Clin. Lab. 2023;69:XXX-XXX


## ORIGINAL ARTICLE

# The Diagnostic Yield and Difficulties of Utilizing Soft-clipped Read Clusters Encountered in Clinical Exome Sequencing

Young-gon Kim [1, 2], Nan Young Lee [1, 3], Ji Yeon Ham [1, 3], Taeheon Lee [4], Chang-Seok Ki [4], Kyung Eun Song [1, 3]

[1] *Department of Clinical Pathology, School of Medicine, Kyungpook National University, Daegu, Republic of Korea*
[2] *Department of Laboratory Medicine & Genetics, Samsung Medical Center, Seoul, Republic of Korea*
[3] *Department of Laboratory Medicine, Kyungpook National University Chilgok Hospital, Daegu, Republic of Korea*
[4] *GC Genome, Yongin, Republic of Korea*

**SUMMARY**

***Background:*** **Despite the wide use of next generation sequencing, there are still many difficulties in detecting structural variants. A split read is one of the clues of structural variants and is represented as a soft-clipped read in the raw sequencing data. Considering that most of the breakpoints of structural variants reside in non-coding regions, split read information has not been routinely used in exome sequencing or targeted panel sequencing. Recently, SCRAMble, a software capable of detecting mobile element insertion (MEI) and deletion based on soft-clipped read clusters (SCRCs), was shown to provide an additional diagnostic yield of 0.03 - 0.25%. SCRAMble is the only software that can be used for exome sequencing or targeted panel sequencing to detect structural variants based on SCRC information. The aim of present study was to establish a working procedure of utilizing SCRC information using SCRAMble in clinical exome sequencing and to assess its diagnostic yield.**

***Methods:*** **Raw sequencing data of clinical exome sequencing were retrospectively analyzed using SCRAMble to search MEIs and deletions. SCRAMble software was installed according to the manufacturer's instructions and default parameters except for one, mei-score, which was adjusted for sensitivity, were used. RefSeq gene annotation was performed for both MEI and deletion calls using ANNOVAR. Blacklist-based filtering was used to reduce candidate MEI/deletion calls. Clinical relevance was manually evaluated for the remaining variant calls.**

***Results:*** **One diagnostic MEI, which is a founder variant in East Asia, was detected in two cases (2/266, 0.75%). In addition, two diagnostic deletions, which had been previously detected in depth-of-coverage (DOC)-based copy number variant (CNV) callers, were detected (2/266, 0.75%). Base-level breakpoints that could not be derived by the DOC-based callers were identified for these two deletions using SCRAMble. Most SCRCs were repetitive among cases and blacklist-based filtering reduced candidate MEI/deletion calls by 49.5 - 94.5%, leaving a considerable number of variant calls to be manually validated.**

***Conclusions:*** **SCRC screening in exome or targeted panel sequencing may provide additional diagnostic yield either by pathogenic MEI detection or reassurance of deletions identified by DOC-based CNV callers. Development of an efficient filtering algorithm is warranted.**

**(Clin. Lab. 2023;69:xx-xx. DOI: 10.7754/Clin.Lab.2022.220731)**

**Correspondence:**
Kyung Eun Song, M.D., Ph.D.
Department of Laboratory Medicine
Kyungpook National University Chilgok Hospital
807, Hoguk-ro, Buk-gu
Daegu 41404
South Korea
Phone: + 82 53-200-2831
Fax: +82-53-200-7299
E-mail: kesong@knu.ac.kr

ORCID: 0000-0003-1628-4741

Manuscript accepted August 3, 2022

**KEYWORDS**

soft-clipped read clusters, mobile element insertion, SCRAMble, clinical exome sequencing

## INTRODUCTION

Despite the development of diverse new technologies, including long-read sequencing, linked-read sequencing [1], and optical mapping [2], massive parallel sequencing based on short reads is still the mainstream sequencing technology used in clinical laboratories. Nonetheless, a limitation of short-read-based sequencing is that it cannot directly detect structural variants. Structural variants can only be inferred from the data produced by short-read sequencing by using approaches such as depth of coverage (DOC), split reads (SR), and read pairs (RP) [3-5]. Dozens of callers that detect structural variants from short-read-based sequencing are available online, and most of them apply one or a combination of three principle approaches: DOC, SR, and RP [3-12]. As the majority of the breakpoints of structural variants reside in non-coding regions, most callers that use SR and RP are developed for whole genome sequencing [5]. In contrast, for exome sequencing or targeted panel sequencing, copy number variant (CNV) callers that only use DOC are typically used [10-12].

Soft-clipped read clusters (SCRCs) are aligner-generated representations of SR (Figure 2, 3, 4). A soft clip represents a contiguous mismatch between the read and reference sequences. In addition to structural variants, sequencing errors, chimeric reads, errors in the reference genome, and genetic polymorphism among individuals may also cause SCRCs [6]. Although SCRCs are occasionally encountered during the interpretation of raw sequencing data, currently, they are not incorporated into automated variant calling pipeline in exome or targeted panel sequencing.

Mobile element insertion (MEI) is a kind of structural variant caused by the insertion of transposable elements, such as Alu, L1, and SVA, into the genome. MEIs are a known cause of human genetic disease but are not routinely targeted in clinical diagnostic testing because of their scarcity and detection difficulties [13]. However, many of the reported pathogenic MEIs are located in exons or flanking introns [13-20], and thus can potentially be detected by exome or targeted panel sequencing. As MEIs does not change the DOC because they represent an insertion of a genetic segment that is not part of the human genome, they cannot be detected by DOC-based CNV callers, which are widely used in exome or targeted panel sequencing.

