

Mutation Analysis of Synthetic DNA Barcodes in a Fission Yeast Gene Deletion Library by Sanger Sequencing

Minho Lee¹, Shin-Jung Choi², Sangjo Han³, Miyoung Nam⁴,
Dongsup Kim⁵, Dong-Uk Kim^{2*}, Kwang-Lae Hoe^{4**}

¹Catholic Precision Medicine Research Center, College of Medicine, The Catholic University of Korea, Seoul 06591, Korea,

²Aging Research Center, Korea Research Institute of Bioscience & Biotechnology (KRIBB), Daejeon 34141, Korea,

³Data Analytics CoE, SK Telecom, Seongnam 13595, Korea,

⁴Department of New Drug Development, Chungnam National University, Daejeon 34134, Korea,

⁵Department of Bio and Brain Engineering, Korea Advanced Institute of Science & Technology (KAIST), Daejeon 34141, Korea

Incorporation of unique barcodes into fission yeast gene deletion collections has enabled the identification of gene functions by growth fitness analysis. For fine tuning, it is important to examine barcode sequences, because mutations arise during strain construction. Out of 8,708 barcodes (4,354 strains) covering 88.5% of all 4,919 open reading frames, 7,734 barcodes (88.8%) were validated as high-fidelity to be inserted at the correct positions by Sanger sequencing. Sequence examination of the 7,734 high-fidelity barcodes revealed that 1,039 barcodes (13.4%) deviated from the original design. In total, 1,284 mutations (mutation rate of 16.6%) exist within the 1,039 mutated barcodes, which is comparable to budding yeast (18%). When the type of mutation was considered, substitutions accounted for 845 mutations (10.9%), deletions accounted for 319 mutations (4.1%), and insertions accounted for 121 mutations (1.6%). Peculiarly, the frequency of substitutions (67.6%) was unexpectedly higher than in budding yeast (~28%) and well above the predicted error of Sanger sequencing (~2%), which might have arisen during the solid-phase oligonucleotide synthesis and PCR amplification of the barcodes during strain construction. When the mutation rate was analyzed by position within 20-mer barcodes using the 1,284 mutations from the 7,734 sequenced barcodes, there was no significant difference between up-tags and down-tags at a given position. The mutation frequency at a given position was similar at most positions, ranging from 0.4% (32/7,734) to 1.1% (82/7,734), except at position 1, which was highest (3.1%), as in budding yeast. Together, well-defined barcode sequences, combined with the next-generation sequencing platform, promise to make the fission yeast gene deletion library a powerful tool for understanding gene function.

Keywords: barcode, fission yeast, gene deletion, growth fitness, mutation

Introduction

The fission yeast *Schizosaccharomyces pombe* is a eukaryotic model organism that has become useful in studies of the cell cycle and chromosome dynamics [1]. According to a phylogenetic tree, fission yeast is separated from the other model yeast—the budding yeast *Saccharomyces cerevisiae*—by an estimated 1,000 million years of evolution [2]. Fission yeast has the smallest genome among eukaryotes, with a haploid genome size of 13.8 Mb, covering ~5,000 open

reading frames (ORFs) compared with >5,500 for budding yeast [3]. Like budding yeast, fission yeast is also genetically tractable. Its popularity in studying mammalian cell biology has led to its nickname, “micro-mammal.”

Genomics has benefited from innovative advances in the field of genetic engineering. Especially, genomewide gene deletion allows the qualitative and quantitative functional analysis of genes *en masse*. The budding yeast is the pioneer model organism for gene deletion collections, facilitating feasible parallel growth analyses [4]. Next to the budding yeast gene deletion library, the fission yeast gene deletion

Received May 11, 2018; Accepted May 16, 2018

*Corresponding author: Tel: +82-42-860-4159, Fax: +82-42-860-4149, E-mail: kimdongu@kribb.re.kr

**Corresponding author: Tel: +82-42-821-8627, Fax: +82-42-821-8927, E-mail: kwanghoe@cnu.ac.kr

Copyright © 2018 by the Korea Genome Organization

© It is identical to the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>).

library has been reported by our team [5], supplying powerful resources for cell cycle and cell shape genes [6]. The fission yeast gene deletion library was constructed using a similar strategy to the one used in budding yeast. In brief,

for each strain, the ORF was replaced and tagged by homologous recombination with a deletion cassette. Each correct deletion strain was confirmed by at least one successful colony PCR out of a pair of colony PCRs from both

