A short report on optimization of nucleic acid probes by DNA microarray synthesis and next generation sequencing

Rachel Yuan Nong Mikkelsen and Tarjei Sigurd Mikkelsen

Department of Immunology, Genetics and Pathology and SciLifeLab, Uppsala University, Uppsala Sweden
Broad Institute of Harvard and MIT, Cambridge, Boston, USA

Correspondence: rachel.nong@gmail.com, tarjei.mikkelsen@gmail.com

Abstract/Introduction

Recent development in oligonucleotide synthesis and DNA sequencing technologies have facilitated advancements for application of the human genome sequencing project into the clinics and routine research thus advance human health and our understanding of human biology. Examples are exome sequencing (1), CRISPR genome editing (2) and MPRA (3,4). Aside from analysis of liquid biopsies, sample-derived cell lines, tissue imaging and investigation of molecular events in situ (5, 6, 7) also provides a means to retrieve biological information however argument on specificity from the secondary labels. Present study built on previous work on utilizing several molecular biology tools, such as restriction enzymes (9) and rolling circle amplification (RCA) for enhancing signal-to-noise ratio in a detection assay (8) – a novel probe-type 2fold probe is developed to generate a single strand circular reporter molecule upon target recognition and multi-step enzymatic conversion. We specifically investigated probe performance by varying a 9-nt sequence within the stem structure at single nucleotide resolution and in library fashion.

Result and Discussion

2fold probe

A 2fold probe consists of target binding sequence and reporter action sequence that anchors each probe to target by dual hybridizations (for nucleic acid target) or bindings (for protein target) followed by sequential washings and enzymatic treatments, 1) MlyI cleaves the single probe to two parts; 2) UNG digests base 'dU' so degrades protection oligo and availables one end of circle reporter; 3) Nb.BtsI acts upon one strand at stem and availables the other end of circle reporter, 4) T4 DNA ligase ligates the open circle templated from strand that anchored by one target binding sequence but remained locally via hybridization to the other target binding sequence (Figure4). Newly formed single strand circle reporter molecule is then amplified by RCA, detected by conventional detection schemes – hybridization by fluorescent labeled oligo, or in situ sequencing methods, and enumerated from a microscope image or in the present study, amplified by PCR and analyzed by next generation sequencing (RCA-seq).

2fold probe library and optimization assay

We computationally designed a 4000-variant probe library, including MlyI-Nb.BtsI and two matched control-sets MlyI-Nb.BbvCI and MlyI-NoNick probe-types, and synthesized these probes following previously developed protocols. The optimization assay was then performed with a solid phase magnetic beads protocol that utilized a synthetic DNA molecule as target. We found that the RCA-seq protocol is reproducible (Figure1) and a striking 'T' bias at first base next to the nicking enzyme binding site (Figure2). In agreement with most other studies probes with balanced G+C/A+T content tend to give higher signals in a one-step assay (Figure3).

Figure1 RCA-seq assay reproducibility were accessed with scatter-plots of two replicated experiments for each probe type. Correlation could be observed in result by probes contained correct enzyme binding sequences (left), less obvious when probes contained unintended enzyme binding site (middle) or no matched enzyme binding sequence (right).

