Application Note

QIAseq™ Targeted Panels: accurately identify genetic variants with ease

Francesco Lescai, Bjarni J Vilhjalmsson, Anne-Mette Hein, Thomas Rask, Rune G. Madsen, Dmitrii Kamaev, Martin M. Johansen, Lise Husted, Raed Samara, Jean-Noel Billaud, Eric Lader, Arne Materna, Leif Schauser and Martin Simonsen

QIAGEN® Bioinformatics – Aarhus, Denmark

Introduction

QIAGEN combines a powerful chemistry based on unique molecular indices (UMIs) with a UMI-aware bioinformatics workflow in the Biomedical Genomics Workbench. Using DNA panels is a cost-effective approach to achieve the high coverage necessary for some challenging applications. In such scenarios, the capability to distinguish between sequencing or amplification errors and real findings is crucial in order to detect biologically relevant mutations at low allele fraction levels. UMIs address this challenge when combined with bioinformatics capable of exploiting their added value. Here we describe the features of the QIAseq Panel Analysis Plugin and how its workflow leverages UMIs to achieve impressive performance in detecting low allele fraction variants. Combining this technology with additional QIAGEN solutions, such as Ingenuity® Variant Analysis (IVA) and QCI™ Interpret, allows further exploration of the results for biological and pathological relevance.

About QIAseq Targeted Panel Analysis

QIAseq Targeted Panels, together with the Biomedical Genomics Workbench, provide a true Sample to Insight® solution for NGS-based, targeted low allele fraction variant detection in both somatic and germline cells.

QIAseq Targeted Panel Analysis uniquely combines state-of-the-art molecular biology with accurate and computationally efficient variant detection, and provides an unparalleled range of different mutation or DNA lesion types supported in a single solution.

The QIAseq Targeted Panel Analysis plugin supports QIAseq Targeted DNA Panels, QIAseq RNAscan Fusion panels and QIAseq RNA Expression panels with tailored workflows and a one-stop user-friendly interface. The analysis can be run locally on various operating systems, on a server or high-performance cluster, or in a virtual private cloud (VPC) environment. This ensures flexibility, scalability and data security – all of which are necessary in advanced research applications.

Detect any variant

Building on well-established CLC QIAGEN solutions, the QIAseq Targeted Panels Analysis plugin is capable to detect a broad range of different types of variants with competitive sensitivity and precision. Supported variant types include:

- Single nucleotide variants (SNVs)
- Multiple nucleotide variants (MNVs)
- Insertions and deletions (InDels)
- Copy number variants (CNVs)
- Gene fusions

Analyzing QIAseq data made easy

The QIAseq Targeted Panel Analysis plugin includes a user-friendly interface that guides the user through all necessary steps when selecting and initializing the analysis of QIAseq Targeted DNA, RNAscan and RNA Panel data (Figure 1).

