

Controlled Integration Site Mapping by NGS-Based Targeted Locus Amplification and Regular PCR

Katharina Koether, Isabel Proneth, Janet Köhler, Jadranka Koehn

Rentschler Biopharma SE, Laupheim Germany, katharina.koether@rentschler-biopharma.com

Introductory Abstract

Advantage of the TurboCell™ system is the single targeted integration of the Gene of Interest (GoI). Integration reveals in stable expression as shown by long-term cultivation studies up to 100 days using the MCB and Producer TurboCells™ (PTC). However, targeted integration cannot exclude clones with additional randomly integrated GoI copies or even the recombinase gene. Therefore, Top10 recombinant clones were genetically characterized using Southern and Northern Blot resulting in some projects with up to 4 clones with additional integrations. This highlights the necessity for an earlier screening step, such that clones with unintended integrations can be excluded before the Top10 selection. Due to the fact that Southern Blot and Northern Blot are comparably laborious they are both not suitable for a routine use during early clone screening.

The development of new PCR-based screening methods needs the knowledge of all intragenomic copy numbers and the exact loci where the targeting cassette integrates (breakpoints). While copy numbers of the MCB are accessible (2 copies of the tagging vector) and the generated PTCs are known (1 copy of the gene of interest (GoI)), breakpoints have to be identified. Therefore, inverse PCR (iPCR) was successfully used to identify the 5'-integration site (BP1) and confirmed using classical PCR. Application of Next Generation Sequencing (NGS)-based targeted locus amplification (TLA)¹ confirmed BP1 and identified the 3'-site (BP2) of the integration locus. Additionally, the results highlighted structural changes around the insertion site. About 70 kb were lost from the original sequence displacing BP2 downstream the originally expected locus.

By knowing the exact site of the integrated tagging vector an integration site PCR (IS-PCR) was developed in combination with a Multiplex TaqMan based qPCR assay to determine the gene copy numbers. This Multiplex TaqMan assay was also used for measuring the mRNA expression of the respective gene. These assays in combination are substantially faster and need less genetic material, which makes them suitable for earlier use during clone selection.

1. Identification of Breakpoints and Fusion Sites

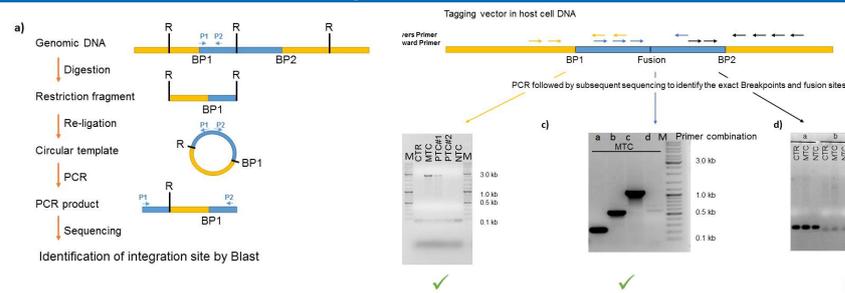


Figure 1: (a) Inverse PCR (iPCR) was used to determine the BPs. The genomic DNA was initially digested using a restriction enzyme (R) known to uniquely cut within the tagging-vector. Subsequently, the DNA was re-ligated and analyzed by a forward and reverse primer, which bind to the known part of the circular DNA for amplifying the unknown genomic DNA. The PCR product was finally sequenced and analyzed using BLAST. This led to the identification of the genomic region of the first breakpoint. (BP1) (b) classical PCR was used to confirm the exact BP1 locus of the 5'-integration. (c) PCR amplification of the region between the two known integrated copies showed, that they have fused. (d) classical PCR using primer walking along the genome could not verify the BP for the 3'-integration site. Neither could iPCR or long-range PCR approaches (data not shown).