The soft-clipped read alignment mapper (SCRAMble) is a recently developed structural variant caller optimized for exome or targeted panel sequencing [13]. SCRAMble aimed to focus on MEI detection based on SR, which is presented as SCRCs. SCRAMble was shown to outperform conventional MEI callers that use both RP and SR, such as MELT [21] and Mobster [22], in detecting MEIs from exome or targeted panel sequencing data. In their study, diagnostic MEI was found in 0.03% of all cases. SCRAMble also supports the detection of large deletions based on soft-clipped reads, although its performance has not been reported yet. To the best of our knowledge, SCRAMble is the only software that is optimized for exome or targeted panel sequencing in detecting structural variants based on SCRCs.

As mentioned above, in addition to the structural variants, there are many other causes of SCRCs leading to highly frequent false positive findings, requiring an effective filtering algorithm. The present study aimed to establish a working procedure of utilizing SCRAMble, including variant filtering, and assess the diagnostic yield of utilizing SCRC information for structural variant detection using SCRAMble.

## MATERIALS AND METHODS

### Participants

This study was approved by the Institutional Review Board of the Kyungpook National University Chilgok Hospital, Daegu, South Korea (IRB No. KNUCH 2021-12-035). Informed consent for genetic testing was obtained from all included participants or their guardians in case of minor patients. Patients (n = 266) who had been referred for CES between January 2020 and December 2021 were included in the study, and their raw sequencing data (BAM files) were retrospectively analyzed. Reasons for CES included autism spectrum disorder (59/266, 22.2%), coagulation disorder (26/266, 9.8%), epilepsy (20/266, 7.5%), inborn errors of metabolism (20/266, 7.5%), and amyotrophic lateral sclerosis (18/266, 6.8%) (Supplemental Table S1).

### Clinical exome sequencing

Sequencing was performed by GC Genome (Yongin, Korea). CES was performed using the G-Mendeliome CES panel (Celemics, Seoul, Korea). Two versions of the CES panels were used: Panel 1, which consisted of 5,447 genes (14.0 Mb), was used for patients referred from January 2020 to December 2020 (n = 44); and Panel 2, which consisted of 7,537 genes (19.8 Mb), was used for patients referred from January 2021 to December 2021 (n = 222). Except for the target regions, the same procedure was applied to both panels.

The experimental procedures were performed using the following steps: genomic DNA was extracted from EDTA-treated whole blood samples. The G-Mendeliome panel (Celemics) was used for library preparation, and sequencing was performed on the MGI-G400 platform (MGI Tech Co., Shenzhen, China). DNA sequence reads were aligned to the reference sequence based on the public human genome build GRCh37/UCSC hg19. Alignment was performed using the BWA-mem algorithm (version 0.7.17), duplicated reads were marked

**Table 1. Genomic distribution of SCRCs, MEI calls, and deletion calls found in Panel 1 and Panel 2 by SCRAMble.**

| Genomic region | Region description * | Panel 1 (n = 44, mean depth = 144.3) | | | | | Panel 2 (n = 222, mean depth = 257.6) | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | SCRC /case (%) | No. of MEI calls/case (%) | | No. of deletion calls/case (%) | | SCRC /case (%) | No. of MEI calls/case (%) | | No. of deletion calls/case (%) | |
| | | | All | AF ≤ 15% | All | AF ≤ 15% | | All | AF ≤ 15% | All | AF ≤ 15% |
| Exonic | Variant overlaps a coding sequence | 985.98 (57.27) | 0 (0) | 0 (0) | 12.7 (64.76) | 4.18 (83.27) | 2,049.52 (34.61) | 0.95 (6.15) | 0.01 (0.21) | 11.15 (34.58) | 4.54 (27.85) |
| Splicing | Variant is within 2 bp of a splicing junction | 13.39 (0.78) | 0 (0) | 0 (0) | 1.09 (5.56) | 0.09 (1.79) | 61.08 (1.03) | 0.01 (0.06) | 0.01 (0.21) | 2.26 (7.01) | 0.3 (1.84) |
| ncRNA | Variant overlaps a transcript without coding annotation in the gene definition | 44.14 (2.56) | 0 (0) | 0 (0) | 0 (0) | 0 (0) | 233.21 (3.94) | 1.07 (6.93) | 0.61 (13.03) | 0.38 (1.18) | 0.38 (2.33) |
| UTR5 | Variant overlaps a 5' - untranslated region | 36.45 (2.12) | 0 (0) | 0 (0) | 0.02 (0.1) | 0.02 (0.4) | 121.71 (2.06) | 0.01 (0.06) | 0.01 (0.21) | 0.03 (0.09) | 0.03 (0.18) |
| UTR3 | Variant overlaps a 3' - untranslated region | 43.52 (2.53) | 0 (0) | 0 (0) | 0.11 (0.56) | 0.11 (2.19) | 122.52 (2.07) | 0.03 (0.19) | 0.03 (0.64) | 0.09 (0.28) | 0.09 (0.55) |
| Intronic | Variant overlaps an intron | 583.68 (33.9) | 2.89 (100) | 0.16 (100) | 5.52 (28.15) | 0.61 (12.15) | 2,326.05 (39.28) | 12.54 (81.22) | 3.43 (73.29) | 9.15 (28.38) | 4.34 (26.63) |
| Upstream | Variant overlaps 1 kb region upstream of the transcription start site | 9 (0.52) | 0 (0) | 0 (0) | 0.16 (0.82) | 0 (0) | 32.82 (0.55) | 0.08 (0.52) | 0.08 (1.71) | 0.63 (1.95) | 0.05 (0.31) |
| Down-stream | Variant overlaps 1 kb region downstream of the transcription end site (use - neargene to change this) | 0.09 (0.01) | 0 (0) | 0 (0) | 0 (0) | 0 (0) | 19.21 (0.32) | 0 (0) | 0 (0) | 0.83 (2.57) | 0.22 (1.35) |
| Intergenic | Variant is in an intergenic region | 5.48 (0.32) | 0 (0) | 0 (0) | 0 (0) | 0 (0) | 956.23 (16.15) | 0.74 (4.79) | 0.49 (10.47) | 7.73 (23.98) | 6.35 (38.96) |
| Total | | 1,721.73 (100) | 2.89 (100) | 0.16 (100) | 19.61 (100) | 5.02 (100) | 5,922.36 (100) | 15.44 (100%) | 4.68 (100) | 32.24 (100) | 16.3 (100) |