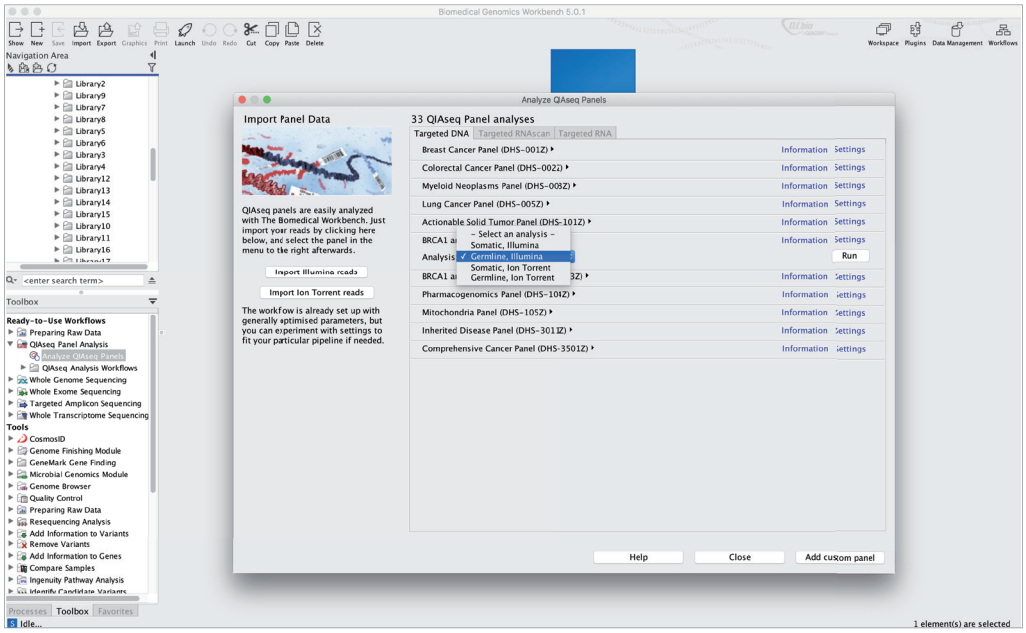

Figure 1. Screenshot of the QIAseq Panel Analysis guide.

For data analysis, the QIAseq Targeted Panel Analysis plugin relies on preconfigured, ready-to-use workflows. The workflow system in CLC Workbenches combines ease-of-use with reproducibility. The plugin comes with workflows designed for each panel type, that have been preconfigured for standard somatic and germline applications, for both Illumina® and Ion Torrent® sequencing platforms. However, users can easily duplicate and modify the workflows, adapting them to other sequencing applications. In addition, custom parameters and settings can be locked by the user to ensure reproducibility. Users can also inspect the reported variants in an interactive manner and perform additional analyses.

Increased variant calling accuracy via unique molecular indices (UMIs) and advanced bioinformatics

For enrichment, QIAseq panels use single primer extension (SPE) to avoid the limitations of 2-primer amplicon and nested PCR designs. With SPE, only a single primer is required to define a genomic target. Amplicon-based enrichment is facilitated by the use of a universal primer that binds to sequences introduced in the library adapters. Using SPE provides greater benefits, such as reducing the number of required primers, increasing enrichment and uniformity of sequencing, and allowing flexibility in designing the panel content. Additionally, NGS library preparation using QIAseq chemistry takes advantage of unique molecular indices (UMIs), which are also known as molecular bar codes. This technique reduces systematic error in the NGS process.

Tagging DNA and RNA with UMIs before any amplification takes place allows reads to be assigned to individual molecules, allowing for a computational correction of amplification bias and sequencing errors. Through the use of UMIs and the advanced UMI-aware algorithms in the QIAseq Targeted Panel Analysis plugin, the NGS error rate is reduced by up to 8-fold.

This combination of molecular biology and smart bioinformatics leads to more accurate reporting of variant allele fraction and expression level estimates.

Improved ability to detect hard-to-find variants

A common challenge for variant callers are the sensitive, yet precise, detection of somatic mutations that occur (for example, in liquid biopsies) at very low allele fractions. The QIAseq Panel Analysis workflow shows competitive performance even at allele fractions as low as 1%, detecting SNVs with a sensitivity of 92.4% and a precision of 96.3%. For insertions/deletions, known to be more difficult to identify, our workflow reaches a sensitivity of 89.8% and a precision of 91.7% (see section on “Benchmarking”, page 8).

Combined with the QIAseq Targeted Panel Analysis plugin, QIAseq Targeted Panels increase the chance of detecting hard-to-find variants. These include disease-specific variants such as the *CEBPA* deletion and the *CALR* deletion.

Example 1:

Single nucleotide variants (SNVs) and insertion/deletion variants (InDels) (Figure 2).