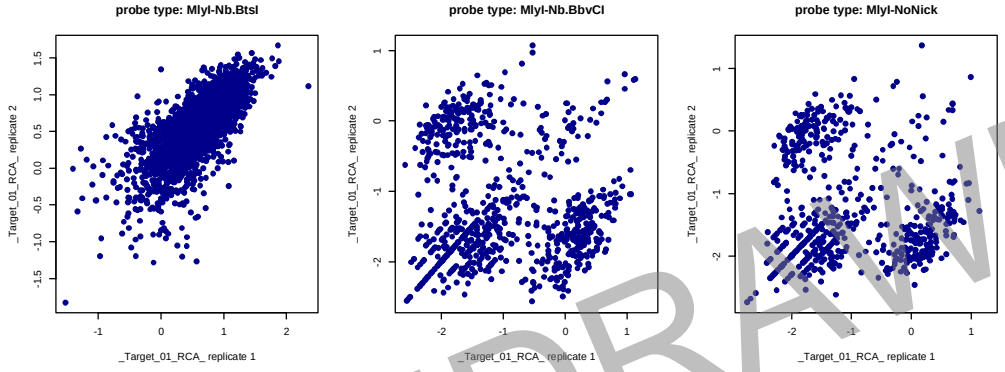

Figure2 Investigation of 9-nt sequence next to the Nb.BtsI binding site (5’ -> 3’) in a 2fold probe was presented by heatmap-plot in experiment conditions with or without protection oligo (denoted as _addProt and _noProt respectively) having target presented versus target absent (denoted as Target and ProbeOnly respectively) followed by RCA or directly to PCR (denoted as _RCA and _noRCA respectively) for thus circle reporters.

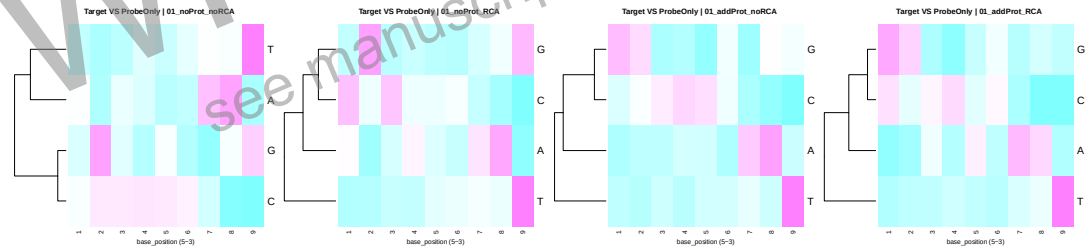

Figure3 Investigation of stem-structure sequence (denoted as carrier) composition in a 2fold probe was presented by boxplot of results from experiments with Nb.BtsI (left) and Nb.BbvCI (right) in parallel following a one-step solid phase protocol.

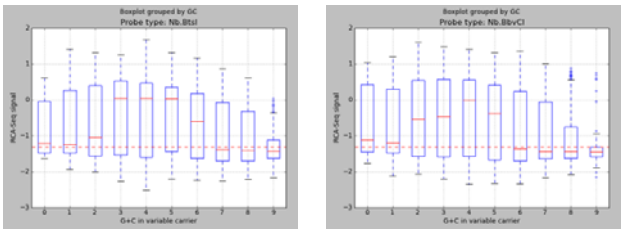

Figure4 Illustration of library probe design scheme: Target binding sequence (color brown), 3’-end of circle reporter sequence (color green), 5’-end of circle reporter sequence, MlyI binding sequence (color orange), Stem-structure sequence (color blue), circle reporter RCA-blob detection sequence (color purple), enzyme binding sequences (text underline) and connected bases (color black).

MlyI-Nb.BtsI
CGACGACGAGCGCGGAAAAGACAGGCAAAGCGAGCAGGAAACAAGGAAGAGTCAAAAATCGCTTTGCCTGTCTYYYYYYYYYGCAGTGGCGGCTCGTTTGCG
GTTCTGAATTCCTTGTTTCCTGCCACTGCXXXXXXXXXGAACGCGGGGCGGACGCG

MlyI-Nb.BbvCI
CGACGACGAGCGCGGAAAAGACAGGCAAAGCGAGCAGGAAACAAGGAAGAGTCAAAAATCGCTTTGCCTGTCTYYYYYYYYYGCCTCAGCGGCTCGTTTGCG
GTTCTGAATTCCTTGTTTCCTGCTGAGGCXXXXXXXXXGAACGCGGGGCGGACGCG

MlyI-NoNick
CGACGACGAGCGCGGAAAAGACAGGCAAAGCGAGCAGGAAACAAGGAAGAGTCAAAAATCGCTTTGCCTGTCTYYYYYYYYYGCCGTAGCGGCTCGTTTGCG
GTTCTGAATTCCTTGTTTCCTGCTACGGCXXXXXXXXXGAACGCGGGGCGGACGCG

Materials and methods

DNA oligonucleotide sequences: 2fold library probes were prepared by facility at the Broad Institute (130106_UUProbe_carrier_fullTileAmp_rev.dat), other sequences were purchased from Integrated DNA Technologies (supplementary table).