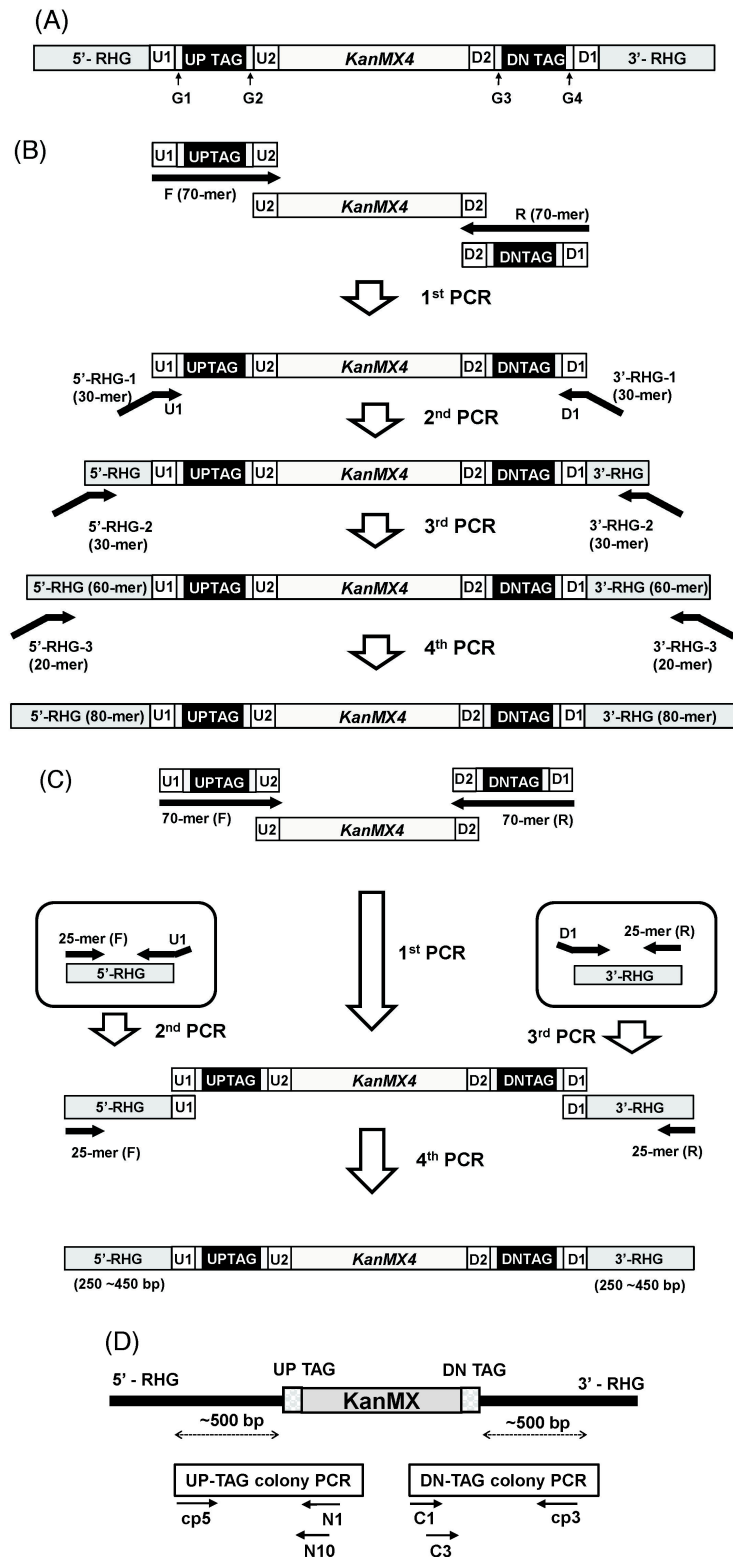


Fig. 1. Construction overview of fission yeast gene deletion strains. (A) Structure of the deletion cassette consisting of the *KanMX4* module and its flanking RHG (Region of Homology to the Gene of interest). The *KanMX4* module is flanked by unique barcodes and RHG on both sides. The deletion cassette replaces the open reading frame of interest by homologous recombination. The length of the RHG is from 80 bp to 250–450 bp, depending on construction method. (B) Serial PCR method. For serial PCR method, 4 rounds of PCR were performed overall. The purpose of the first PCR is for preparation of the *KanMX4* module. Through another 3 rounds of PCR, the RHG of the 80-mer was finally prepared for homologous recombination. (C) Block PCR method. For block PCR method, 4 rounds of PCR were performed overall, as with the serial PCR method. The first PCR is same as the serial PCR method for the preparation of the *KanMX4* module. Two separate PCRs using 51-mer and 25-mer primers and chromosomal DNA as templates were then performed for the preparation of a pair of RHG blocks flanking the *KanMX4* module. Finally, three separated blocks were combined by block PCR. (D) Strategy of colony PCR. For Sanger sequencing, barcode regions were PCR-amplified as indicated. Note that the cp5 and cp3 primers are gene-specific primers. They are around 500 bp apart from the homologous recombination site of each gene deletion cassette. For the sequences of the primers used, refer to “Methods.”

ends of deletion cassettes. In addition, a pair of unique barcode tags was built on each deletion strain to enable the identification of genes that are affected by specific growth conditions without prior knowledge of gene function [7]. For example, the growth fitness of each strain under a chemical treatment can be measured in a highly parallel manner [8].