**AF - allele frequency, MEI - mobile element insertion, ncRNA - non-coding RNA, SCRC - soft-clipped read clusters, UTR - untranslated region.**
*** https://annovar.openbioinformatics.org/en/latest/user-guide/gene/.**

**Table 2. Diagnostic deletions detected by DOC-based CNV analysis performed at the time of CES and corresponding SCRAMble results.**

| Case ID | Reason for the test | Deletions detected by DOC-based CNV detection | CNV algorithm | Diagnosis based on the deletion | SCRAMble result |
|---|---|---|---|---|---|
| P067 | Diabetes mellitus | *HNF1B*, whole gene | exon-level CNV | renal cysts and diabetes syndrome | not detected |
| P094 | Coagulation disorder | *F11*, whole gene | exon-level CNV | *F11* deficiency | not detected |
| P121 | Muscular dystrophy | *DMD*, exons 45 - 47 | exon-level CNV | becker muscular dystrophy | not detected |
| P151 | Coagulation disorder | chr22: 18893877_20786306del | XHMM | 22q11.21 deletion syndrome | not detected |
| P170 | Short stature | chr16: 15771647_16317301del | XHMM | 16p13.11 microdeletion | not detected |
| P213 | Dyslipidemia | *LDLR*, exon 7 | exon-level CNV | familial hypercholesterolemia | chr19:11219133_11222209del (deletion of *LDLR* exon 7 and a part of exon 8) |
| P220 | Diabetes mellitus | chr17: 34842523_36104885del | XHMM | 17q12 microdeletion syndrome | Not detected |
| P229 | Developmental delay | *MECP2*, exon 3 | exon-level CNV | rett syndrome | chrX:153296037_153302498del (deletion of *MECP2* exon 3 and a part of exon 4) |

**CES - clinical exome sequencing, CNV - copy number variants, DOC - depth of coverage, XHMM - eXome-Hidden Markov Model software.**

with the biobambam2 toolkit, and base quality recalibration was performed using the Genome Analysis Toolkit (version 4.1.8). DOC-based CNV analysis was performed by two methods: a segmentation approach, in which the eXome-Hidden Markov Model (XHMM) software (Version 1.0) [23] was used, and exon-level CNV, in which an in-house algorithm that utilized normalized DOC was used.

### Interpretation of SCRAMble results

SCRAMble software (version 1.0.2, https://github.com/GeneDx/scramble) was installed on Ubuntu 18.04, according to the manufacturer's instructions. Default parameters of SCRAMble were used except for the 'meiscore' option, for which a lower value 30 was used because default value 50 failed to detect the previously detected LINE-1 insertion [20]. RefSeq gene annotation was performed for both MEI/deletion calls using ANNOVAR (version 2019Oct24) [24]. In-house allele frequency of MEI/deletion calls was calculated using Linux shell script (Ubuntu 18.04) and python programming language (version 3.9.1) with pandas library (version 1.2.0). Based on the in-house frequency, variants that were called more frequently than 15% of the cases in the same panel were incorporated into the blacklists. After removing these variants, the remaining variant calls were manually reviewed in terms of clinical relevance based on available information, such as disease-gene relationships, genomic regions (coding/non-coding, etc.), previous reports of the same variant, in-house variant frequency, and BAM file findings viewed through the Integrative Genomics Viewer (version 2.11.9).

## RESULTS

### Distribution of SCRCs, MEI calls, and deletion calls

The distribution of SCRCs and MEI/deletion calls according to the genomic region is presented in Table 1. Panel 1 showed an average of 1,721.73 detected SCRCs, among which 2.89 (0.17%) and 19.61 (1.14%) were called MEIs and deletions, respectively. Most detected SCRCs were located in exonic regions (57.27%), followed by intronic regions (33.9%), non-coding RNA regions (2.57%), and 3'-untranslated regions (2.54%). All MEI calls were from intronic regions, whereas the majority of deletion calls were located in exonic (64.76%) and splicing (5.56%) regions. Panel 2 showed an average of 5,922.36 detected SCRCs, among which 15.44 (0.26%) and 32.24 (0.54%) were called MEIs and deletions, respectively. SCRCs were most frequently located in intronic regions (39.28%), followed by exonic regions (34.61%), intergenic regions (16.15%), and non-coding RNA regions (3.94%). Most (81.22%) MEI calls were from intronic regions, whereas deletion calls showed a wider distribution across exonic (34.58%), intronic (28.38%), intergenic (23.98%), and splicing (7.01%) regions.

