

Tailor-made Solutions Exemplified with the High-throughput 5' RACE Kit



Patrick J. Biggs

Patrick J. Biggs^{*1}, Rolf Reichhuber², Yalem Z. Bekele¹, Wendy J. Mustill¹, Elliott S. I. Harrison¹, Markus Anton², William C. Skarnes¹, and Thomas Metzler²

¹Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambs, UK

²Roche Diagnostics GmbH, Roche Applied Science, Penzberg, Germany

*Corresponding author: pjb@sanger.ac.uk

Rapid amplification of 5' cDNA ends (5' RACE), a PCR-based method designed to characterize the 5' ends of mRNAs [1], has been adapted for the identification of endogenous loci associated with

insertions of gene trap vectors [2]. Direct sequencing of 5' RACE products [3] has been used by several large-scale gene trapping efforts to generate a library of insertional mutations in mouse embryonic stem (ES) cells [4]. In close collaboration with scientists at Roche Diagnostics, we have optimized 5' RACE and developed a kit for high-throughput characterization of gene trap ES cell lines. The tailor-made HTS 5' RACE kit* provides a significant improvement in the quality and length of sequence tags generated by 5' RACE, and consequently, more of the sequences can be annotated to the mouse genome.

Introduction

With the completion of the human and mouse genome sequences, attention is now focusing on the functional analysis of the ~25,000 genes in the mammalian genome [5, 6]. Gene trapping in mouse embryonic stem (ES) cells provides a cost-effective approach to creating insertional mutations in a substantial fraction of genes in mice [7]. Gene trap vectors contain a splice acceptor site upstream of a promoterless β geo (β -galactosidase/neomycin phosphotransferase fusion) reporter gene that is activated following insertions in genes expressed in ES cells. To identify the endogenous gene associated with each gene trap insertion, 5' RACE is used to amplify the cDNA sequence upstream of the β geo reporter gene followed by automated fluorescent sequencing to generate a short sequence tag for each insertion event [3]. Application of this method on a high-throughput scale proved to be problematic. Here we describe the results of a collaboration with Roche Diagnostics to develop a custom high-throughput 5' RACE kit. This Roche HTS 5' RACE kit* simplifies the protocol and dramatically improves the length and quality of sequence tags generated from gene trap cell lines.