However, its extensive usage combined with microarray has been hampered by the high rate of mutations in barcode regions that are introduced during strain construction. According to a previous analysis of barcode sequences in budding yeast, at least 18% to 31% of the genome-integrated barcodes contain differences from those originally synthesized, depending on the sequencing method [9, 10]. Also, in a preliminary study, the fission yeast deletion library has been reported to contain ~30% mutations in barcode sequences [11]. These mutations cause problems of anomalous hybridization during microarray, resulting in false positive signals. Supposing only a 1% hybridization error, there exist >50 false-positives among ~5,000 ORF genes in fission yeast. To circumvent the problem caused by barcode mutations, next-generation sequencing (NGS) technology has been applied to replace the microarray method [10]. No internal mutations in barcode regions matter, because the innovative NGS protocol directly counts each barcode in a complex sample via genomic deep sequencing. For adaptation of the NGS technology, a prerequisite is to characterize the entire sequences of all barcodes.

In the study, for the first time, we assessed the prevalence of mutations of all genome-incorporated barcodes in the fission yeast deletion library by Sanger sequencing. The results in this work, combined with the NGS platform, promise to harness the fission yeast gene deletion library for understanding gene function.

Methods

Oligonucleotides, medium, the gene deletion library, and DNA samples

All synthetic DNA oligonucleotides were obtained from Bioneer (Daejeon, Korea). Cultures were grown in YES medium (0.5% yeast extract, 3% glucose, and the appropriate amino acid supplements) at 30°C, following the instructions [12]. Genomic DNA from the fission yeast gene deletion library was extracted using the Quick-DNA Fungal/Bacterial kit (Zymo Research Co., Irvine, CA, USA) and then resuspended in distilled water until use.

The gene deletion library used in this study was constructed based on the principle of homologous recombination, as previously reported [5]. In brief, for each strain, the ORF was replaced and tagged by homologous recombination with a

deletion cassette consisting of the KanMX module [the selectable resistance gene KanMX4 and a pair of unique 20-mer molecular barcodes (up-tag and down-tag) on both sides flanking the KanMX4 gene] and its flanking homologous regions to the chromosome (Fig. 1A). Deletion cassettes were prepared by two different methods: serial or block PCR (Fig. 1B and 1C). The latter one was designed to have longer flanking homologous regions, resulting in a higher recombination rate than former. Among the 4,354 strains used in this study, 2,886 and 1,468 strains were made by the block and serial PCR methods, respectively.

The oligonucleotide sequences used in the construction of the KanMX4 module were as follows: UP-tag (20-mer), unique barcode sequence assigned to each ORF; DN-tag (or DOWN-tag, 20-mer), unique barcode sequence assigned to each ORF; U1 (20-mer), common priming site 1 for amplifying UP-tags, 5'-CGCTCCCGCCTTACTTCGCA-3' (sense); U2 (18-mer), common priming site 2 for amplifying UP-tags and homologous to the 5' region of the KanMX4 module, 5'-TTAGCTTGCCTCGTCCCC-3' (sense); D2 (18-mer), common priming site 2 for amplifying DN-tags and homologous to the 3' region of the KanMX4 module, 5'-TTTCGACACTGGATGGCG-3' (sense); D1 (20-mer), common priming site 1 for amplifying DN-tags, 5'-TTGCGTTGCGTAGGGGGGA T-3' (sense); G1 (6-mer), gap sequence 1, 5'-TTTAAA-3' (sense); G2 (6-mer), gap sequence 2, 5'-GATATC-3' (sense); G3 (6-mer), gap sequence 3, 5'-AGTATC-3' (sense); G4 (6-mer), gap sequence 4, 5'-TTTAAA-3' (sense).

The oligonucleotide sequences used in the serial or block PCR method were as follows: UP-tag forward (70-mer), 5'-U1-G1-(20-mer UP-tag)-G2-U2-3'; 70-mer DN-tag reverse, 5'-D1-G4-(20-mer DN-tag)-G3-D2-3'.

Colony PCR and DNA sequencing

All barcode regions were amplified by colony PCR and subjected to Sanger sequencing. Up-tag or down-tag regions were amplified using a pair of cp5 and N1 primers or C1 and cp3 primers and sequenced using N10 or C3 as a sequencing primer, respectively (Fig. 1D). The gene-specific cp5 or cp3 primer was designed to be located approximately 500 bp from the insertion sites of the deletion cassettes. The oligonucleotide sequences were as follows: N1 5'-CGTC TGTGAGGGGAGCGTTT-3', C1 5'-TGATTTTGTATGACGA GCGTAAT-3', N10 5'-GATGTGAGAACTGTATCCTAGCA AG-3', C3 5'-GGCTGGCCTGTTGAACAAGTCTGGA-3'. Sequencing was performed by SolGent Co. (Daejeon, Korea).