Figure 1 shows the in-house allele frequency distribution of MEI/deletion calls found in this study. In Panel 1, 94.5% of the MEI calls showed an in-house allele fre-

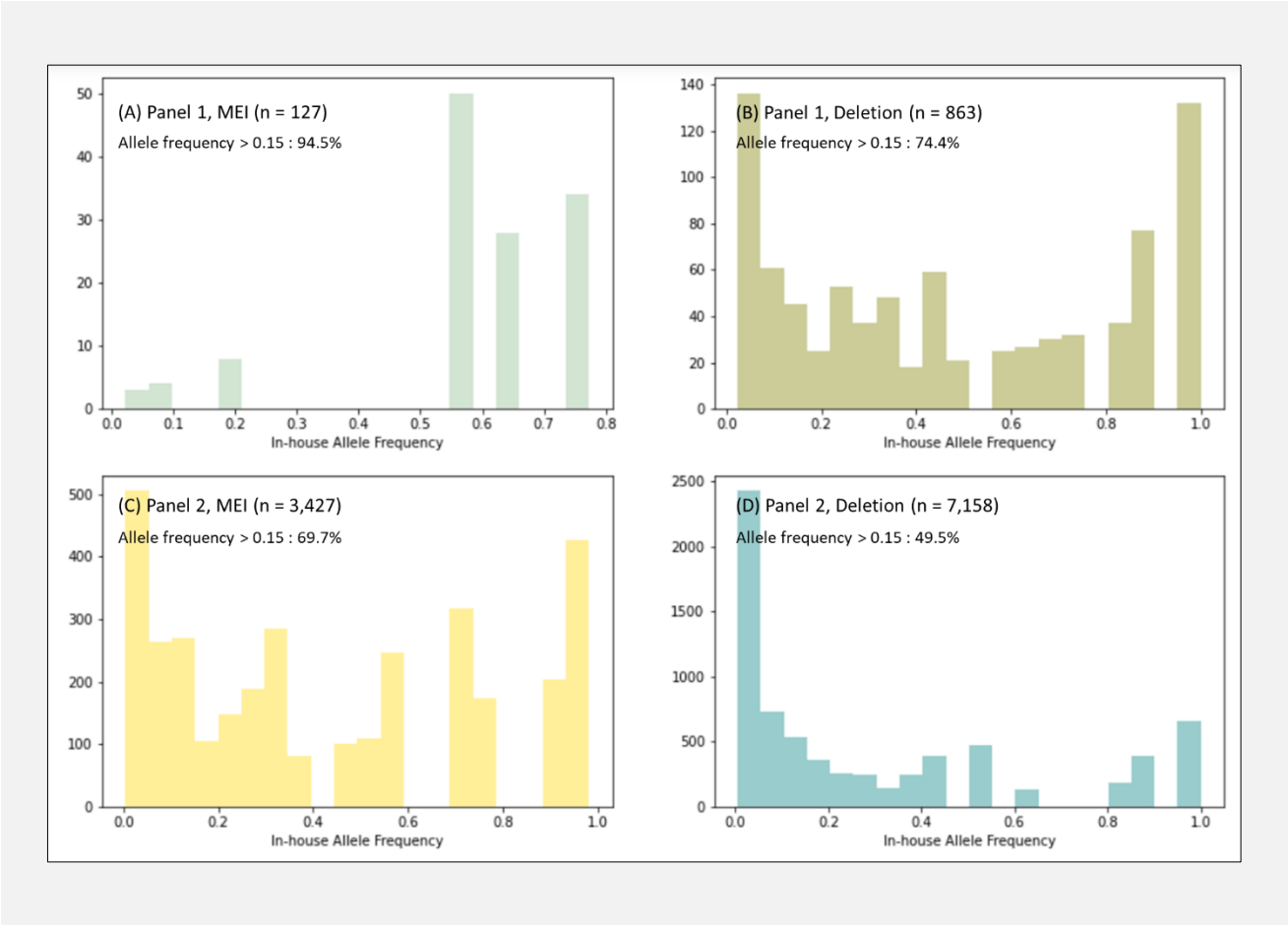


**Figure 1. In-house allele frequency distribution of mobile element insertion and deletion calls from two types of panels (Panel 1 and Panel 2).**

quency higher than 15%, which corresponds to the threshold for the blacklist, and 74.4% of the deletion calls showed allele frequencies higher than 15%. Panel 2 showed 69.7% of MEI calls with in-house allele frequencies higher than 15%, and 49.5% of deletion calls with allele frequencies higher than 15%. The blacklist of variant calls is provided in Supplemental Tables S2 - S5. The number of MEI/deletion calls after the blacklist was applied and whose in-house allele frequencies were ≤ 15% are listed in Table 1. The average number of MEI calls to be evaluated after blacklist-based filtering were 0.16 and 4.68 in Panel 1 and Panel 2, respectively, and the average number of deletion calls to be evaluated were 5.02 and 16.3, respectively.

**Diagnostic MEIs**

Among the 266 patients screened, intronic LINE-1 insertion in *SLCO1B3* (NM_019844.3), which is responsible for Rotor syndrome that causes benign fluctuating hyperbilirubinemia [20,25], was detected in 28 patients (10.5%, Supplemental Table S1). Based on the BAM file findings, five were homozygous. Rotor syndrome is inherited in a digenic recessive manner that it is caused by complete inactivation of two genes, *SLCO1B1* and *SLCO1B3* [26]. Among five patients who were detected with homozygous S*LCO1B3* LINE-1 insertion, two of them (P018 and P023, Figure 2) also carried a homozygous *SLCO1B1* pathogenic variant (NM_006446.4, c.1738C>T, p.R580*), completing the diagnosis of Rotor syndrome. CES was requested for these two patients because of cholestasis. They were suspected to have Rotor syndrome based on their phenotype and the *SLCO1B1* pathogenic variant found in CES. The diagnosis of Rotor syndrome, which could not have been confirmed because of the lack of the *SLCO1B3* variant, was confirmed based on the results of this study. As most of MEI calls were from non-coding regions (Table 1), their clinical significance could not be precisely evaluated, except for the *SLCO1B3* LINE-1 insertion. No other MEI calls identified in this study have been previously reported.

