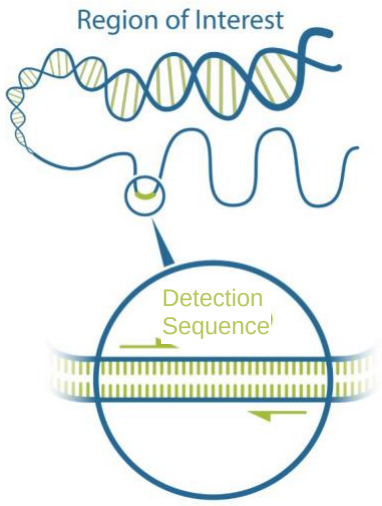
Verifying Genome Editing with Xdrop: Indirect Sequence Capture and Detection of a Transgene in a Mouse Line

Background

Zygote transformation procedures have been used for decades to create transgenic animal models that express exogenous genes. Since 1981, when the first transgenic mice were produced by microinjection of DNA into single cell embryos, almost 6,500 transgenic alleles have been created in mouse lines expressing reporter systems or key genes involved in diseases.[1] Notoriously, injected transgenes insert randomly into the zygote genome and the exact insertion site and pattern is unknown for numerous lines. Traditional methods to identify insertion site, like DNA FISH or mate pair sequencing, are expensive or fail to reveal critical deletions and rearrangements. The long DNA fragments enriched by Xdrop® addresses these issues and makes insertion site detection a straightforward process.

Xdrop enriches long (~100 kb) target DNA regions selected by amplifying a Detection Sequence corresponding to a small portion of the Region of Interest (ROI) or in flanking regions – in this case the transgene. This amplicon is exclusively used to detect and select droplets containing the ROI, thereby enriching it for capture and sequencing.



We evaluated the Xdrop technology for the purpose of identifying the integration site of a Pax8-CreERT2 transgenic mouse line (see graphic below).

The Xdrop Technology

The Xdrop technology combines droplet PCR (dPCR) with high-resolution droplet sorting followed by Multiple Displacement Amplification in droplets (dMDA).

Firstly, Xdrop partitions high molecular weight genomic DNA into millions of double emulsion droplets. Droplets containing the DNA molecules with the ROI are identified by a ~150 bp dPCR specific to a Detection Sequence within or adjacent to the region of interest.

The detection and sorting of droplets are performed using a standard cell sorter, which allows the PCR positive droplets containing the ROI to be collected. The sorted long DNA fragments are finally amplified in droplets (dMDA) to ensure unbiased DNA amplification.

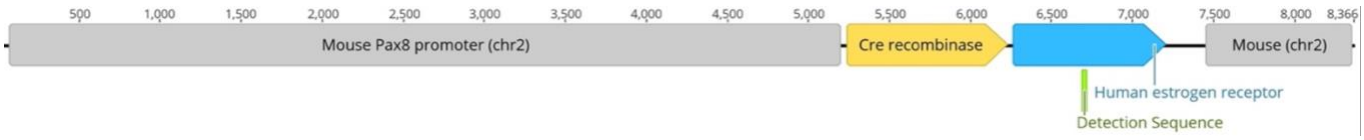
The Xdrop enrichment and amplification technology is compatible with both long- and short-read library preparation and sequencing.

Main Applications of Xdrop targeted enrichment

- Structural Variations
- Tandem Repeats
- GC-rich Regions
- Gap-closing
- CRISPR edit verification
- Integration of transgenes and viral DNA

Generating the transgenic mouse line

The transgenic mouse line was developed at the Center for Advanced Preclinical Research at the U.S. National Cancer Institute. A plasmid with the Pax8-rtTA sequence[2] was modified by replacing the rtTA sequence with the CreERT2 sequence from a pCAG-CreERT2 plasmid (Addgene). Here, the mutated ligand-binding domain of the human estrogen receptor allows tamoxifen-induced Cre recombinase activity. The Pax8 part of the construct was derived from the promotor of the mouse *Pax8* gene. The linearized transgene was microinjected into fertilized mouse eggs from C57Bl/g mice (The Jackson Laboratory) and mouse lines were identified with the desired expression pattern indicating integration.



Enrichment and read mapping

We enriched DNA fragments from 5 ng input material using a single primer set to amplify a Detection Sequence located within the human estrogen receptor part of the construct. After dMDA, we measured enrichment by qPCR to be >100x. We then sequenced libraries prepared from the enriched samples on Oxford Nanopore's MinION, which resulted in approximately 150,000 reads.