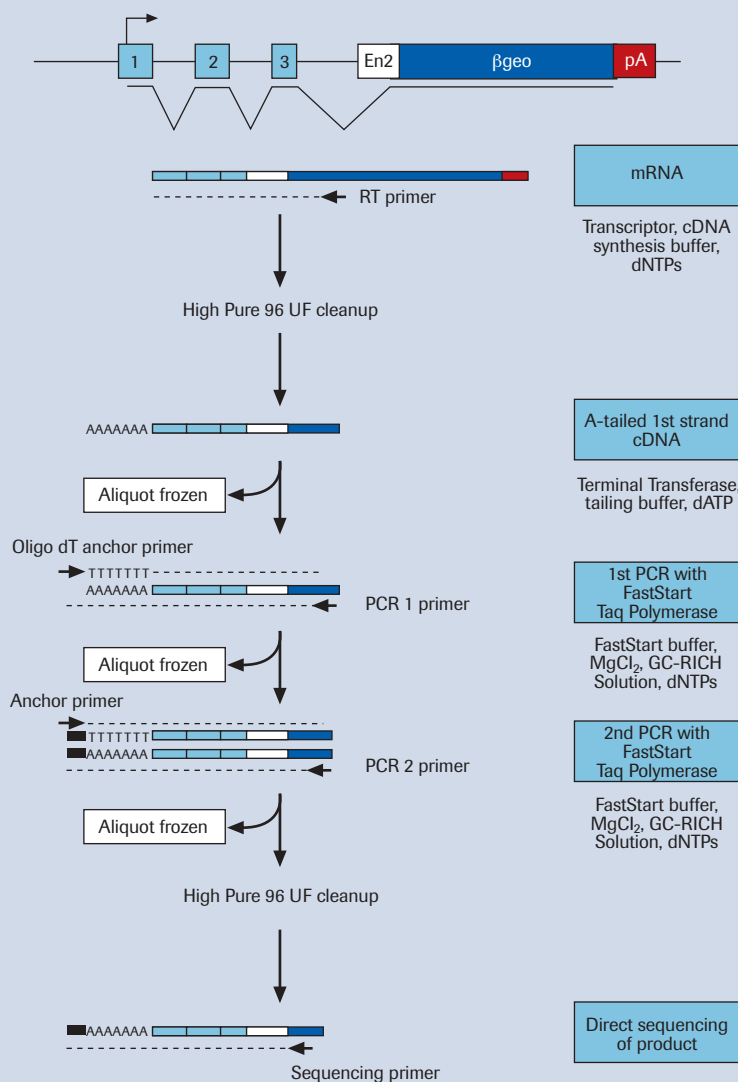


Figure 1: Schematic representation of the new RACE methodology. A detailed description of the method can be found in the text.

*exclusive for the members of the International Gene Trap Consortium (www.genetrapp.org)

Materials and Methods

First, 100 µg of linearized pGT01xr vector was electroporated into 10⁸ ES cells and selected in 200 µg/ml G418 for up to 9 days. Colonies containing random gene trap insertions were picked into 48-well tissue culture plates and when confluent, re-arrayed into 24-well tissue culture plates. At confluence, total RNA was prepared from the ES cells using a commercially available total RNA isolation system together with a Beckman FX robotic liquid handler (Beckman Coulter). The RNA samples were arrayed into a standard 96-well format and all subsequent manipulations were performed in 96-well plates.

The protocol by Townley *et al.* [3] was used as a starting point for optimization, and modifications to the published protocol are shown in Figure 1. The primer sequences used are listed in Table 1. First strand cDNA synthesis was performed in a 20-µl volume using 5 µg (standard deviation of 2.45 µg) of total RNA as the template, 12.5 pmol of RT primer, 1 mM dNTPs, and 25 units of Transcriptor Reverse Transcriptase in the recommended buffer at 55°C for 60 minutes, followed by a 5-minute incubation at 80°C. Single-stranded cDNA was purified with High Pure 96 UF Cleanup Plates by applying a vacuum (-400 to -600 mbar) through a Whatman Univac vacuum manifold for 10–15 minutes until dry, and then applying the vacuum for a further minute. The samples were resuspended in 30 µl PCR-grade water on a microplate shaker for 5 minutes with moderate shaking (700 rpm). Then, 19 µl of the samples were removed and placed into a new 96-well plate.

Prior to the cDNA tailing reaction, the samples were denatured at 94°C for 3 minutes, and then placed on ice. Eighty units of recombinant Terminal Transferase, 200 µM dATP and the recommended buffer were mixed in a volume of 6 µl and immediately added to the denatured cDNA, mixed, and incubated at 37°C for 20 minutes, followed by 10 minutes at 70°C. For the first round of PCR, 5 µl of the reaction was used and the remaining DNA was stored at -20°C.

Two rounds of nested PCR were performed, each using 1 unit of Fast Start Taq DNA Polymerase in a total volume of 25 µl. For the first round, samples were denatured at 95°C for 5 minutes and then subjected to 40 cycles of denaturation (95°C for 30 seconds), annealing (55°C for 30 seconds) and extension (72°C for 2 minutes) using 18.75 pmol of oligo dT anchor primer, 6.25 pmol of PCR 1 primer and dNTPs to a final concentration of 200 µM in the appropriate buffer with MgCl₂ and GC-RICH solution. For the second PCR step, 1 µl of the first round PCR product was taken and the remaining DNA was stored at -20°C. For the second round, samples were

amplified as described above at an annealing temperature of 60°C using 6.25 pmol of the anchor and PCR 2 primers.