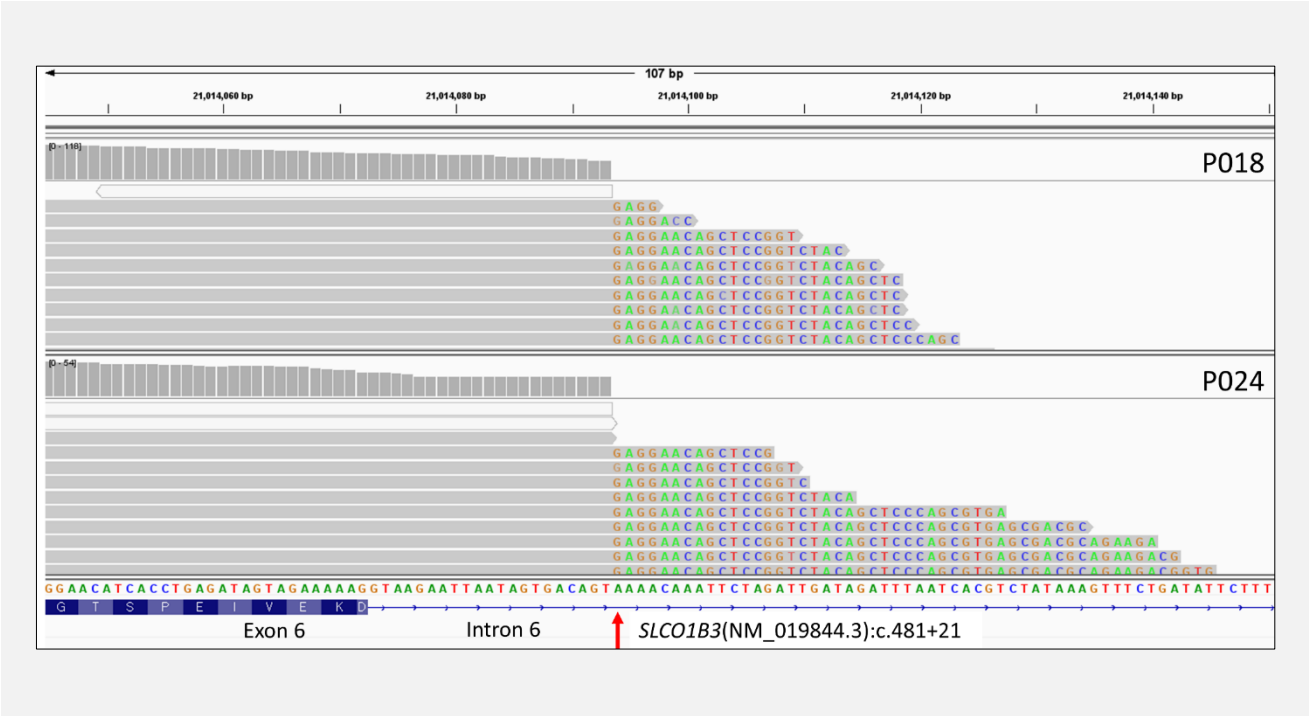


**Figure 2. BAM file findings of homozygous *SLCO1B3* LINE-1 insertion identified in two patients with Rotor syndrome (P018, upper; P024, lower).**

**Arrow denotes the breakpoint of insertion. All the reads covered at the junction of the insertion are soft-clipped on their right side and the clipped-reads share the same sequence, which is the sequence of inserted LINE-1. This is the same finding that was previously discovered in patients with Rotor syndrome [20].**

### Diagnostic deletions

Among the 266 patients included in the study, eight had diagnostic deletions previously detected by DOC-based CNV analysis, of which two cases (P213 and P229) were also detected by SCRAMble (Table 2). No additional diagnostic deletions were found only in SCRAMble. The difference between those detected by SCRAMble and those that were not was the location of the breakpoints. In cases where the deletion was detected by SCRAMBle, one of their breakpoints was exonic and thus was covered by the CES target region (Figures 3 and 4). Using the DOC-based exon-level CNV algorithm, the exact breakpoints could not be identified, and the deletions were called single-exon deletions (*LDLR* exon 7 deletion in P213 and *MECP2* exon 3 deletion in P229). However, after base-level breakpoints were identified from soft-clipped reads analysis, both deletions were proven to extend to an adjacent exon that contained the breakpoint (Figures 3 and 4). P213 had elevated LDL cholesterol levels that ranged between 159 - 302 mg/dL and was diagnosed with familial hypercholesterolemia based on the *LDLR* deletion found in CES. P229 was suspected to have Rett syndrome before the CES test, and the diagnosis was confirmed by the detection of *MECP2* deletion.

## DISCUSSION

In this study, the additional diagnostic yield from MEI detection was 0.75% (2/266), which was much higher than the rates reported in other studies, which ranged from 0.03 - 0.25% [13,18,27,28]. This was evident because of the highly prevalent *SLCO1B3* LINE-1 insertion that contributed to the diagnosis of two patients. In a previous study, the estimated carrier and allele frequencies of *SLCO1B3* LINE-1 insertion were of 13.1% and 6.8%, respectively, in the Korean population [20]. This exceptionally high frequency of a pathogenic variant can be explained by the digenic recessive mode of inheritance of *SLCO1B1* and *SLCO1B3*, which requires complete inactivation of both genes to cause a disease. In the current study, the estimated carrier and allele frequencies of the same insertion were 10.5% (28/266) and 6.2% (33/532), respectively.

