Application Note

Detection of *FLT3* internal tandem duplications with QIAseq® Targeted Panels and QIAGEN® CLC Genomics Workbench analysis workflows

Anne-Mette K. Hein[1], Matthijs Vynck[2] and Friedel Nollet[2]

[1] QIAGEN Aarhus A/S, Silkeborgvej 2, DK-8000 Aarhus C, Denmark

[2] Molecular Hematology Unit, Department of Laboratory Medicine, AZ Sint-Jan Brugge-Oostende AV – Campus Brugge Ruddershove 10, B-8000 Brugge, Belgium

Introduction

Targeted DNA panels offer a cost-effective approach to meet the high coverage demand that is inherent to certain challenging applications, such as detection of specific variants. The use of unique molecular indices (UMIs) allows the removal of PCR amplification artifacts, resulting in highly accurate and precise detection. QIAGEN's QIAseq panels and corresponding analysis tools combine powerful single-primer extension chemistry and UMIs with UMI-aware bioinformatics workflows. QIAGEN CLC Genomics Workbench provides an easy-to-use graphical user interface for the analysis of genomics data, including QIAseq Targeted Panel data, enabling the calling of complex variants and visual inspection of results in the context of read mappings.

Here we present the results of a study evaluating the detection performance of internal tandem duplications (ITDs) in the *fms related receptor tyrosine kinase 3 (FLT3)* gene using next-generation sequencing (NGS). The study used a set of human bone marrow samples with internal tandem duplications in exons 13 to 15 of the *FLT3* gene, which were confirmed by PCR followed by capillary electrophoresis (PCR-CE). Sample DNA NGS libraries were prepared using a QIAseq Targeted DNA Custom Panel, sequenced on Illumina® instruments and analyzed with the "QIAseq Targeted DNA analysis ready-to-use workflow" available in the Biomedical Genomics Analysis plugin of QIAGEN CLC Genomics Workbench.

Detection of *FLT3* ITDs with QIAseq Targeted Panels and QIAGEN CLC Genomics Workbench

A *FLT3* gene ITD is a common driver mutation that confers a poor prognosis in patients with acute myeloid leukemia (AML) and occurs in approximately 30% of all AML cases. ITD-status is often used as guidance to understand therapeutic options.

The Biomedical Genomics Analysis plugin for QIAGEN CLC Genomics Workbench comes with ready-to-use workflows for the analysis of all off-the-shelf QIAseq panels, including QIAseq Leukemia Multimodal Panel. For custom-designed panels, the user is prompted by the software to provide the primer and target regions defining the panel design. The workflows are pipelines of algorithms with preconfigured parameter settings for proper handling of QIAseq Targeted DNA Panel read structure, including UMIs as well as components for trimming, mapping, realignment, variant calling, filtering and annotation of the detected variants and quality control reporting.

The workflow detects variants from two types of evidence. The Low Frequency Variant Detection tool detects variants by examining the aligned parts of the reads in the read mappings and the Structural Variant Caller supplements these calls by inferring variants from the unaligned parts of reads in the read mappings. Single nucleotide variants and shorter indels, including short tandem duplications, are typically present in aligned parts of the reads and are detected by the Low Frequency Variant Detection tool (Figure 1). These variants are presented in the "variants passing filters" output track.

Due to the limited read length (150 bp) of the Illumina reads, longer indels, including long tandem duplications, will not be present in the aligned parts of the reads; they will however be partially present in unaligned parts of mapped reads. The Structural Variant Caller examines read mappings for positions that show a significant proportion of reads with unaligned parts, calculates consensus sequences of those unaligned parts, performs de novo assembly between pairs