Bioinformatics analysis of barcode sequences

Barcode sequences obtained from sequencing were aligned with the designed using *needle* in the EMBOSS package [13], which implements the Needleman-Wunsch

algorithm [14] of pairwise sequence alignment.

Results

Validation of fidelity on sequenced barcodes

In the study, 8,708 barcodes from 4,354 deletion strains covering 88.5% (4,354/4,919) of all ORF genes (4,919) were amplified by colony PCR and used for the analysis of barcode sequences. At that time, 565 strains (11.5%) were under construction by gene synthesis method [15], and they were not available for barcode analysis. As shown in Table 1, out of the 8,708 barcodes from the 4,354 strains tested, 7,734 barcodes (88.8% in total, block PCR 87.7% and serial PCR 91.0%) were validated to be inserted at the correct positions on both sides—so called “high-fidelity.” The remaining 974 barcodes (11.2%, so called “medium- or low-fidelity”) were characterized only on either side.

Evaluation of defects in barcodes

An examination of the DNA sequences obtained from the 7734 high-fidelity barcodes revealed that 6,695 barcodes (86.6%) had no sequence mutation within them (Table 2). In detail, 2,920 strains (5,840 barcodes, 75.5%) had no mutation on either barcode, and 855 strains (855 barcodes, 11.1%) had mutation(s) only on either barcode. In other words, 1,039 barcodes (13.4%) had mutation(s) in the 20-mer molecular barcodes. It is noteworthy that the mutation rate of barcodes made by block PCR was higher by 4.3% than by serial PCR (14.9% vs. 10.6%). It is likely that the key process in block PCR, block-joining, is complicated and creates mutation(s). Among the 1,039 mutated barcodes, 855 mutations were found on either barcode, and 184 mutations were found on both barcodes.

Within the 1,039 mutated barcodes out of the 7,734

Table 1. Validation of sequenced barcodes

| | No. of strains/barcodes (%) | | | Total |
|-----------------------------|-----------------------------|---------------------|------------------|---------------------|
| | Block-PCR | Serial-PCR | Gene-synthesis | |
| Entire ORF in fission yeast | | | | 4,919/9,838 (100.0) |
| Under strain construction | - | - | 565/1,300 (11.5) | 565/1,300 (11.5) |
| Used for barcode sequencing | 2,886/5,772 (58.7) | 1,468/2,936 (29.8) | - | 4,354/8,708 (88.5) |
| Used for barcode sequencing | 2,886/5,772 (100.0) | 1,468/2,936 (100.0) | - | 4,354/8,708 (100.0) |
| Confirmed on both sides | 2,531/5,062 (87.7) | 1,336/2,672 (91.0) | - | 3,867/7,734 (88.8) |
| Not confirmed on both sides | 355/710 (12.3) | 132/264 (9.0) | - | 487/974 (11.2) |

ORF, open reading frame.

Table 2. Overview of defects in barcodes and mutations

| | No. of defects (%) | | |
|--|--------------------|---------------|---------------|
| | Block-PCR | Serial-PCR | Total |
| Barcodes confirmed by sequencing | 5,062 (100.0) | 2,672 (100.0) | 7,734 (100.0) |
| Barcodes having mutation(s) | 756 (14.9) | 283 (10.6) | 1,039 (13.4) |
| Barcodes without mutation(s) | 4,306 (85.1) | 2,389 (89.4) | 6,695 (86.6) |
| No mutation on both barcodes | 3,696 (73.0) | 2,144 (80.2) | 5,840 (75.5) |
| No mutation on either barcode | 610 (12.1) | 245 (9.2) | 855 (11.1) |
| Barcodes having mutation(s) | 756 (14.9) | 283 (10.6) | 1,039 (13.4) |
| Mutation(s) on either barcode | 610 (12.1) | 245 (9.2) | 855 (11.0) |
| Mutation(s) on both barcodes | 146 (2.8) | 38 (1.4) | 184 (2.4) |
| Mutations within barcode tags ^a | 931 (18.4) | 353 (13.2) | 1,284 (16.6) |
| Mutation(s) in UT | 475 (9.4) | 199 (7.4) | 674 (8.7) |
| Mutation(s) in DT | 456 (9.0) | 154 (5.8) | 610 (7.9) |
| Type of mutations within barcode tags | | | |
| Substitution | 654 (12.9) | 191 (7.1) | 845 (10.9) |
| Deletion | 187 (3.7) | 132 (4.9) | 319 (4.1) |
| Insertion | 90 (1.8) | 31 (1.2) | 121 (1.6) |

UT, up-tag; DT, down-tag.

^aThe number of mutations within the barcodes (1,284) is more than the number of defective barcodes (1,039) because of multiple mutations in a single barcode.