Other founder MEIs are frequently observed in patients with a specific phenotype. For example, SVA insertion in the 3'-untranslated region of *FKTN* has been observed in most Fukuyama-type congenital muscular dystrophy patients in Japan (87.5 - 100%) and Korea (75%) [29-31]. Recently, a founder Alu insertion in exon 4 of *RP1* was detected in Japanese patients with retinitis pigmentosa. Of the 26 patients with a hetero-

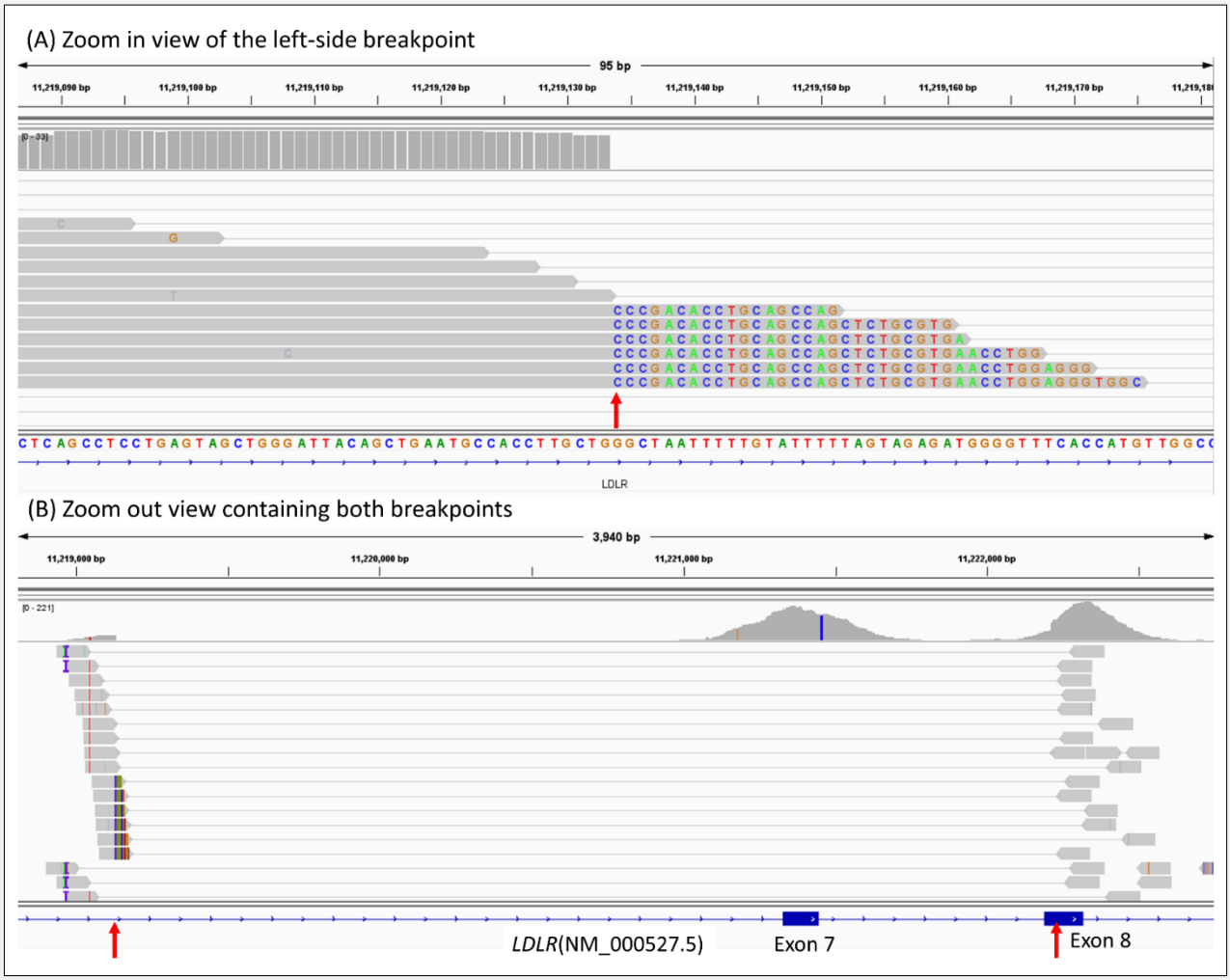


**Figure 3. BAM file findings of LDLR exon 7, 8 (part of exon 8) deletion identified in P213.**

**The left-side breakpoint was located in a deep intronic region that is not targeted by CES, but was covered by SCRAMble because the mate of the paired-reads were on a target region, exon 8. Soft-clipped reads are observed from the left-side breakpoint (A, B) and discordant read-pairs spanning the deletion range are observed (B).**

zygous pathogenic variant in *RP1*, 5 (19.2%) were found to carry an additional heterozygous Alu insertion [32]. In Korea, the same Alu insertion was detected in 1.8% of patients with inherited retinal disease [19]. In the current study, patients with inherited retinal disease were not included, and the 3'-untranslated region of *FKTN* was not included in the target region of CES. For the diagnosis of diseases with known founder MEIs, routine screening for SCRCs is warranted in exome or targeted panel sequencing. Additionally, founder MEIs located in non-coding regions could be added to the target region.

Although the exact breakpoints could be identified from the detected cases, the detection of deletions based only on soft-clipped reads showed limited sensitivity. SR-based CNV detection can be used adjunctively in exome or targeted panel sequencing along with DOC-based CNV detection algorithms. Deletions detected from both DOC-based CNV analysis and SR-based analysis must have higher positive predictive values than those detected only from DOC-based algorithms.

The 15% threshold for blacklist generation was determined from the frequency of *SLCO1B3* LINE-1 insertion, based on the assumption that no other pathogenic MEIs or deletions could have a higher carrier frequency. Use of 15% threshold can be supported by the updated recommendation for the benign stand-alone criterion, in which 13.7% is described as the highest minor


**Figure 4. BAM file findings of *MECP2* exon 3, 4 (part of exon 4) deletion identified in P229.**

**Right-side breakpoint was located in a deep intronic region that is not target of CES, but was covered by SCRAMble because the mate of the paired-reads were on a target region, exon 4. Soft-clipped reads are observed from both breakpoints (A, B) and discordant read-pairs spanning the deletion range are observed (B).**

allele frequency observed from a pathogenic variant [33]. Considering the repetitive nature of false-positive findings caused by sequencing errors, the blacklist should contain both the false-positive findings and benign variants. Thus, the blacklist-based filtering was used in a way that filters both the false-positive calls and benign variants while minimizing the chance of removing true pathogenic variants. However, a considerable number of variant calls remained after blacklist filtering, necessitating a tedious manual validation procedure.