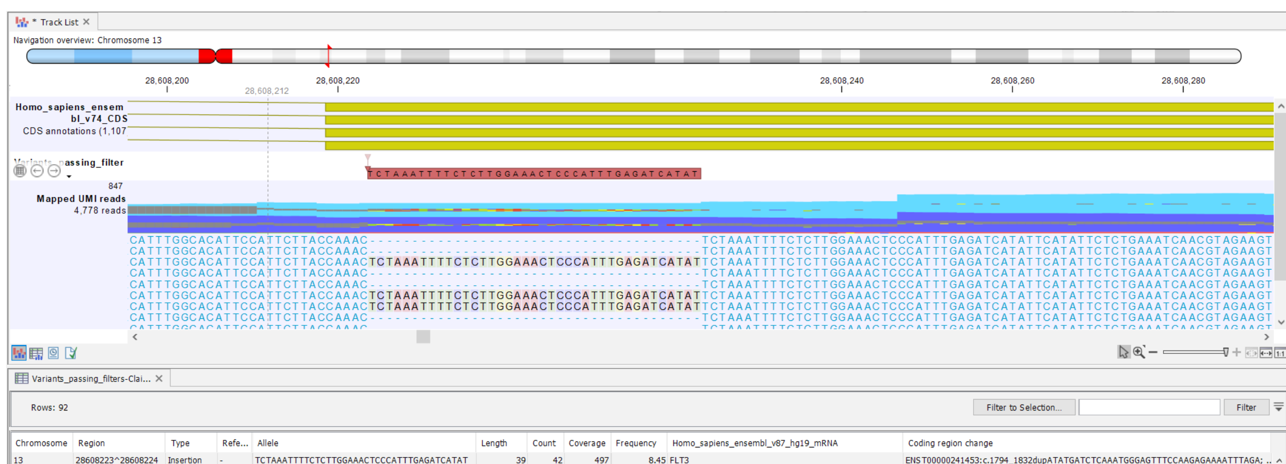

Figure 1. A 39 bp *FLT3* ITD called with the Low Frequency Variant Detection tool. Note the fully aligned reads and the absence of unaligned parts of reads.

of obtained consensus sequences and broken pair reads, and infers structural variants by comparing the observed alignment patterns to those expected for different types of structural variants (Figure 2, Figure 3).

Indels, including tandem duplications detected by the Structural Variant Caller are presented in the "Indels indirect evidence" track.

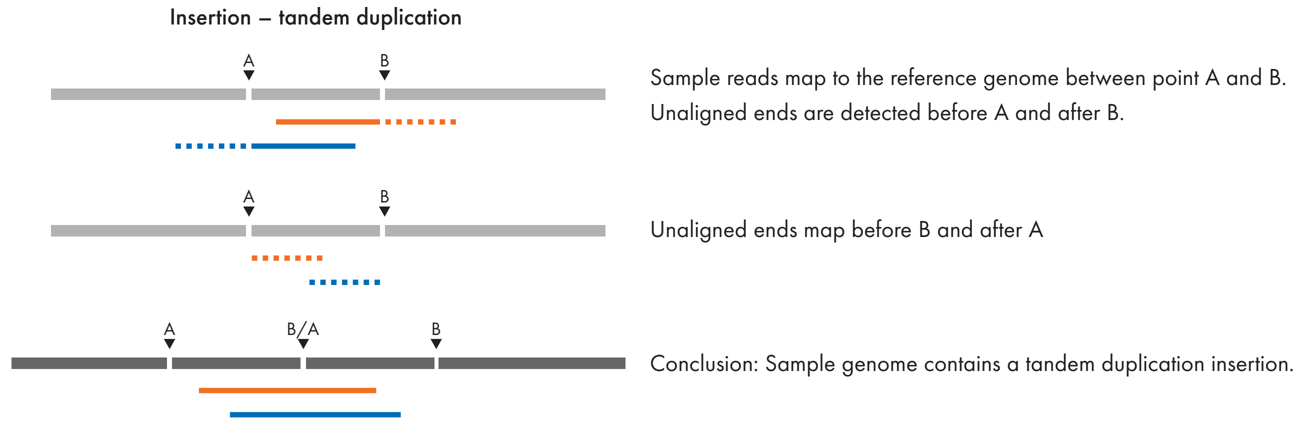

Figure 2. Steps in the detection of tandem duplication signatures (from the QIAGEN CLC Genomics Workbench manual).