Table 3. Types of defects in 1,039 barcodes having mutation(s)

| | No. of barcodes (%) | | | | | | | | | | | |
|-------------------------------|-----------------------------------|--------------------|----------------------------------|-------------------|-----------------------------------|--------------------|-----------------------------|-------------------|-----------------------------------|--------------------|----------------------------------|-------------------|
| | Block-PCR (n = 756, 72.8%) | | | | Serial-PCR (n = 283, 27.2%) | | | | Total (n = 1,039) | | | |
| | Either tag (n = 610, 58.7%) | | Both tags (n = 146, 14.1%) | | Either tag (n = 245, 23.6%) | | Both tags (n = 38, 3.7%) | | Either tag (n = 855, 82.3%) | | Both tags (n = 184, 17.7%) | |
| | UT (n = 316) | DT (n = 294) | UT (n = 73) | DT (n = 73) | UT (n = 140) | DT (n = 105) | UT (n = 19) | DT (n = 19) | UT (n = 456) | DT (n = 399) | UT (n = 92) | DT (n = 92) |
| Substitution (n = 702, 67.6%) | 233 | 198 | 55 | 55 | 81 | 59 | 10 | 11 | 314 | 257 | 65 | 66 |
| 1 bp | 215 | 176 | 50 | 51 | 75 | 55 | 9 | 9 | 290 | 231 | 59 | 60 |
| 2 bp (in a row) | 3 | 2 | 0 | 1 | 1 | 0 | 0 | 0 | 4 | 2 | 0 | 1 |
| 2 bp (1 + 1) | 11 | 18 | 3 | 3 | 4 | 4 | 1 | 2 | 15 | 22 | 4 | 5 |
| 3 bp (in a row) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 3 bp (1 + 1 + 1) | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 |
| 3 bp (1 + 2) | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| 4 bp (in a row) | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 0 | 0 |
| 4 bp (1 + 1 + 1 + 1) | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| 4 bp (1 + 2 + 1) | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 0 | 0 |
| 4 bp (2 + 2) | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| Deletion (n = 177, 17.0%) | 36 | 48 | 5 | 11 | 35 | 30 | 7 | 5 | 71 | 78 | 12 | 16 |
| 1 bp | 31 | 41 | 5 | 9 | 29 | 22 | 5 | 4 | 60 | 63 | 10 | 13 |
| 2 bp (in a row) | 3 | 4 | 0 | 1 | 0 | 3 | 1 | 0 | 3 | 7 | 1 | 1 |
| 2 bp (1 + 1) | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| 3 bp (in a row) | 1 | 1 | 0 | 0 | 5 | 3 | 1 | 1 | 6 | 4 | 1 | 1 |
| 3 bp (1 + 1 + 1) | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| 4 bp (in a row) | 0 | 1 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 3 | 0 | 0 |
| 5 bp (in a row) | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| 6 bp (in a row) | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| Insertion (n = 79, 7.6%) | 19 | 28 | 6 | 4 | 9 | 10 | 0 | 3 | 28 | 38 | 6 | 7 |
| 1 bp | 18 | 27 | 6 | 4 | 9 | 10 | 0 | 3 | 27 | 37 | 6 | 7 |
| 2 bp (in a row) | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 |
| Complex (n = 55, 5.3%) | 17 | 14 | 6 | 2 | 11 | 3 | 2 | 0 | 28 | 17 | 8 | 2 |
| S1 + I1 | 6 | 4 | 2 | 1 | 2 | 2 | 0 | 0 | 8 | 6 | 2 | 1 |
| S1 + D1 | 6 | 7 | 0 | 0 | 4 | 1 | 0 | 0 | 10 | 8 | 0 | 0 |
| I1 + D1 | 5 | 3 | 4 | 1 | 5 | 0 | 2 | 0 | 10 | 3 | 6 | 1 |
| Unclassified (n = 26, 2.5%) | 11 | 6 | 1 | 1 | 4 | 3 | 0 | 0 | 15 | 9 | 1 | 1 |

UT, up-tag; DT, down-tag; S1, 1-bp substitution; I1, 1-bp deletion; D1, 1-bp deletion.

sequenced barcodes, 1,284 mutations were found in total, resulting in a mutation rate of 16.6%. There was no significant difference in mutation rate between up-tags and down-tags. When type of mutation was considered for the 1284 mutations (16.6% in total), substitutions accounted for 845 mutations (10.9%), deletions accounted for 319 mutations (4.1%), and insertions accounted for 121 mutations (1.6%). The overall mutation rate was similar between both yeasts (16.6% vs. 18%). Insertions and deletions occurred at a much lower frequency than substitutions. The frequency of insertions (1.6%) was 1 order of magnitude lower than that of substitutions (10.9%).