Lowered 'mei-score' options of SCRAMble presumably have resulted in more false-positive calls, but we recommend the value we used, 30, for this option because a higher value, such as 40, failed to detected known *SLCO1B3* LINE-1 insertions that had been previously identified in the 1000 Genomes Project [20]. The numbers of variant calls that remained after the application of the blacklist were 0.16 MEI and 5.02 deletion calls for Panel 1 and 4.68 MEI and 16.30 deletion calls for Panel 2. Most MEI/deletion calls evaluated after the blacklist-based filtering were determined as false positives based on the BAM file findings. Variant validation based on manual evaluation of BAM file is widely used [34,35]. False-positive calls typically show poorly aligned soft-clipped reads, poor quality scores, and sim-

ilar BAM file findings to many other samples. Since the blacklist-based filtering should be used in a conservative way, the manual validation of numerous variants would be the biggest hurdle in adopting SCRAMble into routine practice.

The difference in the number of MEI/deletion calls between the two panels was larger than expected based on the difference in panel size (Panel 1: 14.0 Mb, Panel 2: 19.8 Mb). This was possibly due to the difference in the mean DOC (Panel 1: 144.3, Panel 2: 257.6), considering that only SCRCs with a sufficient number of supporting reads (five in SCRAMble, by default) were further analyzed.

This study had several limitations. First, owing to the retrospective nature of the study, confirmation tests for the detected variants could not be performed. However, in a previous study, all cases that were suspected to have *SLCO1B3* LINE-1 insertions based on the same BAM file findings identified in this study were confirmed by long-range or gap PCR tests (100%, 95/95) [20]. In addition, the frequency of *SLCO1B3* LINE-1 insertion variant was consistent with the previous study. Thus, it can be assumed that the *SLCO1B3* LINE-1 insertions found in this study are true variants. Second, the number of cases was small compared with previous studies that performed similar analyses [13,18]. A prospective study comprising a larger number of participants would be helpful to provide insights into the maximum potential of SCRC screening in exome or targeted panel sequencing. Third, the performance of the SCRAMble could not be compared to other software tools since SCRAMble was the only software optimized for exome sequencing or targeted panel sequencing.

In conclusion, SCRC screening using SCRAMble in exome or targeted panel sequencing may provide additional diagnostic yield either by pathogenic MEI detection or reassurance of deletions identified by DOC-based CNV callers. Development of new tools with advanced features such as variant scoring and filtering will be of great help for the incorporation of SCRC based structural variant detection in exome sequencing or targeted panel sequencing.

**Acknowledgment:**

The authors thank all the laboratory staff members who contributed to the production of the data used in this study.

**Declaration of Interest:**

The authors declare no potential conflict of interest.

**References:**

1. Elyanow R, Wu HT, Raphael BJ. Identifying structural variants using linked-read sequencing data. Bioinformatics 2018 Jan 15; 34(2):353-60. (PMID: 29112732)

2. Chan S, Lam E, Saghbini M, et al. Structural Variation Detection and Analysis Using Bionano Optical Mapping. Methods Mol Biol 2018;1833:193-203. (PMID: 30039375)

3. Tattini L, D'Aurizio R, Magi A. Detection of Genomic Structural Variants from Next-Generation Sequencing Data. Front Bioeng Biotechnol 2015;3:92. (PMID: 26161383)

4. Cameron DL, Di Stefano L, Papenfuss AT. Comprehensive evaluation and characterisation of short read general-purpose structural variant calling software. Nat Commun 2019 Jul 19;10(1):3240. (PMID: 31324872)

5. Kosugi S, Momozawa Y, Liu X, Terao C, Kubo M, Kamatani Y. Comprehensive evaluation of structural variation detection algorithms for whole genome sequencing. Genome Biol 2019 Jun 3; 20(1):117. (PMID: 31159850)

6. Schroder J, Hsu A, Boyle SE, et al. Socrates: identification of genomic rearrangements in tumour genomes by re-aligning soft clipped reads. Bioinformatics 2014 Apr 15;30(8):1064-72. (PMID: 24389656)

7. Nguyen HT, Boocock J, Merriman TR, Black MA. SRBreak: A Read-Depth and Split-Read Framework to Identify Breakpoints of Different Events Inside Simple Copy-Number Variable Regions. Front Genet 2016;7:160. (PMID: 27695476)

8. Zook JM, Hansen NF, Olson ND, et al. A robust benchmark for detection of germline large deletions and insertions. Nat Biotechnol 2020 Nov;38(11):1347-55. (PMID: 32541955)

9. Cameron DL, Baber J, Shale C, et al. GRIDSS2: comprehensive characterisation of somatic structural variation using single breakend variants and structural variant phasing. Genome Biol 2021 Jul 12;22(1):202. (PMID: 34253237)

10. Gordeeva V, Sharova E, Babalyan K, Sultanov R, Govorun VM, Arapidi G. Benchmarking germline CNV calling tools from exome sequencing data. Sci Rep 2021 Jul 13;11(1):14416. (PMID: 34257369)

11. Kim MJ, Lee S, Yun H, et al. Consistent count region-copy number variation (CCR-CNV): an expandable and robust tool for clinical diagnosis of copy number variation at the exon level using next-generation sequencing data. Genet Med 2022 Mar;24(3): 663-72. (PMID: 34906491)