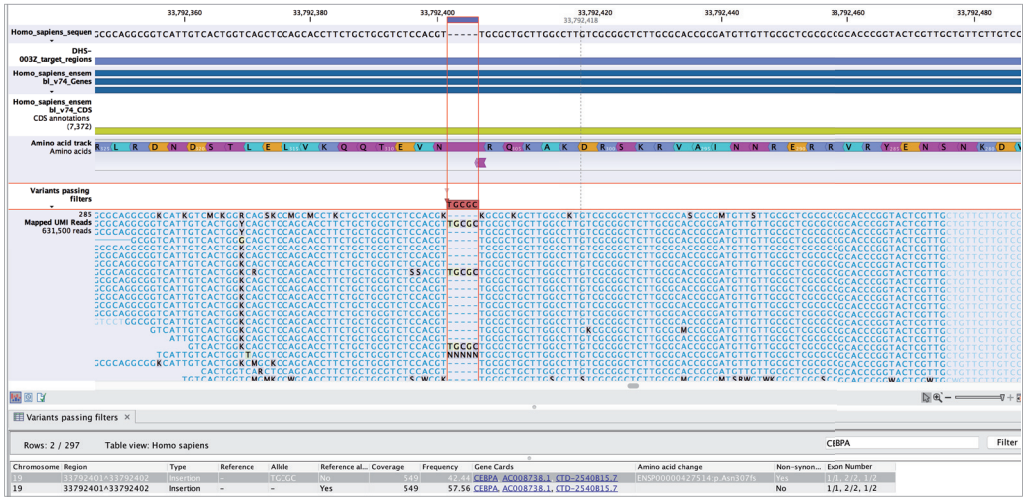

Figure 2. The Biomedical Genomics Workbench. Precisely detects single nucleotide variants (SNVs), multi nucleotide variants (MNVs) and, as depicted here, insertion/deletion type variants (InDels) as a result of competitive global and local alignment, and variant calling algorithms.

Example 2:

Large structural variants (Figure 3).

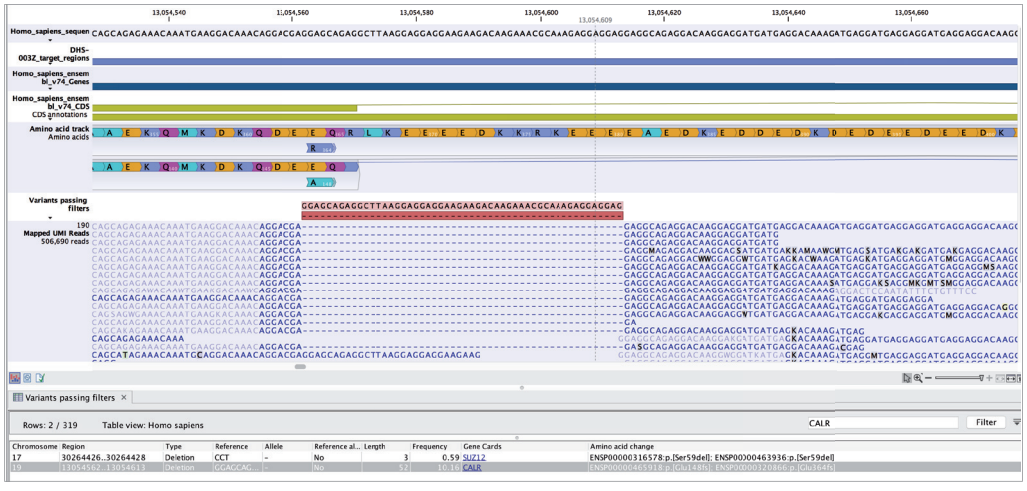

Figure 3: The Biomedical Genomics Workbench. Detects both the *CALR* (this figure) and the *CEBPA* (Figure 2, above) deletions from QIAseq Targeted DNA Panel data, despite being located in regions of low sequence complexity.