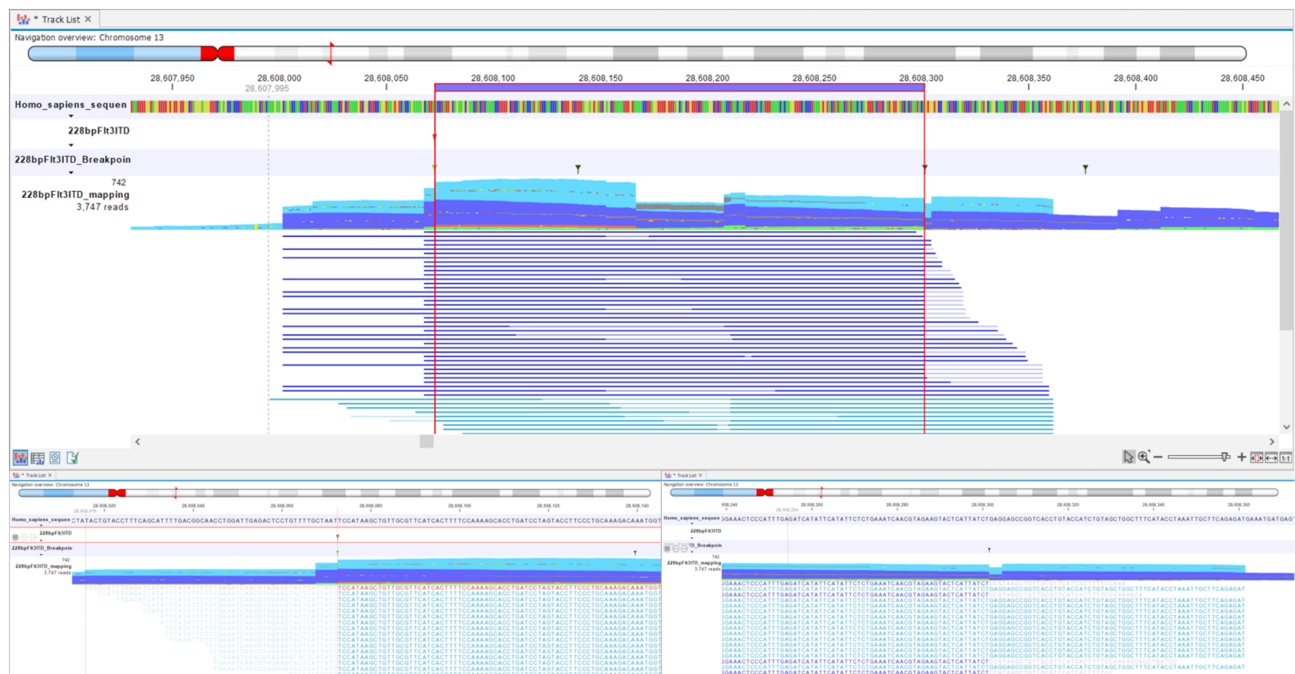

Figure 3. The signature of tandem duplication insertions in real read mappings. Note the unaligned parts of reads in the upper view before resolution of the ITD in the lower view.

Data

NGS libraries representing DNA from 34 human bone marrow samples from the Molecular Hematology Unit (Department of Laboratory Medicine, AZ Sint-Jan Brugge-Oostende AV – Campus Brugge; data courtesy of Matthijs Vynck and Friedel Nollet), were used to evaluate the performance of the *FLT3* ITD detection. The NGS libraries were prepared with a QIAseq Targeted DNA Custom Panel. The panel targets 21 genes of diagnostic and/or prognostic importance, including exons 13 to 15 of the *FLT3* gene, which were sequenced on MiSeq® or NextSeq® 550 Illumina instruments using 2 x 150 bp read-length protocols. The status of putative tandem duplications in the *FLT3* gene was established with PCR-CE. Insertions were detected in 24 samples, and the lengths of the insertion varied between 12 and 298 bp, whereas an additional 10 samples were negative for *FLT3* insertions.

