Simplifying transgene insertion analysis with Xdrop® Sort

Edward J. O'Neill[1], Cecilie Nyholm Andersen[2] and Lydia Teboul[1]

[1]Mary Lyon Centre, MRC Harwell, Oxfordshire, OX11 0RD, United Kingdom

[2]Samplix ApS, Bregnerødvej 96, 3640 Birkerød, Denmark

Summary

- Analysis of transgene insertion sites and patterns remains challenging and laborious using current technologies.
- Here, with a single primer set and low amount of input DNA, we identify the insertion sites and characterize the surrounding genomic landscapes of a transgene
- The Xdrop Sort workflow combined with long-read sequencing simplifies the identification of both insertion sites and patterns.

Introduction

Additive transgenesis is an affordable and reliable way to generate transgenic animals expressing a gene of interest. The methodology for microinjection of the gene of interest into the pronucleus of fertilized eggs is well described and the success rates for generation of transgenic animals is fair. Unfortunately, identifying the transgene insertion site and pattern remains challenging with existing technologies.

Here, we used Xdrop Sort and an Oxford Nanopore® GridION to enrich and sequence DNA from the insertion(s) and the flanking genomic region(s). This allowed the characterization of both the insertion regions and the insertion patterns.

Generation of the transgenic mouse line

The transgenic mouse line was developed at the Mary Lyon Centre at MRC Harwell. A purified 6.8-kb DNA fragment generated from an HpaI and XbaI double digest of pK18-hACE2[1] was microinjected into the pronucleus of one-cell mouse embryos. The genotypes of the founder animals and two subsequent generations were analyzed using droplet digital PCR to estimate the transgene copy numbers. An animal with an estimated 7 transgene copies was selected for further insert analysis using Xdrop Sort.

Enrichment

Using Xdrop Sort and an Xdrop DE20 Cartridge, 2 ng of mouse genomic DNA was encapsulated in DE20 droplets with PCR reagents and a PCR primer set (forward primer: TGTCACTTTCTGCAGCCACA; reverse primer: TGGCAGAGTCCCAACAATCG) specific for the hACE2 transgene generating a 131 bp amplicon.

Then, using Xdrop Sort and an Xdrop DE20 Sort Cartridge, we sorted 739 stained amplicon-positive droplets (Figure 1), followed by droplet breaking. Amplification of the recovered DNA was performed via droplet-based MDA in SE85 droplets generated using Xdrop Sort and an Xdrop SE85 Cartridge.[2]

Enrichment was measured using qPCR based on the enriched versus non-enriched DNA. Over 700x enrichment was found. This was considered acceptable since the DNA was from an animal with multiple transgene insertions. Next, we sequenced the libraries prepared from the enriched samples on an Oxford Nanopore Technology GridION using a single flowcell (R9.4.1), which yielded approximately 220,000 reads.

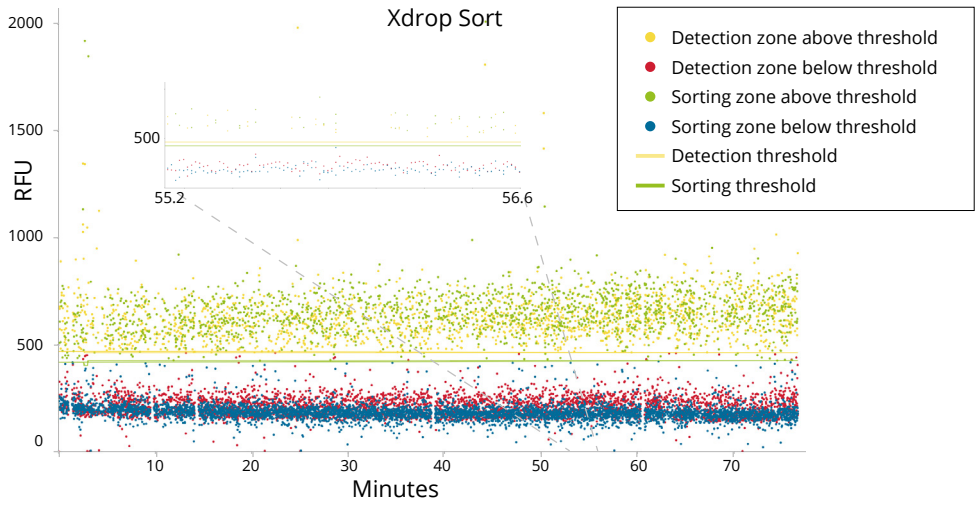

Figure 1. The Xdrop Sort Droplet Viewer analysis of droplet sorting shows that during the 76 minutes of sorting, 739 positive droplets were sorted. The insert shows a zoom-in for the time interval from 45.5 to 46.4 minutes. In Droplet Viewer, each sorted positive droplet appears as 4 dots, corresponding to 2 individual detections of each droplet in the detection zone (yellow) and 2 individual detections in the sorting zone (green). When support files are collected by the user, Xdrop Sort stops recording for a short period but continues sorting, this will apppear in the Droplet Viewer as gaps in the sort.

Read mapping

Using MiniMap2,[3] we generated a version of the mouse genome with the construct added as a pseudo-chromosome and masked regions highly homologous to the construct. Reads mapping to the construct were extracted and mapped back to the masked mouse genome and construct. Reads were found to map to 81 regions of the mouse genome, but only one region showed a high relative number of reads (Figure 2).