Detect gene fusions with confidence

A novel fusion gene detection algorithm is implemented in the QIAseq Targeted Panel Analysis plugin. The algorithm relies on two primary sources of evidence: reads that map across fusion gene breakpoints and paired reads where the left and right read map to each of the two genes. Leveraging the acclaimed QIAGEN CLC RNA-seq aligner (1), the algorithm offers high accuracy (F score = 0.95, see Table 2, page 8).

Ingenuity Variant Analysis-powered interpretation

Besides providing accurate variant detection, the Biomedical Genomics Workbench makes it easy to explore variants down to the read level. Once the variants have been detected, using a plugin in the Biomedical Genomics Workbench, the user will be able to easily send these variants for further biological exploration in Ingenuity Variant Analysis (IVA).

IVA is a secure (HIPAA- and Safe Harbor-compliant) web platform for annotating and comparing comprehensively sequenced human genomes in order to quickly shortlist candidate variants in studies of matched or unmatched tumors, disease kindreds, single- or multi-proband sets, or large case-control cohorts. Integration with IVA enables the user to characterize the identified variants with valuable disease insight, leveraging the QIAGEN knowledge-base. A few simple questions are asked at the start of analysis and after which, the platform will sensibly parameterize filters for finding credibly rare, appropriately functionally-suspect variants based upon study design, focus and assumptions. Spotting likely disease-causing drivers requires sensible filters to accurately answer three key questions about each putative variant: Is it real? How common is it among other tumors and in the world at large? And how might it affect physiology through gene product sequence and/or expression? IVA uses a default-configured, yet customizable, series of filters to answer these questions in order to shortlist candidate variants, genes and pathways (see **https://www.qiagenbioinformatics.com/products/ingenuity-variant-analysis/**).

The results can be uploaded directly (or via IVA) to QIAGEN Clinical Insight (QCI) for further evaluation.

Visualization tools

The Biomedical Genomics Workbench enables creation of straightforward visualizations of the user's results in the Genome Browser View (Figure 4). With a simple drag-and-drop, multiple tracks can be combined in order to view the results of variant filtering and annotation, targeted regions and information from a wide range of databases.

Figure 4. Genome Browser View. A 1.4% allele fraction deletion is displayed in its UMI reads context, targeted region and annotated databases like dbSNP and COSMIC, where it was previously described.

Benchmarking

QIAseq workflows provide competitive performance compared to other solutions in its ability to detect low allele fraction variants, which are particularly important in liquid biopsy applications and tumor sequencing. We benchmarked our solution with 2 other well-known applications and showed an improvement in the accuracy of the analysis.

Table 1 shows benchmarking results for a custom evaluation using the QIAseq DNA Panel on a NA0030 sample with 272 well-characterized variants, with 1% allele fraction in the target region (2). For the analysis, we used the QIAseq Targeted Panel Analysis plugin. MuTect and VarDict parameters were chosen to optimize results.

Table 1. Benchmarking results

Variant type	Method	TPs	FPs	FNs	Sensitivity (TPR)	Precision (PPV)	F1 Score (accuracy)
SNVs	**QIAseq Analysis**	206	8	17	92.4%	96.3%	0.94
	MuTect	214	58	9	96.0%	78.7%	0.87
	VarDict	204	169	19	91.5%	54.7%	0.68
InDels	**QIAseq Analysis**	44	4	5	89.8%	91.7%	0.91
	VarDict	43	100	6	87.8%	30.1%	0.45

Table 2 shows the results for detecting fusions in 7 samples using 4 catalog QIAseq Targeted RNAscan Panels and one custom panel. Five of the samples were HD784 and the other two were Seraseq™ FFPE Tumor Fusion RNA Reference Material v1 and v2. QIMERA is a fusion detection method based on STAR (3) and implemented in the GeneGlobe® pipeline for QIAseq Targeted RNAscan data.