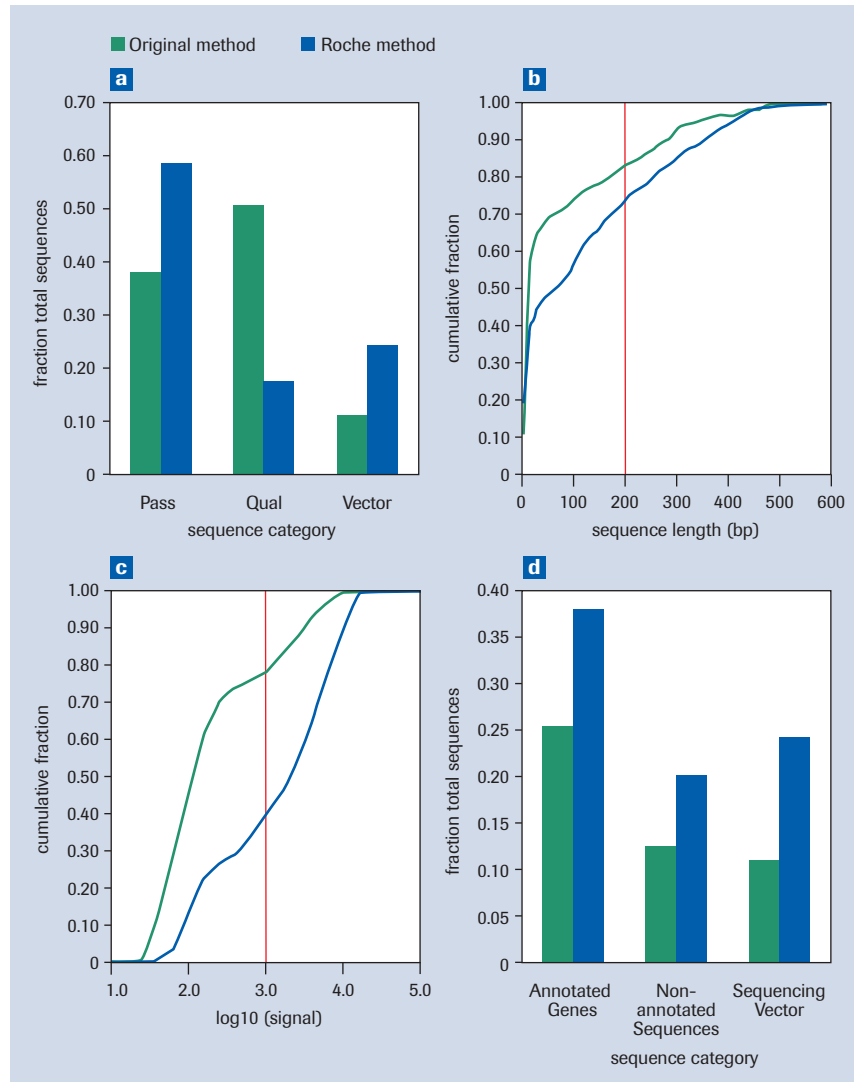


Figure 2: Summary of sequencing results based on 576 sequences comparing the data from the new kit with that from our old method. Each panel represents each of the four criteria used, be it sequence quality (a), sequence length (b), average intensity of the signal (c), or annotation (d).

Table 1: Primer sequences used in the HTS 5' RACE kit

Step	Primer name	Sequence (5' to 3')
First strand cDNA synthesis	RT primer	TAATGGGATAGGTCACGT
First-round PCR	oligo dT anchor primer	GGTTGTGAGCTCTTCTAGATGGTTTTTTTTTTTTTTTTTTT
First-round PCR	PCR 1 primer	AGTATCGGCCTCAGGAAGATCG
Second-round PCR	anchor primer	GGTTGTGAGCTCTTCTAGATGG
Second-round PCR	PCR 2 primer	PCRATTCAGGCTGCGCAACTGTTGGG
Sequencing	sequencing primer	CGACGGGATCCTCTAGAGT

PCR products were then dephosphorylated by adding 1 unit of recombinant alkaline phosphatase to the reaction and incubating at 37°C for 60 minutes, followed by 15 minutes at 70°C. These products were then purified with High Pure 96 UF Cleanup Plates as described above. Subsequently, 2 µl of this reaction was used as the template for a 10-µl cycle sequencing reaction (initial denaturation at 96°C for 30 seconds followed by 45 cycles of denaturation [92°C for 15 seconds], annealing [52°C for 15 seconds] and extension [60°C for 2 minutes]) in buffer containing 1% glycerol, 20 pmol of the sequencing primer, and 0.5 µl BigDye v3.1 (ABI). Following ethanol precipitation, RACE products were run on an ABI 3730 capillary sequencer and the sequences were analysed and annotated to the Ensembl gene build of the NCBI m33 assembly (www.ensembl.org/Mus_musculus/) using Wellcome Trust Sanger Institute software (ASP – Automated Sequence Preprocessing and MapTag, our custom annotation software).