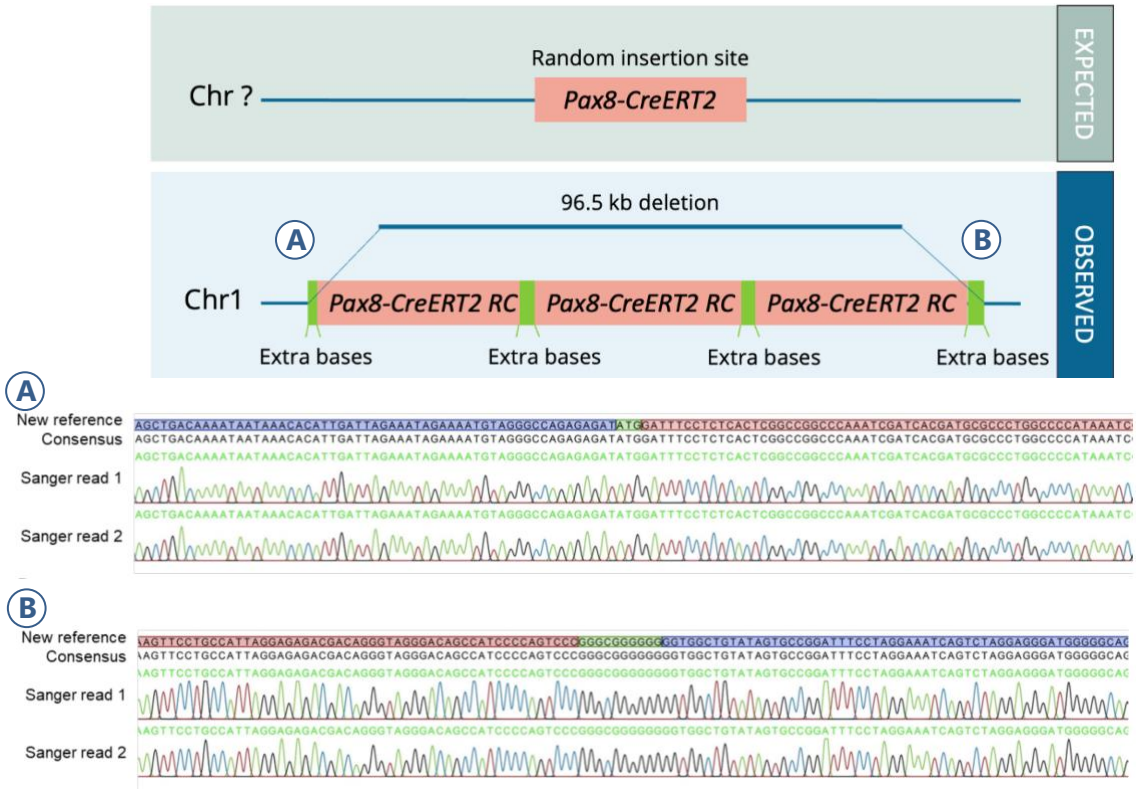
A total of 2,544 reads mapped to the full Pax8-CreERT2 construct with a mean coverage of 220-fold. These reads were mapped back to the mouse genome to identify reads that would span the breakpoint between insert and genome. The reads mapped to four locations: (1) the Pax8 gene on chromosome 2; (2) a region on chromosome 4 sharing homology with part of the construct; (3) the mouse estrogen receptor homolog on Chromosome 10; and (4) a region on chromosome 1 where 45 reads mapped.

Complex structure of insert at chromosome 1

The reads that mapped to chromosome 1 showed the two borders of the construct insertion at Chr1: 93,575,236 and Chr1: 93,671,773. Furthermore, the data indicated that the insertion resulted in a 96.5 kb deletion in chromosome 1, and that the construct was inserted three times in series with variable truncations and insertions at the borders (see figure below). We confirmed the identified insertion site on chromosome 1 by Sanger sequencing of PCR amplified border regions (see A and B below). Both Sanger sequencing and Oxford Nanopore reads corroborate the presence of two inter-transgenic regions differing by 28 bp in length.

Conclusions

With simple design and Indirect Sequence Capture of long DNA fragments, Xdrop facilitates identifying transgene insertion sites, which are often difficult and costly to find. Random integration and the presence of host genomic sequences in constructs complicate integration site identification. Xdrop makes validating genome engineering straightforward.



Acknowledgements

This project was carried out in collaboration with Dr. Ludmila Szabova at the NCI and received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 848497. The original work can be found in the journal Methods.

References and notes

1. Srivastava, A., et al. 2014. Discovery of transgene insertion sites by high throughput sequencing of mate pair libraries. BMC Genomics 15: 367.
2. The plasmid was obtained from Dr. Robert Koesters from INSERUM/Université Pierre et Marie Curie, France.