Table 2. Detecting fusions results

Pipeline	TPs	FPs	FNs	Sensitivity (TPR)	Precision (PPV)	F1 Score (accuracy)
QIAseq Analysis	36	2	2	94.7%	94.7%	0.95
QIMERA	31	4	7	81.6%	88.6%	0.85

SNV detection as a function of coverage and workflow parameters

To study the impact of read coverage on the ability to detect low allele fraction variants, we analyzed three datasets with respect to sensitivity and two with respect to precision, including the one used in the SNV benchmark (Table 1, page 8). The reference samples used for sensitivity measures contained known variants with expected variant allele frequencies (VAF) down to 1%.

With default parameters, we show impressive sensitivity levels, even for variants expected around 1%, at coverage levels above 1000x (Figure 5). Specificity was calculated based on the dataset published by Xu et al. (2).

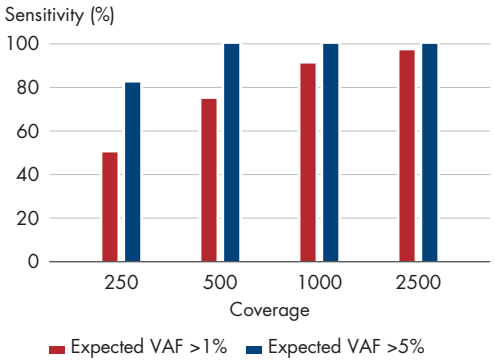

Figure 5. Sensitivity in calling variants at different levels of variant allele frequency (VAF) and coverage. The sensitivity of QIAseq Targeted Panel Analysis for DNA Panel, summarized across 2 datasets. The expected VAF depends on the fraction of DNA carrying the variant injected in the standard sample used for testing. The sensitivity is shown as a function of UMI-read coverage. Default parameters (i.e., frequency and QUAL thresholds of 0.5 and 200, respectively) were used. Both frequency and QUAL thresholds can be adjusted in the ready-to-use workflows provided in the QIAseq Targeted Panel Analysis plugin.

Similarly, the QIAseq Targeted Panels workflow performs very well in terms of precision. Figure 6 shows that the QIAseq Targeted Panels plugin workflow is capable of calling variants with precision levels in the range of 90–95% at variant allele frequency thresholds as low as 0.5% when the dataset has an appropriate UMI-read coverage. In our analyses, we recommend a UMI-read coverage of at least 1000x with default parameters (i.e., quality threshold 200 and frequency threshold 0.5%). In order to reach the desired UMI-read coverage, the user should aim at a higher sequencing coverage, which depends on the amount of input DNA and the amplification process (Source: *QIAseq Targeted DNA Panel Handbook*, 2nd ed. May 2017).

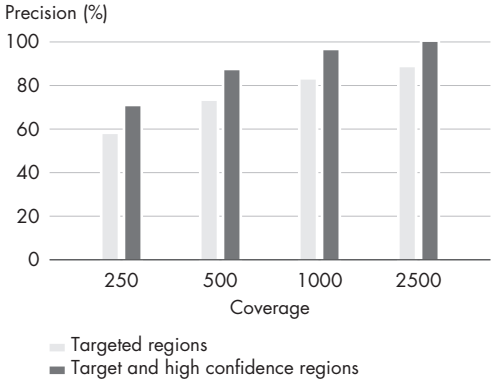

Figure 6. Precision in calling variants in different regions and coverage. The precision of the QIAseq Targeted Panel Analysis for DNA Panel, summarized across two datasets. The reported gold standard variant counts have been adjusted in order to take into account overlaps with the target regions captured by the QIAseq Panels (light gray), and those overlapping both targeted capture regions and gold standard high confidence regions (dark gray). When looking at high confidence regions, the precision is above 95% for UMI-read coverage above 1000x and a variant allele frequency threshold of 0.5% (default setting of the plugin).

Appendix: Methods

Samples

In order to assess the performance of the QIAseq DNA Targeted Panels workflow, both standard NA samples as well as commercially available samples have been used.