Results and Applications

The protocol by Townley *et al.* [3] is a robust method that has been used successfully to characterize many hundreds of gene trap cell lines. Approximately 60% of gene trap cell lines produce high quality sequence tags using this method ([3] and unpublished results). However, multiple size selection steps carried out on individual microdialysis filters are not well-suited for high-throughput

5' RACE in a 96-well format. Our attempts to substitute microdialysis for other methods such as size exclusion columns have increased the throughput of the method, but at a lower efficiency: only 20–40% of gene trap cell lines produced high-quality sequence tags. Therefore, in close collaboration with Roche Diagnostics, we set out to develop a method that could be carried out in 96-well plates without a reduction in overall efficiency.

For the evaluation of the HTS 5' RACE kit*, 576 RNA samples were used and reactions were compared with reactions using our method based on the protocol by Townley *et al.* [3]. Four criteria – quality, length, signal intensity, and annotation – were used for our analysis of the sequence tags generated by 5' RACE and a marked improvement was observed for all criteria using the HTS 5' RACE kit* (Figure 2). The automated sequence analysis software (ASP) designates high-quality sequences as “Pass”, vector-derived sequences as “Vector”, and poor-quality sequences as “Qual”. Using the HTS 5' RACE kit*, 58% of the reactions produced high-quality sequence tags and fewer sequences failed (18%) compared with our original protocol (38% and 51%, respectively). Length analyses were performed on sequences following automated vector trimming where the base quality score using Phred was

*exclusive for the members of the International Gene Trap Consortium (www.genetrapp.org)

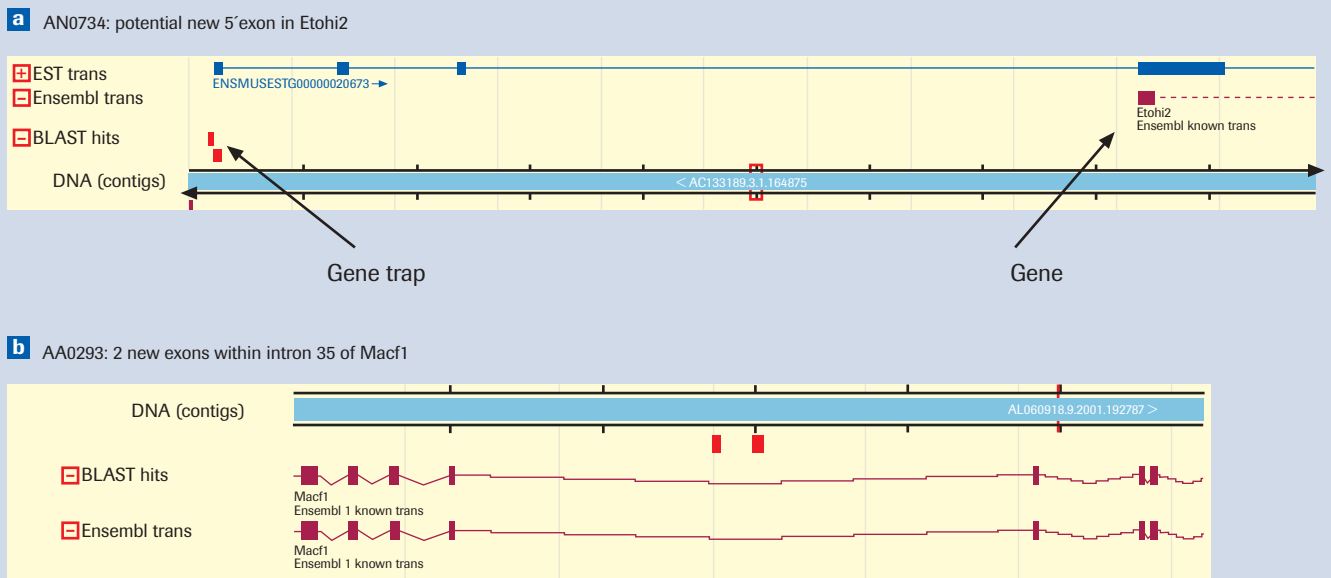


Figure 3: The use of the HTS 5' RACE kit in genomic annotation. RACE product sequences are shown in *red* as “Blast hits” and genes are shown in *dark red* as “Ensembl trans” in these views from Ensembl. The sequence data from some cell lines shows previously unknown exons in genes, either at the 5' end (*a*) or within (*b*).

above 15. The average sequence length increased from 77 base pairs (bp) to 123 bp with the new method. Sequence lengths were also plotted as the cumulative fraction of the sequence length distribution (Figure 2b) and more of the resulting sequences are longer, for example, 27% of the sequences with the new protocol exceed 200 bp compared with 17% using the old method. On average, signal strengths (the average signal detected for a given nucleotide) were four times higher with the new method (3516 units) than with the original method (878 units). The signal intensity, plotted as a cumulative fraction against the log 10 of the signal (Figure 2c), shows that the new method outperforms the original method as 60% of the sequences have an intensity of more than 1000 units compared with only 22% using the original protocol.