Types of defects in mutations

The presence of 1,284 mutations within 1,039 barcodes prompted us to examine the types of barcode defects in detail. As shown in Table 3, substitutions accounted for 702 mutated barcodes (67.6%), versus deletions for 177 (17.0%), insertions for 79 (7.6%), and others for 81 (complex and unclassified, 7.8%). When compared with budding yeast (~28%), the overall occurrence of substitution was unexpectedly high (refer to Discussion).

Mutation rate by position within 20-mer barcodes

To estimate the mutation rate by position within 20-mer

barcodes, 1,284 mutations from 7,734 Sanger-sequenced barcodes were analyzed. As shown in Table 4 and Fig. 2, there was no significant difference between up-tags and down-tags at a given position. The mutation frequency at a given position was similar at most positions, ranging from 0.4% (32/7,734) to 1.1% (82/7,734), except at position 1, which was highest (3.1%). Intriguingly, the mutation frequency showed an increasing tendency from position 20 to position

1 in a similar manner as budding yeast [9]. The phenomenon is consistent with a decrease in the fidelity of the solid-phase chemical synthesis with distance from the first position at the 3'-end of each oligonucleotide.

Table 4. Statistics of 1,284 mutations by position within 20-mer barcodes

| PCR type | Mutation | Tag | Base position | | | | | | | | | | | | | | | | | | | | Sum |
|--------------------|-------------------------------------|-----|---------------|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|-------|
| | | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | |
| Block | Substitution (n = 654, 50.9%) | UT | 112 | 11 | 15 | 11 | 22 | 12 | 12 | 25 | 14 | 13 | 10 | 9 | 5 | 10 | 17 | 13 | 10 | 5 | 8 | 17 | 351 |
| | | DT | 82 | 14 | 17 | 16 | 9 | 10 | 11 | 14 | 13 | 8 | 13 | 11 | 10 | 13 | 12 | 11 | 7 | 6 | 9 | 17 | 303 |
| | Deletion (n = 187, 14.6%) | UT | 6 | 8 | 11 | 2 | 2 | 4 | 5 | 3 | 4 | 5 | 3 | 3 | 2 | 3 | 3 | 1 | 3 | 0 | 2 | 8 | 78 |
| | | DT | 12 | 16 | 12 | 4 | 4 | 5 | 6 | 9 | 5 | 4 | 7 | 3 | 6 | 3 | 5 | 3 | 1 | 0 | 1 | 3 | 109 |
| | Insertion (n = 90, 7.0%) | UT | 5 | 2 | 5 | 2 | 3 | 1 | 2 | 0 | 3 | 1 | 4 | 2 | 1 | 3 | 4 | 1 | 5 | 0 | 1 | 1 | 46 |
| DT | | 0 | 2 | 2 | 2 | 4 | 4 | 2 | 1 | 4 | 1 | 2 | 3 | 4 | 1 | 3 | 2 | 2 | 3 | 2 | 0 | 44 | |
| | Sum | | 217 | 53 | 62 | 37 | 44 | 36 | 38 | 52 | 43 | 32 | 39 | 31 | 28 | 33 | 44 | 31 | 28 | 14 | 23 | 46 | 931 |
| Serial | Substitution (n = 191, 14.9%) | UT | 7 | 6 | 8 | 8 | 4 | 7 | 4 | 5 | 5 | 9 | 7 | 5 | 8 | 4 | 4 | 1 | 4 | 6 | 3 | 6 | 111 |
| | | DT | 4 | 2 | 4 | 5 | 6 | 5 | 5 | 4 | 1 | 0 | 3 | 1 | 4 | 6 | 3 | 4 | 5 | 9 | 3 | 6 | 80 |
| | Deletion (n = 132, 10.3%) | UT | 7 | 3 | 5 | 8 | 3 | 7 | 3 | 5 | 5 | 3 | 2 | 8 | 0 | 2 | 1 | 2 | 1 | 1 | 2 | 4 | 72 |
| | | DT | 4 | 1 | 2 | 3 | 6 | 4 | 3 | 4 | 7 | 3 | 5 | 5 | 1 | 1 | 3 | 2 | 2 | 2 | 1 | 0 | 59 |
| | Insertion (n = 31, 2.4%) | UT | 1 | 5 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | 2 | 1 | 2 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 16 |
| DT | | 1 | 1 | 1 | 0 | 0 | 2 | 0 | 1 | 0 | 0 | 2 | 3 | 0 | 1 | 3 | 0 | 0 | 0 | 0 | 0 | 15 | |
| | Sum | | 24 | 18 | 20 | 24 | 19 | 25 | 16 | 19 | 19 | 16 | 21 | 23 | 15 | 14 | 15 | 9 | 12 | 18 | 10 | 16 | 353 |
| Total ^a | | | 241 | 71 | 82 | 61 | 63 | 61 | 54 | 71 | 62 | 48 | 60 | 54 | 43 | 47 | 59 | 40 | 40 | 32 | 33 | 62 | 1,284 |

UT, up-tag; DT, down-tag.

