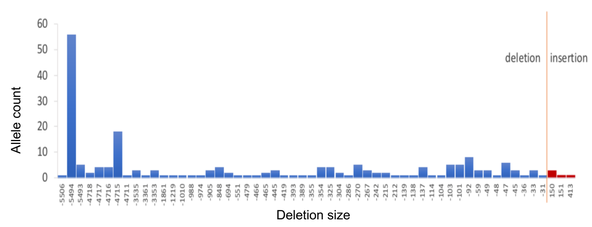
Sequencing summary

Gene	type	Amplicon size	Sequencing platform	Read count	Reads with UMI	UMI groups for variant calling (>= 5 reads)	Somatic SNV count	Somatic SNV load Per megabase	SV groups
PANX1	Control of exon 1	7077 bp	Nanopore	403,885	111,146	1816	51	4.0	0
PANX1	Control of exon 3	6595 bp	Nanopore	233,164	75,219	2213	69	4.8	0
PANX1	Cas9 editing in exon 1	7077 bp	Nanopore	2,761,805	613,147	3,566	293	11.6	200 (5.6 %)
PANX1	Cas9 editing in exon 3	6595 bp	Nanopore	3,078,165	1,042,582	8,870	843	14.4	232 (2.6 %)

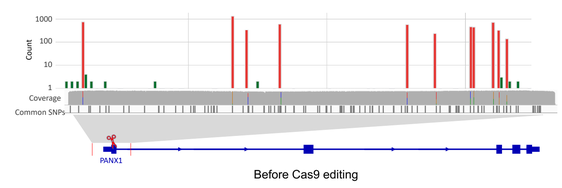
Large SVs detected in Cas9 edited exon 1 sequencing

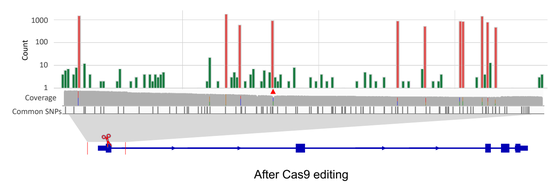






SNVs detected in Cas9 edited exon 1 sequencing





Find our work in bioRxiv

Chongwei Bi, et al. (2020). Long-read Individual-molecule Sequencing Reveals CRISPR-induced Genetic Heterogeneity in Human ESCs. bioRxiv 2020.02.10.942151.

doi: https://doi.org/10.1101/2020.02.10.942151

Laboratory of Stem Cell and Regeneration
King Abdullah University of Science and Technology

Contact: *chongwei.bi@kaust.edu.sa*
mo.li@kaust.edu.sa
Lab website: *regeneration.kaust.edu.sa*