Solid phase assay: (single probe assay, steps adjusted to library probe assay would be noted) 25 amol of capture probes were immobilized of 1e7 MyOne T1 beads (Life technologies 65602) for each 25 microliter reaction following manufactural instruction and blocked by blocking buffer (1 mg/ml BSA (NEB), 1 mM biotin (Life Technologies B-1595), 0.1 µg/µl salmon sperm DNA (Life Technologies 15632-011) in 1x PBS) at room temperature (RT) for 30 min on an end-to-end table rotator, collected and stored at 4˚C in storage buffer (1x PBS and 0.1% BSA (Sigma-Aldrich)) after 2x washing by washing buffer (1x PBS, 0.05% Tween20 (Sigma-Aldrich P9416)) at RT. Assay was started by probe preparation that 0,2 µM probes and 20 µM protection oligo were incubated in 10 µl 1x NEB4 buffer and 0.5 µg/µl BSA for 30-60 min at 37˚C and diluted 1:20 in blocking buffer. 25 µl of probes mix was adding to 1e7 (or 25e7 for library probe assay) beads and incubate at RT for 60 min. A multi-step enzymatic treatment were then carried out on beads in 25 µl of: 1) 0,5 U/µl MlyI (NEB R0610L), 5 µM protection oligo, 0.5 µg/µl BSA, 1x NEB4 buffer by incubation at 37˚C for 30 min; 2) 0,5 U/µl Nb.BtsI (NEB R0707L) or Nb.BbvCI (NEB R0631L) or ddH2O, 0.5 µg/µl BSA, 1x NEB4 buffer by incubation at 37˚C for 30 min; 3) 0,05 U/µl UNG (Thermo Fisher Scientific EN0362), 0.5 µg/µl BSA, 1x NEB4 buffer by incubation at 37˚C for 30 min; 4) 0,3 U/µl T4 DNA ligase (Thermo Fisher Scientific EL0013), 0,5 mM ATP (Thermo Scientific R0441), 0.5 µg/µl BSA, 1x NEB4 buffer by incubation at 37˚C for 30 min; 5) 0,5 U/µl phi29 DNA polymerase (Thermo Fisher Scientific EP0094 for single probe assay, NEB for library probe assay), 0.25 mM dNTP (Thermo Fisher Scientific R0182), 1x phi29 buffer by incubation at 37˚C for 30 min; with 2x washing by washing buffer at RT. Sequencing library preparation were performed with a NEB High Fidelity DNA polymerase (NEB) and high concentration PCR primers. Resulted products were sequenced at a HiSeq2000 sequencer with its rapid run program.

Data analysis was performed in R version 3.6.2 (Figure4 was plotted in a Python environment with matplotlib). Counts for each probe sequence and each experimental condition was extracted from original HiSeq2000 fastq files, followed by normalization:
normalizing_factor = each_probe_count / sum(each_probe_count) and normalized_data = each_probe_assay_read * normalizing_factor. Mean of each triplicate were then calculated and log2 transformed, of which median value for one of the four DNA bases at each of the nine nucleotide positions was computed prior reading into heatmap to generate heatmap plots (figure2). RCA-seq signal for each probe = log10(# RCA count / # control count) for figure1.

Acknowledgements

The authors would like to thank discussion and support from all staffs at the Landegren lab and the Nilsson lab in Sweden since 2005 and the Broad Technology Labs at the Broad Institute at Cambridge over the years and at many exciting cross-continent collaborations. This work has been supported by

NIH grant R21CA126727 to the Landegren lab and R01HG006785 to Tarjei S Mikkelsen, and a VR international postdoc grant VR637-2013-7373 to Rachel YN Mikkelsen.