^aThe number of mutations within the barcodes (1,284) is more than the number of defective barcodes (1,039) because of multiple mutations in a single barcode.

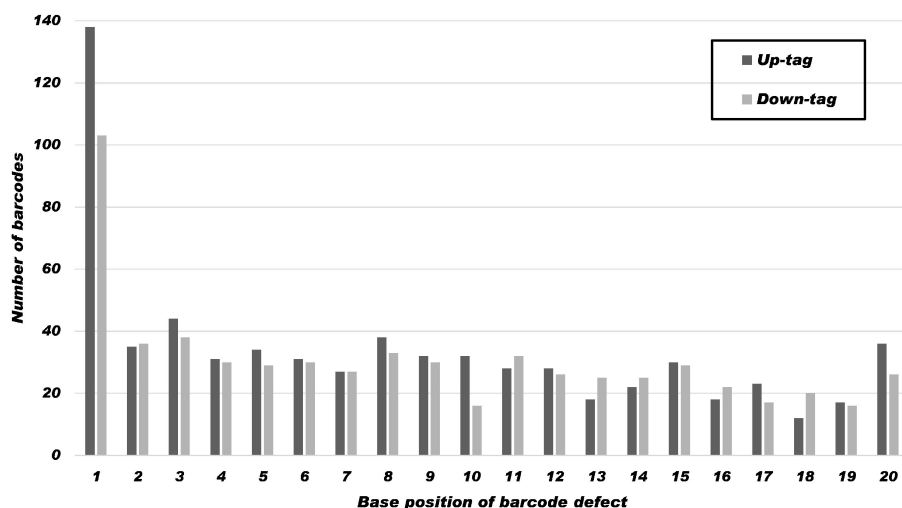


Fig. 2. Occurrence of sequence variation by position within tags. The results show a summary of the 1,284 nucleotide sequence mutations in 7,734 Sanger-sequenced up-tags (dark grey) and down-tags (light grey) at a given position. The mutation frequency at a given position in the 20-mer barcode is similar at most positions, ranging from 0.4% (32/7,734) to 1.1%, except at position 1, which is the highest (3.1%). It shows a tendency to increase from position 20 to position 1. For the statistics in detail, refer to Table 4.

Discussion

The 11.2% failure rate (974 barcodes) of validation was unexpectedly high. The phenomenon happened simply because we only analyzed high-fidelity barcodes. When barcodes of lower fidelity that was confirmed only on either side were counted, the success rate increased to 95%. The results suggest that all of the sequenced barcodes did not locate properly because of many causes, including imperfect genome sequence information, redundancy among ORFs (e.g., *wtf*), and ambiguous PCR products due to sequence defects in PCR priming sites [9]. Therefore, usage of another sequencing method, such as NGS, would increase the success rate of the barcode validation up to >95%, as with budding yeast [10], because NGS technology directly counts barcode sequences without PCR amplification.

The overall mutation frequency (16.6%) in the fission yeast barcodes was similar to that in budding yeast (~18%) [10]. In detail, the overall rates of deletions and insertions were within reasonable ranges, compared with previous results in budding yeast [9, 10]. However, the substitution rate (702 mutated barcodes, 67.6%) in fission yeast was unexpectedly high compared with budding yeast (~28%). Moreover, it was well above the predicted error frequency from the raw Sanger sequencing data [16]: 0.1% occurrence per each base position or ca. 2% occurrence within a 20-mer tag. In addition, the high substitution rate might come from intrinsic errors during the solid-phase oligonucleotide chemical synthesis. For example, a single-nucleotide deletion might be expected to occur randomly due to periodic de-blocking failure during the oligonucleotide synthesis. In contrast, substitutions and insertions are less likely to arise during oligonucleotide synthesis, because they should come from mechanical or software errors. However, contamination of the reagents would give rise to a substitution rate of <0.2% per base position or a 4% occurrence within a 20-mer tag. Another explanation for a substitution rate (16.6%) higher than 6% might come from nucleotide misincorporation error ($1/10^3$ – 10^4) of *Taq* polymerase during PCR amplification of the barcode regions. In addition, extra errors might come from errors in base-calling, because only Sanger sequencing technology was used. This will be clarified by additional analyses using NGS technology, which are underway at the moment.

Intriguingly, the sequence mutation of each base position showed a gradient difference within a 20-mer barcode in a similar manner as in budding yeast [9]. Generally, the phenomenon is consistent with a decrease in the fidelity of the oligonucleotide chemical synthesis, depending on the distance from the first position at the 3'-end of each synthetic oligonucleotide. However, the prominent high occurrence at

position number 1 is difficult to explain. There is no reason why position 1 had a peculiarly high mutation rate in common, because the 20-mer barcode sequences were located inside the 75-mer oligonucleotide primers used for PCR amplification of the KanMX module in both budding and fission yeasts.