In particular, different mixtures of NA12878 and NA24385 (2% and 10%) were used to measure precision, due to the widely available information on the genetic background of the samples.

Commercially available Horizon™ Discovery samples HD780 (carrying variants at VAF of 1% and 5%) and HD700 (carrying variants at VAF between 0.1 and 10%) have been used to measure sensitivity, in order to specifically test the workflow against well-known disease-specific mutations.

In order to assess the performance of the QIAseq Targeted RNAscan Panels, commercially available samples have been used: HD784 from Horizon Discovery (3 gene fusions reported), Seraseq v1 (12 gene fusions reported), and Seraseq v2 (16 gene fusions reported, 13 of which overlap with our panels) from SeraCare®.

Targeted capture

The following QIAseq panels have been used in targeted DNA analysis: 101Z (actionable solid tumor pane, 22 genes), 003Z (Myeloid Neoplasm panel, 141 genes), 3501Z (comprehensive cancer panel, 275 genes). A custom panel (N0030) described in Xu et al. (2) has been used for the benchmarking of the workflow with other software (Table 1, page 8).

The following QIAseq panels have been used in targeted RNAscan analysis: FHS-001Z, FHS-002Z, FHS-003Z, FHS-3001Z and a custom panel (CFHS-10017Z-1198).

Analysis

The following standard measures have been used to assess performance, both in DNA and RNA workflows:

Precision (or PPV – Positive Predictive Value) = TP/(TP + FP)

Sensitivity (or Recall – also called TPR, True Positive Rate) = TP/(TP + FN)

True and false positives, as well as false negatives, counts have been made by overlapping the data with the target regions of the different panels. The final counts represent the aggregated counts of the different combinations between panels and samples, in both DNA and RNA workflows.

When counting false positives in NA reference samples for the DNA workflow, we classified as being “true positives” those variants consistently called in the samples and reported as common variants (i.e., normally shared in the population) by dbSNP.

To assess the relationship between precision and sensitivity with UMI-read coverage in the DNA workflow, the original sequencing data (at approximately 4000x UMI-coverage) have been random-sampled at different depths between 200x and 2500x. Three replicas at each coverage level have been created by independent random sampling and the same workflow has been applied to each dataset. The results at each coverage level show the average among the replicas.

In the RNAscan workflow, the fusion HACL1_COLQ 5′: 3 -(15563357^15563357), 3′: 3 –(15489637^15489637) has been consistently found in the same sample with different panels, and it has therefore been considered a true positive.

References

1. Baruzzo, G. et al. (2017) Simulation-based comprehensive benchmarking of RNA-seq aligners. Nat Methods **14**, 135–139.
2. Xu, C., Nezami Ranjbar, M. R., Wu, Z., DiCarlo, J. and Wang, Y. (2017) Detecting very low allele fraction variants using targeted DNA sequencing and a novel molecular barcode-aware variant caller. BMC Genomics **18**:5.
3. Dobin, A. et al. (2013) STAR: ultrafast universal RNA-seq aligner. Bioinformatics **29** (1), 15–21.

Discover more at **www.qiagenbioinformatics.com**.

To learn more, have a look at these informative tools:

Web resources

Biomedical Genomics Workbench:
www.qiagenbioinformatics.com/products/biomedical-genomics-workbench/

CLC Genomics Workbench
www.qiagenbioinformatics.com/products/clc-genomics-workbench/

CLC Microbial Genomics Module
www.qiagenbioinformatics.com/products/clc-microbial-genomics-module/

Tutorials

www.qiagenbioinformatics.com/support/tutorials/

For up-to-date licensing information and product-specific disclaimers, see the respective QIAGEN kit handbook or user manual. QIAGEN kit handbooks and user manuals are available at www.qiagen.com or can be requested from QIAGEN Technical Services or your local distributor.

Ordering **www.qiagen.com/shop** | Technical Support **support.qiagen.com** | Website **www.qiagen.com**

1113909 05/2018