Reference

1. Johansson, H., Isaksson, M., Sörqvist, E.F., Roos, F., Stenberg, J., Sjöblom, T., Botling, J., Micke, P., Edlund, K., Fredriksson, S., et al. (2011) Targeted resequencing of candidate genes using selector probes. *Nucleic Acids Res.*, **39**.

2. Wang, T., Wei, J.J., Sabatini, D.M. and Lander, E.S. (2014) Genetic screens in human cells using the CRISPR-Cas9 system. *Science (80-.).*, **343**, 80–84.

3. Melnikov, A., Murugan, A., Zhang, X., Tesileanu, T., Wang, L., Rogov, P., Feizi, S., Gnirke, A., Callan, C.G., Kinney, J.B., et al. (2012) Systematic dissection and optimization of inducible enhancers in human cells using a massively parallel reporter assay. *Nat. Biotechnol.*, **30**, 271–277.

4. Melnikov, A., Rogov, P., Wang, L., Gnirke, A. and Mikkelsen, T.S. (2014) Comprehensive mutational scanning of a kinase in vivo reveals substrate-dependent fitness landscapes. *Nucleic Acids Res.*, **42**, 1–8.

5. Söderberg, O., Gullberg, M., Jarvius, M., Ridderstråle, K., Leuchowius, K.-J., Jarvius, J., Wester, K., Hydbring, P., Bahram, F., Larsson, L.-G., et al. (2006) Direct observation of individual endogenous protein complexes in situ by proximity ligation. *Nat. Methods*, **3**, 995–1000.

6. Ke, R., Mignardi, M., Pacureanu, A., Svedlund, J., Botling, J., Wählby, C. and Nilsson, M. (2013) In situ sequencing for RNA analysis in preserved tissue and cells. *Nat. Methods*, **10**, 857–860.

7. Lee, J.H., Daugharthy, E.R., Scheiman, J., Kalhor, R., Amamoto, R., Peters, D.T., Turczyk, B.M. and Marblestone, A.H. (2014) Sequencing in Situ.

8. Jarvius, J., Melin, J., Göransson, J., Stenberg, J., Fredriksson, S., Gonzalez-Rey, C., Bertilsson, S. and Nilsson, M. (2006) Digital quantification using amplified single-molecule detection. *Nat. Methods*, **3**, 725–727.

9. Landegren, U., Kaiser, R., Sanders, J. and Hood, L. (1988) A ligase-mediated gene detection technique. *Science*, **241**, 1077–1080.

Data availability:
GEO accession upload: Counts files
Initial single 2fold probe was designed at the Landegren lab in Uppsala University, and the following experiments were performed at the Broad Institute at the Mikkelsen lab: 12000 2fold probes were synthesized by a B3 Custom Array DNA synthesizer in their reverse complementary oligonuleotide-order. After stripped off from microarrays, synthesized products were pooled, emulsion-PCR amplified, lambda exonuclease and restriction enzymes treated and resulted in single stranded 2fold probes – briefly, the 9 nucleotides positions next to an nicking enzyme binding site at the stem structure within a probe were computationally designed to be variable thus a set of 4000 probe sequences, namely MlyI-Nb.BtsI in parallel with two matched control-sets: MlyI-Nb.BbvCI and MlyI-NoNick. A solid phase (magnetic beads) assay protocol was utilized and end-result products were subsequently analyzed in a HiSeq2000 sequencer whereas variable sequences were read and served as tags to generate a count file. GEO submission 2021-May-03
GEO accession number: GSE173791

Code availability: analyze_2fold.py process_1_2fold_lib.py
Supplementary file: 130106_UUProbe_carrier_fullTileAmp_rev.dat

#bioRxiv.org
#academic.oup.com Nucleic Acids Research Oxford Academic