Incorporation of unique barcode tags into the fission yeast gene deletion library has enabled the identification of gene functions by parallel growth fitness analysis [8]. However, the assay system requires renovation of either mutated barcodes or mismatched array sequences, which carry significant up-front costs. To circumvent the expensive situation, an NGS approach is preferred over fixing the previous microarray platform. NGS analysis directly counts each barcode in a complex sample via sequencing using a protocol, called “Barcode Analysis by Sequencing,” or “Bar-seq.” Together, well-defined barcode sequences, combined with the NGS platform, promise to make the fission yeast gene deletion library a powerful tool for understanding gene function. In this regard, the results in this study are a basis for upgrade of the parallel analysis from the previous microarray technology to NGS technology.

ORCID: Minhoo Lee: <https://orcid.org/0000-0002-0168-9546>; Shin-Jung Choi: <https://orcid.org/0000-0003-3549-121X>; Sangjo Han: <https://orcid.org/0000-0002-7644-3671>; Miyoung Nam: <https://orcid.org/0000-0002-7465-4029>; Dongsup Kim: <https://orcid.org/0000-0002-5916-6799>; Dong-Uk Kim: <https://orcid.org/0000-0003-4088-0893>; Kwang-Lae Hoe: <https://orcid.org/0000-0002-3943-4549>

Authors' contributions

Conceptualization: DK, DUK, KLH
 Data curation: MN, SH, SJC
 Funding acquisition: ML, DUK, KLH
 Experiments: MN, SJC
 Data analysis: ML, SH
 Writing – original draft: ML, DUK, KLH
 Writing – review & editing: ML, KLH

Acknowledgments

This work was supported by a grant from Chungnam National University.

References

1. Nurse P, Thuriaux P, Nasmyth K. Genetic control of the cell division cycle in the fission yeast *Schizosaccharomyces pombe*. *Mol Gen Genet* 1976;146:167-178.

2. Hedges SB. The origin and evolution of model organisms. *Nat Rev Genet* 2002;3:838-849.
3. Wood V, Gwilliam R, Rajandream MA, Lyne M, Lyne R, Stewart A, *et al.* The genome sequence of *Schizosaccharomyces pombe*. *Nature* 2002;415:871-880.
4. Winzeler EA, Shoemaker DD, Astromoff A, Liang H, Anderson K, Andre B, *et al.* Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis. *Science* 1999;285:901-906.
5. Kim DU, Hayles J, Kim D, Wood V, Park HO, Won M, *et al.* Analysis of a genome-wide set of gene deletions in the fission yeast *Schizosaccharomyces pombe*. *Nat Biotechnol* 2010;28: 617-623.
6. Hayles J, Wood V, Jeffery L, Hoe KL, Kim DU, Park HO, *et al.* A genome-wide resource of cell cycle and cell shape genes of fission yeast. *Open Biol* 2013;3:130053.
7. Shoemaker DD, Lashkari DA, Morris D, Mittmann M, Davis RW. Quantitative phenotypic analysis of yeast deletion mutants using a highly parallel molecular bar-coding strategy. *Nat Genet* 1996;14:450-456.
8. Han S, Lee M, Chang H, Nam M, Park HO, Kwak YS, *et al.* Construction of the first compendium of chemical-genetic profiles in the fission yeast *Schizosaccharomyces pombe* and comparative compendium approach. *Biochem Biophys Res Commun* 2013;436:613-618.
9. Eason RG, Pourmand N, Tongprasit W, Herman ZS, Anthony K, Jejelowo O, *et al.* Characterization of synthetic DNA bar codes in *Saccharomyces cerevisiae* gene-deletion strains. *Proc Natl Acad Sci U S A* 2004;101:11046-11051.
10. Smith AM, Heisler LE, Mellor J, Kaper F, Thompson MJ, Chee M, *et al.* Quantitative phenotyping via deep barcode sequencing. *Genome Res* 2009;19:1836-1842.
11. Han TX, Xu XY, Zhang MJ, Peng X, Du LL. Global fitness profiling of fission yeast deletion strains by barcode sequencing. *Genome Biol* 2010;11:R60.
12. Forsburg SL, Rhind N. Basic methods for fission yeast. *Yeast* 2006;23:173-183.
13. Rice P, Longden I, Bleasby A. EMBOSS: the European molecular biology open software suite. *Trends Genet* 2000;16:276-277.
14. Needleman SB, Wunsch CD. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J Mol Biol* 1970;48:443-453.
15. Nam M, Lee SJ, Han S, Kim D, Lee M, Kang EJ, *et al.* Systematic targeted gene deletion using the gene-synthesis method in fission yeast. *J Microbiol Methods* 2014;106:72-77.
16. Richterich P. Estimation of errors in "raw" DNA sequences: a validation study. *Genome Res* 1998;8:251-259.