In-depth analysis of the aligned reads on Chromosome 1 showed multicopy integration in an intron of the gene *Cspp1* along with a 162-bp deletion (Figure 3). The copies flanking the insertion site showed an inverted integration pattern. The upstream and downstream border regions between the chromosome and the construct were validated using Sanger sequencing of PCR amplicons generated across the borders.

Conclusion

The Xdrop Sort workflow simplifies the process of identifying transgenes and characterizing transgene insertion sites. Just a single primer set is needed to enrich long DNA fragments containing the transgene and the neighboring chromosomal region. The enriched long DNA fragments are readily sequenced on long-read sequencing platforms and the availability of long-read sequencing information makes the data analysis easier. This provides a simple, fast, and affordable method for transgene insert analysis compared to whole genome sequencing.

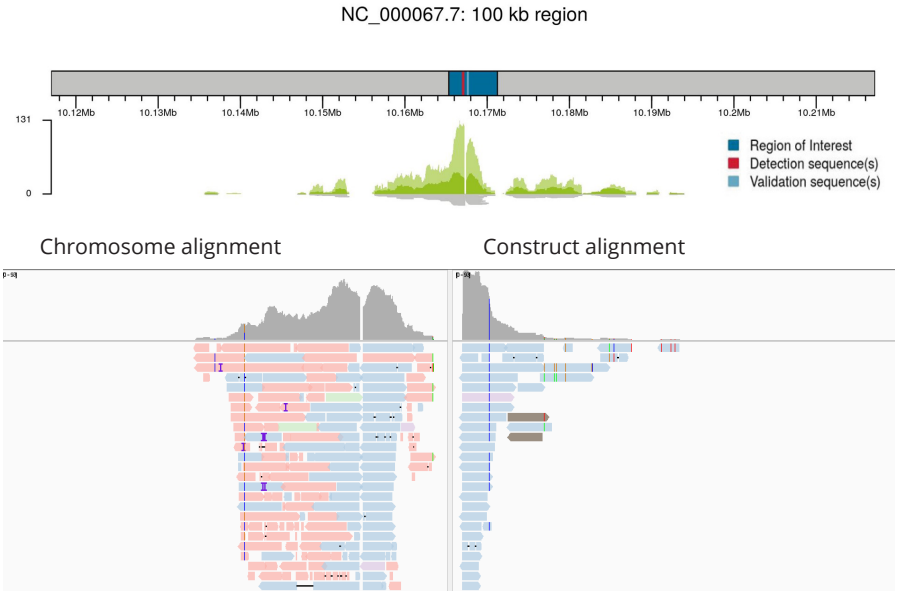

Figure 2. Top: Read alignment in a 100-kb region on Chromosome 1, where the flanking genomic regions bordering the insert are visible along with a 162-bp deletion (the dip in the green coverage diagrams). The vertical scale to the left shows fold coverage. Dark green shows the coverage of primary reads and light green coverage of supplementary reads.

Bottom: A genome browser (IGV[4]) view of individual reads mapping to both the insert (right) and the mouse chromosome (left) are shown in blue. Reads only mapping to the mouse chromosome are shown in pink. Other read colors indicate a few reads that also have homology to regions elsewhere in the genome.

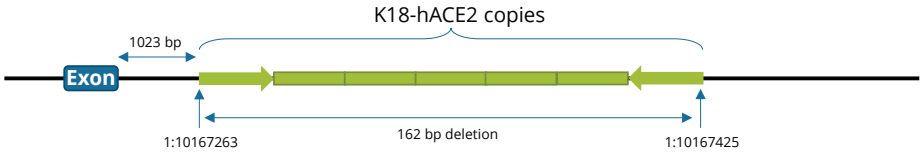

Figure 3. The intronic integration site in the gene *Cspp1* with the presence of a 162-bp deletion. Multiple copies were indicated with inverted copies flanking the insertion.

How Xdrop Sort works

Xdrop Sort partitions DNA fragments into millions of double-emulsion (DE) droplets, then isolates the droplets with DNA from the region of interest. The DE droplets containing the region of interest are identified based on a 120- to 160-bp amplicon specific to that region (the detection sequence). The isolated long DNA fragments are finally amplified via multiple displacement amplification in droplets (dMDA). This ensures unbiased DNA amplification. The DNA fragments generated with Xdrop Sort enrichment and amplification technology are compatible with both long-read and short-read library preparation and sequencing.

Learn more about Xdrop at samplix.com or contact us at samplix.com/contact.

References and notes

1. McCray, P.B. et al. 2007. Lethal infection of K18-hACE2 mice infected with severe acute respiratory syndrome coronavirus. J. Virol. 81(2): 813–821. doi: 10.1128/JVI.02012-06
2. Blondal, T. et al. 2021. Verification of CRISPR editing and finding transgenic inserts by Xdrop indirect sequence capture followed by short- and long-read sequencing. Methods 191: 68-77. doi: 10.1016/j.ymeth.2021.02.003
3. Li, H. 2018. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics 34(18): 3094-3100
4. Robinson, J.T. et al. 2011. Integrative genomics viewer. Nature Biotechnology 29: 24–26