12. Singh AK, Olsen MF, Lavik LAS, Vold T, Drablos F, Sjursen W. Detecting copy number variation in next generation sequencing data from diagnostic gene panels. BMC Med Genomics 2021 Aug 31;14(1):214. (PMID: 34465341)

13. Torene RI, Galens K, Liu S, et al. Mobile element insertion detection in 89,874 clinical exomes. Genet Med 2020 May;22(5):974-8. (PMID: 31965078)

14. Hancks DC, Kazazian HH Jr. Roles for retrotransposon insertions in human disease. Mob DNA 2016;7:9. (PMID: 27158268)

15. Qian Y, Mancini-DiNardo D, Judkins T, et al. Identification of pathogenic retrotransposon insertions in cancer predisposition genes. Cancer Genet 2017 Oct;216-217:159-69. (PMID: 29025590)

16. Bouras A, Leone M, Bonadona V, Lebrun M, Calender A, Boutry-Kryza N. Identification and Characterization of New Alu Element Insertion in the BRCA1 Exon 14 Associated with Hereditary Breast and Ovarian Cancer. Genes (Basel) 2021 Oct 29;12(11):1736. (PMID: 34828342)

17. Chu C, Borges-Monroy R, Viswanadham VV, et al. Comprehensive identification of transposable element insertions using multiple sequencing technologies. Nat Commun 2021 Jun 22;12(1): 3836. (PMID: 34158502)

18. Demidov G, Park J, Armeanu-Ebinger S, et al. Detection of mobile elements insertions for routine clinical diagnostics in targeted sequencing data. Mol Genet Genomic Med 2021 Dec; 9(12):e1807. (PMID: 34491624)

19. Won D, Hwang JY, Shim Y, et al. In Silico identification of a common mobile element insertion in exon 4 of RP1. Sci Rep 2021 Jun 28;11(1):13381. (PMID: 34183725)

20. Kim YG, Sung H, Shin HS, et al. Intronic LINE-1 insertion in SLCO1B3 as a highly prevalent cause of rotor syndrome in East Asian population. J Hum Genet 2022 Feb;67(2):71-7. (PMID: 34354231)

21. Gardner EJ, Lam VK, Harris DN, et al. The Mobile Element Locator Tool (MELT): population-scale mobile element discovery and biology. Genome Res 2017 Nov;27(11):1916-29. (PMID: 28855259)

22. Thung DT, de Ligt J, Vissers LE, et al. Mobster: accurate detection of mobile element insertions in next generation sequencing data. Genome Biol 2014;15(10):488. (PMID: 25348035)

23. Fromer M, Purcell SM. Using XHMM Software to Detect Copy Number Variation in Whole-Exome Sequencing Data. Curr Protoc Hum Genet 2014 Apr 24;81:7.23.1-7.23.21. (PMID: 24763994)

24. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res 2010 Sep;38(16):e164. (PMID: 20601685)

25. Kagawa T, Oka A, Kobayashi Y, et al. Recessive inheritance of population-specific intronic LINE-1 insertion causes a rotor syndrome phenotype. Hum Mutat 2015 Mar;36(3):327-32. (PMID: 25546334)

26. van de Steeg E, Stranecky V, Hartmannova H, et al. Complete OATP1B1 and OATP1B3 deficiency causes human Rotor syndrome by interrupting conjugated bilirubin reuptake into the liver. J Clin Invest 2012 Feb;122(2):519-28. (PMID: 22232210)

27. Chen JM, Chuzhanova N, Stenson PD, Ferec C, Cooper DN. Meta-analysis of gross insertions causing human genetic disease: novel mutational mechanisms and the role of replication slippage. Hum Mutat 2005 Feb;25(2):207-21. (PMID: 15643617)

28. Gardner EJ, Prigmore E, Gallone G, et al. Contribution of retrotransposition to developmental disorders. Nat Commun 2019 Oct 11;10(1):4630. (PMID: 31604926)

29. Kobayashi K, Nakahori Y, Miyake M, et al. An ancient retrotransposal insertion causes Fukuyama-type congenital muscular dystrophy. Nature 1998 Jul 23;394(6691):388-92. (PMID: 9690476)

30. Lim BC, Ki CS, Kim JW, et al. Fukutin mutations in congenital muscular dystrophies with defective glycosylation of dystroglycan in Korea. Neuromuscul Disord 2010 Aug;20(8):524-30. (PMID: 20620061)

31. Kobayashi K, Kato R, Kondo-Iida E, et al. Deep-intronic variant of fukutin is the most prevalent point mutation of Fukuyama congenital muscular dystrophy in Japan. J Hum Genet 2017 Nov; 62(11):945-8. (PMID: 28680109)

32. Nishiguchi KM, Fujita K, Ikeda Y, et al. A founder Alu insertion in RP1 gene in Japanese patients with retinitis pigmentosa. Jpn J Ophthalmol 2020 Jul;64(4):346-50. (PMID: 32193659)

33. Ghosh R, Harrison SM, Rehm HL, Plon SE, Biesecker LG; ClinGen Sequence Variant Interpretation Working Group. Updated recommendation for the benign stand-alone ACMG/AMP criterion. Hum Mutat 2018 Nov;39(11):1525-30. (PMID: 30311383)

34. Robinson JT, Thorvaldsdóttir H, Wenger AM, Zehir A, Mesirov JP. Variant review with the integrative genomics viewer. Cancer Res 2017 Nov;77(21):e31-4. (PMID: 29092934)

35. Camp SY, Kofman E, Reardon B, et al. Evaluating the molecular diagnostic yield of joint genotyping–based approach for detecting rare germline pathogenic and putative loss-of-function variants. Genet Med 2021 May;23(5):918-26. (PMID: 33531667)