The synthetically-spiked SeraSeq® Myeloid Mutation DNA Mix sample was used to generate an NGS library with the QIAseq Pan-cancer Multimodal Panel (cat no UHS-5000Z), sequenced at QIAGEN A/S and analyzed using the corresponding workflow in QIAGEN CLC Genomics Workbench.

Results

Samples were analyzed with the "Identify QIAseq DNA somatic variants (Illumina)" ready-to-use workflow in QIAGEN CLC Genomics Workbench 21.0.3 (Biomedical Genomics Analysis plugin, version 21.0.1). For each sample, the "Variants passing filters" and "Indels-indirect evidence" tracks produced were inspected for calls of ITDs in the *FLT3* gene. A "CLC *FLT3* ITD status call" was defined as "positive" (+) if an ITD in the *FLT3* gene was present in either the "Variants passing filters" or the "Indels indirect evidence" track. If no ITD in the *FLT3* gene was present in either of the tracks, the "CLC *FLT3* status call" was defined as "negative" (–).

Table 1 summarizes the calls, along with the PCR-CE results, for each of the 34 samples. The results indicate the following:

- There is 100% concordance between the *FLT3* ITD status calls achieved with PCR-CE and with NGS using the QIAseq Targeted DNA Custom Panel and QIAGEN CLC Genomics Workbench
- For 24 *FLT3* ITD-positive samples, there is very good agreement between the lengths of the ITDs detected with the two methods:
    - In 19 of 20 samples with ITDs lengths below 150 bp, the reported lengths are within 1 bp of each other
    - In 3 of 4 samples with ITD lengths above 150 bp, the reported lengths are within 5% of each other

Note that, for the 22 samples with good agreement in reported *FLT3* ITD lengths, 20 of the lengths reported by QIAGEN CLC Genomics Workbench are in frame (divisible by 3), whereas only 6 lengths determined by PCR-CE are in frame. Studies suggest that *FLT3* ITDs tend to be in-frame (1, 2). The lengths of the PCR-CE-reported ITDs are often 1 bp shorter than the lengths obtained through sequencing, likely an artifact of the PCR-CE method.

Table 1. Concordance of *FLT3* ITD status calls by orthogonal PCR-CE and NGS methods*

Sample	PCR-CE *FLT3* ITD length	PCR-CE *FLT3* ITD status	CLC *FLT3* ITDs detected (direct evidence)	CLC *FLT3* ITDs detected (indirect evidence)	CLC *FLT3* ITD status	Detected ITD length difference	Status call concordance	Length concordance†
1	12, 112	+	24	201, 81	+	12, 31, 89	+	No
2	65	+	66, 21	66, 105	+	1	+	Yes
3	56	+	–	57	+	1	+	Yes
4	256	+	–	126	+	130	+	No
5	94	+	–	96	+	–2	+	Yes
6	39	+	39	–	+	0	+	Yes
7	84	+	–	84	+	0	+	Yes
8	78	+	–	78, 74	+	0	+	Yes
9	235	+	–	228	+	7	+	Yes
10	66	+	–	66	+	0	+	Yes
11	68	+	–	69	+	–1	+	Yes
12	23	+	24	–	+	–1	+	Yes
13	52	+	–	51	+	1	+	Yes
14	51	+	–	51	+	0	+	Yes
15	51	+	52‡, 51, 24	–	+	0	+	Yes
16	17	+	18	–	+	–1	+	Yes
17	26	+	27‡	–	+	–1	+	Yes
18	160	+	–	154	+	6	+	Yes
19	103	+	–	105	+	–2	+	Yes
20	23	+	24	–	+	–1	+	Yes
21	47	+	48‡	43	+	–1	+	Yes
22	23	+	24	43	+	–1	+	Yes
23	26	+	27	–	+	–1	+	Yes
24	298	+	–	284	+	14	+	Yes
25	–	–	–	–	–	n.a.	+	Yes
26	–	–	–	–	–	n.a.	+	Yes
27	-	–	–	–	–	n.a.	+	Yes
28	–	–	–	–	–	n.a.	+	Yes
29	–	–	–	–	–	n.a.	+	Yes
30	–	–	–	–	–	n.a.	+	Yes
31	–	–	–	–	–	n.a.	+	Yes
32	–	–	–	–	–	n.a.	+	Yes
33	–	–	–	–	–	n.a.	+	Yes
34	–	–	–	–	–	n.a.	+	Yes