3. Replacing Southern Blot by IS-PCR and qPCR

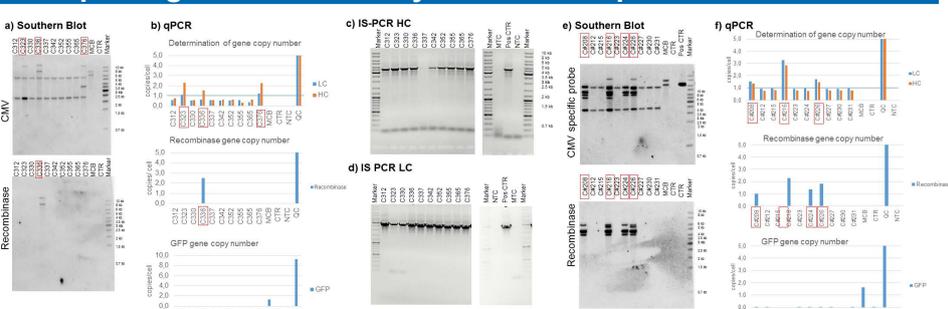


Figure 3: Based on the integration site data results the new PCR and qPCR based techniques for genetic characterization were compared to a conventional Southern Blot. (a) Southern Blot analysis using a CMV or recombinase specific probe. Three clones from project A were identified with additional signals for the CMV promoter and one clone with an integrated recombinase. (b) Gene copy numbers for the GoIs (LC/HC), Recombinase and GFP were determined by absolute qPCR. As for the Southern Blot three clones showed additional copies of the GoI and one clone showed the integration of the recombinase. Except for the MCB none of the clones were positive for GFP. (c, d) Integration specific PCR (IS-PCR) for (c) LC and (d) HC for all clones showed the successful integration into the described locus. Only one clone does not show the expected integration for the HC, repeating the experiment showed that this was due to a technical error.

The comparison between qPCR and Southern Blot was also done for Project B. (e) Southern Blot analysis identified four clones with additional signals for the CMV promoter. The same clones also showed signals for the integration of the recombinase. (f) qPCR analysis of the GoI, the recombinase and GFP confirmed the recombinase integration for all four clones, but showed that only three of the four clones also had additional copies of the GoI. None of the tested clones was positive for GFP.

Summary and Conclusions

- The integration locus of the TurboCell™ was successfully sequenced using a combination of PCR (classical and inverse), sequencing and NGS-based TLA¹
- The TurboCell™ contains two copies of the tagging vector, which are directly fused to each other.
- Downstream of the integration site genetic rearrangements can be observed (deletion of 70 kb of genomic sequence) and BP2 is localized in a highly repetitive genomic area, which explains the difficulties in sequencing the BP2.
- qPCR-based gene copy determination in combination with integration site-specific PCR was shown to be able to successfully replace southern blotting.
- qRT-PCR-based analysis of mRNA expression of the GoI is sufficient for the genetic characterization of PTCs. mRNA expression between the clones is very similar and additional integrations are excluded based on the IS-PCR and qPCR data.

Outlook

- Development of direct PCR assays for IS-PCR, qPCR and qRT-PCR assays for a successful screening at earlier stages of the CLD process (before Top 10 decision)

2. NGS to determine the IS

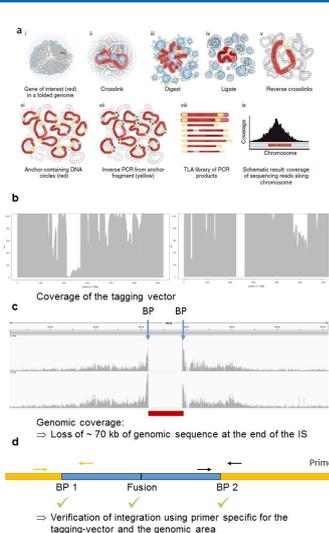


Figure 2: NGS-based targeted locus amplification (TLA)¹ was used to identify unknown regions around the integration sites (IS). (a) TLA is an extended iPCR that crosslinks various parts of the genome to the integrated vector. The region around the IS shows enriched readings due to the crosslinking of regions in spatial proximity. Genomic DNA digestion was followed by circularization. Subsequently, the crosslinking was reversed. Two primer sets specific for the integrated vector were used for PCR amplification and library generation. NGS analysis shows (b) coverage of the integrated vector as well as the (c) surrounding genomic sequence.

The genomic sequence confirmed the known BP1 for the 5'-integration site and the fusion of the two vector copies.

(b) coverage of the integrated vector as well as the (c) surrounding genomic sequence. The genomic sequence confirmed the known BP1 for the 5'-integration site and the fusion of the two vector copies.

This analysis also showed a 70 kb deletion of the genomic DNA at the BP2 for the 3'-integration site. The highly repetitive sequence of the BP2 explains why the iPCR was not successful.

(d) Finally the IS was fully sequenced using an approach of PCR, sequencing and TLA.

4. mRNA Detection - Northern Blot vs. qRT-PCR

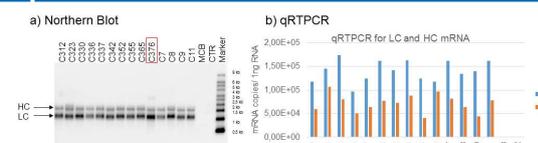


Figure 4: mRNA analysis for selected clones from project A was performed by (a) Northern Blot and (b) qRT-PCR. The Northern Blot showed clear bands for LC and HC also reflecting on the stronger expression of the LC. For C376 an additional band could be observed for one of the clones with additional genetic insertions. This signal was attributed to a weak expressing additional locus. qRT-PCR also showed the different expression for LC and HC, and confirmed the stronger expression of the LC.

References

- de Vree PJP, et al.; Targeted sequencing by proximity ligation for comprehensive variant detection and local haplotyping; Nature Biotechnology 2014;