The improvement in sequence quality, length and signal resulted in a significant improvement in the annotation of these tags. We obtained a 50% increase in annotation to known Ensembl genes, or RefSeq cDNAs (Figure 2d) and a 60% increase in the number of high quality sequences that match the unannotated sequences in the genome. These "novel" sequences are likely to represent insertions in novel transcribed genes or to exons of known genes that are not yet annotated such as the 5' exon-most exons of genes or alternatively-spliced exons (Figure 3). Thus, the HTS 5' RACE kit* will substantially facilitate gene annotation.

Summary

The Roche HTS 5' RACE kit* has been developed for high-throughput 5' RACE in a 96-well format without a loss in overall efficiency. This kit was applied to the characterization of gene trap insertions in mouse ES cell lines and produced high quality sequence tags for 60% of gene trap cell lines, comparable to the previously published, lower-throughput method. The HTS 5' RACE kit* will greatly aid large-scale gene trapping efforts currently underway to generate insertional mutations in most genes in mouse ES cells. Equally, this method can be applied on a genome-wide scale to help characterize all transcribed exons that comprise the mammalian transcriptome. ■

References

1. Frohman MA, Dush MK, Martin GR (1988) Proc Natl Acad Sci USA 85:8998-9002
2. Skarnes WC, Auerbach BA, Joyner AL (1992) Genes Dev 6:903-918
3. Townley DJ *et al.* (1997) Genome Res 7:293-298
4. BayGenomics: baygenomics.ucsf.edu
SIGTR: www.sanger.ac.uk/genetrap
GGTC: tikus.gsf.de/
5. Auwerx J *et al.* (2004) Nat Genet 36:925-927
6. Austin CP *et al.* (2004) Nat Genet 36:921-924
7. Skarnes WC *et al.* (2004) Nat Genet 36:543-544

Industrial Services: Enjoy the complete solution!

A variety of thousands of raw materials, intermediates and consumables combined with the expertise in raw material and kit manufacturing enables Roche Applied Science to create an almost unlimited number of tailor-made solutions. Roche Applied Science Industrial Services stands ready to help you design a kit to match your high-throughput application.

Instead of using a potpourri of reagents from various sources, benefit from a dedicated design that gives consistent performance and saves time and costs. Concentrate your precious resources on research and leave the repetitive tasks to us. If you are interested, contact your local representative and markus.anton@roche.com to discuss your individual needs. We are looking forward to delivering your customized high-throughput solution!



Product	Pack Size	Cat. No.
FastStart Taq DNA Polymerase	100 U	12 032 902 001
	2 x 250 U	12 032 929 001
	4 x 250 U	12 032 937 001
	10 x 250 U	12 032 945 001
	20 x 250 U	12 032 953 001
Transcriptor Reverse Transcriptase	250 U (25 reactions)	03 531 317 001
	500 U (50 reactions)	03 531 295 001
	2,000 U (4 x 500 U) (200 reactions)	03 531 287 001
Terminal Transferase, recombinant	8,000 U (for 20 tailing or 3'-end labeling reactions)	03 333 566 001
	24,000 U (for 60 tailing or 3'-end labeling reactions)	03 333 574 001
	4 x 25 µmol (4 x 250 µl)	11 969 064 001
Deoxynucleoside Triphosphate Set, PCR Grade	4 x 125 µmol (4 x 1,250 µl)	03 622 614 001
	Water, PCR Grade	25 ml (25 vials of 1 ml)
25 ml (1 vial of 25 ml)		03 315 959 001
100 ml (4 vials of 25 ml)		03 315 959 001
High Pure 96 UF Cleanup Plates (10 x 96 purifications)		04 422 716 001
Geneticin (G418)		11 464 990 001
Alkaline Phosphatase, bulk**		04 571 550 103

* exclusive for the members of the International Gene Trap Consortium (www.genetrap.org)

** available in large quantities only, please inquire