* CLC *FLT3* ITDs detected (direct evidence) refers to variants reported in the "Variants passing filters" table, whereas CLC *FLT3* ITDs detected (indirect evidence) refers to variants reported in the "Indels-indirect evidence" table produced by the workflow.

† Green: reported lengths are within +/– 1 bp for lengths below 150 bp or are within +/–5% of PCR-CE determined length for lengths above 150 bp; Red: Reported lengths are NOT within +/–1 bp for lengths below 150 bp or are NOT within +/–5% of PCR-CE determined length for lengths above 150 bp.

‡ Delins (not exact tandem duplication).

n.a.: Not applicable.

For the 34 bone marrow samples, the expected Variant Allele Frequency (VAF) is not known. To demonstrate that low allele frequency detection of *FLT3* ITDs is feasible using QIAGEN CLC Genomics Workbench, we also analyzed a synthetically-spiked SeraSeq Myeloid Mutation DNA Mix sample, which contains *FLT3* ITDs of lengths 33 bp at 10% VAF and 42 bp at 5% VAF, with the off-the-shelf QIAseq multi-modal panel and accompanying workflow. Both variants were detected by the **"Perform QIAseq Multimodal Analysis (Illumina)" workflow** (Figure 4 and 5).

Conclusions

NGS data generated from DNA libraries created using a QIAseq Targeted Panel in combination with bioinformatics analysis using QIAGEN CLC Genomics Workbench provides a powerful and easy-to-use solution for the analysis of ITDs in the *FLT3* gene. The analysis achieves 100% sensitivity and specificity in comparison with PCR-CE results. The sizes of the detected ITDs range from 18 to 284 nucleotides in length, and precision in the estimation of exact length using NGS likely exceeds that achieved by PCR-CE.

Chromosome	Region	Type	Reference	Allele	Reference al...	Length	Zygosity	Count	Coverage	Frequency	Probability	Forward rea...	Reverse rea...	Forward rea...	Reverse rea...	Forward/rev...	Average quality
13	28034112^2...	Insertion	-	TTTGAGATCA...	No	33	Heterozygous	120	1793	6.69	1.00	75	102	1086	1393	0.42	45.38
13	28034112^2...	Insertion	-	-	Yes	0	Heterozygous	1672	1793	93.25	1.00	1011	1290	1086	1393	0.44	45.12
13	28035676^2...	Insertion	-	A	No	1	Heterozygous	25	703	3.56	1.00	17	16	606	329	0.48	35.72

Figure 4. A 33 bp *FLT3* ITD in a SeraSeq Myeloid Mutation DNA Mix sample detected from direct evidence with the "Perform QIAseq Multimodal Analysis (Illumina)" workflow and present in the "Variants_passing_filters" track. Note the fully aligned reads and the absence of unaligned parts of reads.

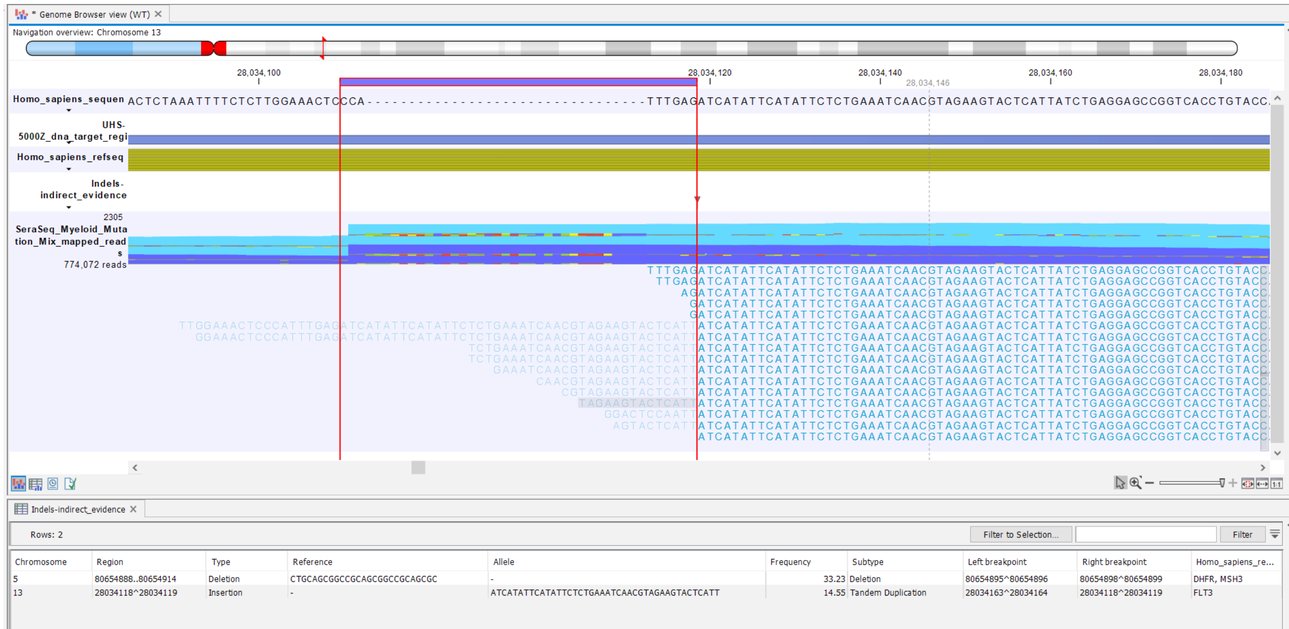

Figure 5. A 42 bp *FLT3* ITD in a SeraSeq Myeloid Mutation DNA Mix sample detected from indirect evidence with the "Perform QIAseq Multimodal Analysis (Illumina)" and present in the "indels-indirect_evidence" track. Note the presence of unaligned parts of reads. The ITD is detected by recognizing that the unaligned part of the reads, to the left of the selected area, align perfectly to the selected area.

References

1 Louise, M., et al. (2002). *FLT3* internal tandem duplication mutations associated with human acute myeloid leukemias induce myeloproliferative disease in a murine bone marrow transplant model. Blood **99**, 310.

2 Nakao, M., et al. (2000). Tandem duplication of the *FLT3* gene in acute lymphoblastic leukemia: a marker for the monitoring of minimal residual disease. Leukemia **14**, 522.

QIAseq Targeted Panels and QIAGEN CLC Genomics products are intended for molecular biology applications. These products are not intended for the diagnosis, prevention or treatment of a disease.

For up-to-date licensing information and product-specific disclaimers, see the respective QIAGEN CLC Genomics product site. Further information can be requested from ts-bioinformatics@qiagen.com or by contacting your local account manager.

Trademarks: QIAGEN®, Sample to Insight®, QIAseq® (QIAGEN Group); llumina®, MiSeq®, NextSeq® (Illumina, Inc.); SeraSeq® (SeraCare Life Sciences, Inc). Registered names, trademarks, etc. used in this document, even when not specifically marked as such, may still be protected by law. 09/2021 1125726 PROM-18837-001 © 2021 QIAGEN, all rights reserved.

Ordering **www.qiagen.com/shop/analytics-software** |
Technical Support **digitalinsights.qiagen.com/support** | Website **digitalinsights.qiagen.com**