Tutorial

Fusion Detection Using QIAseq RNAscan Panels

August 23, 2017

Sample to Insight

QIAGEN Aarhus · Silkeborgvej 2 · Prismet · 8000 Aarhus C · Denmark
Telephone: +45 70 22 32 44 · www.qiagenbioinformatics.com ·
AdvancedGenomicsSupport@qiagen.com

Fusion Detection Using QIAseq RNAscan Panels

This tutorial uses the capacities of the Biomedical Genomics Workbench and the FusQIAseq Targeted RNAscan Panel Analysis plugin to detect fusion events in sequences generated using a QIAseq Targeted RNAscan Panel.

This tutorial covers in just a few steps all the following:

- Import Illumina paired reads in the workbench.
- Create a trim adaptor list to remove the sequencing primers off the reads.
- Find fusion events with the **Detect QIAseq RNAscan Fusions** ready-to-use workflow.

This tutorial makes use of a subset (reads mapping only to chromosome 10) of the sequencing reads obtained using a QIAseq Targeted RNAscan Panel.

Prerequisites

For this tutorial, you must be working with the Biomedical Genomics Workbench 4.0 or higher. You must also have installed the QIAseq DNA V3 Panels Analysis plugin and the QIAseq Targeted RNAscan Panel Analysis plugin version 0.8.1 or higher. How to install plugins is described here: http://resources.qiagenbioinformatics.com/manuals/biomedicalgenomicsworkbench/current/index.php?manual=Installing_plugins.html.

Download and import data

Go through the following steps to download and import the data into the Workbench.

1. Download the sample data from our website: http://resources.qiagenbioinformatics.com/testdata/Fusion_Gene_Detection_tutorial.zip and unzip it in the location of your choice on your computer.
2. Start the *CLC Genomics Workbench*.
3. Import the reads via the toolbar: **File** | **Import** () | **Illumina**
   - Select the two fastq paired files.
   - Under "General Options" section, ensure the **Paired reads** and **Discard read names** checkboxes are checked.
   - In the "Paired read orientation" section, ensure the **Paired-end (forward-reverse)** option is checked.
   - Set the **Minimum distance** to 1 and the **Maximum distance** to 1000 (default values).
   - Click **Next**.
4. Click on the button labeled **Save** in the wizard page that appears, choose the folder you wish to save the reads to (you can create a new folder dedicated to this tutorial for example) and click **Finish**.

Data management configuration

In order to do this tutorial, you need to use the Reference Data Set made specifically for this tutorial:

1. Click on the Data Management icon in the top right corner of the workbench (1) (figure 1).
2. In the QIAGEN Reference Data Library (2), open the Tutorial Reference Data Sets tab (3), select the **Fusion Detection Using QIAseq RNAscan Panels** data set, and click on the button **Download** (4).

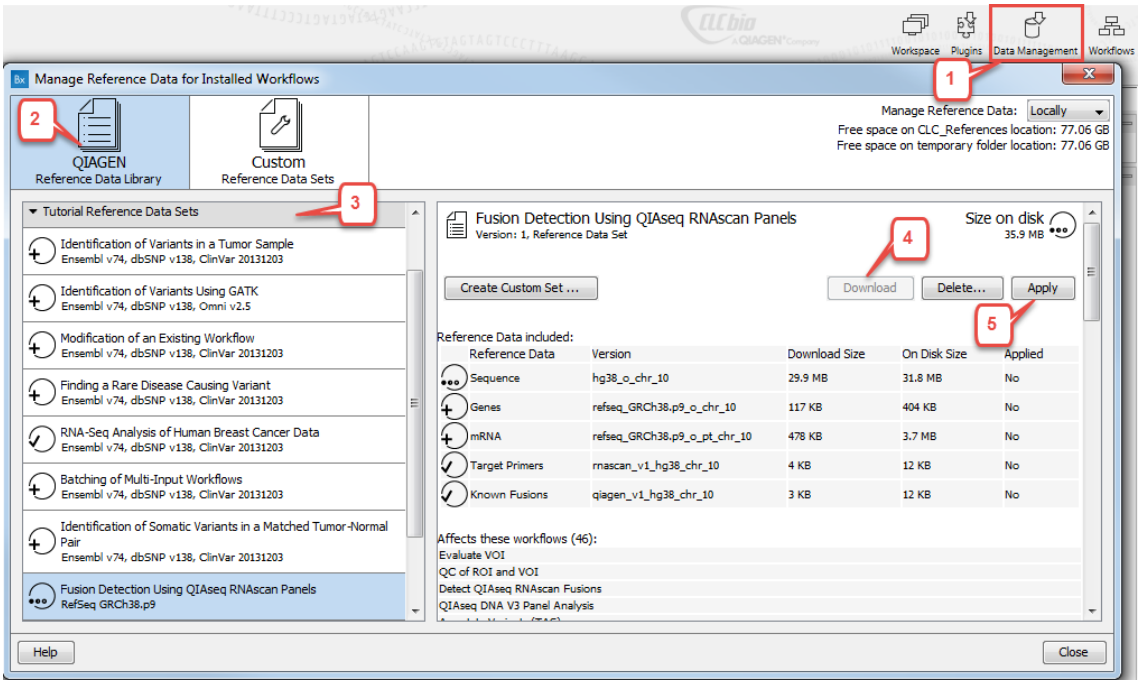


Figure 1: *Downloading the reference data.*

3. When the data is downloaded (characterized by check icons next to each item included in the dataset), click "Apply" (5).

The workflow is now set up to work with the adequate references. In addition, QIAseq Targeted RNAscan Panels specific primer tracks are now available in the CLC_References folder located in the Navigation Area.

Prepare your Adaptor trim list

As remaining adaptor sequences in the reads might lead to bias in downstream data analysis, we recommend to trim them off using a Trim adaptor list.

To create the list, use the oligonucleotide sequences disclosed in the Illumina Customer Sequence Letter found at this address: http://support.illumina.com/downloads/illumina-customer-sequence-letter.html. [1]

[1] Oligonucleotide sequences © 2007-2013 Illumina, Inc. All rights reserved. No sponsorship or affiliation. Link provided for convenience. QIAGEN not responsible for content at link.

1. Open the workbench and go to **File** | **New** | **Trim Adapter List (**)

2. Click on the button **Add Row** ( ) found at the bottom of the View Area in the New Adapter Trim List.

3. Name the adapter Index 1. Then copy and paste from the letter the Nextera Transposase Adapter Read 1 sequence. Adjust "Strand" to **Minus** and "Action" to **Remove adapter**. Set **Mismatch cost** at 2 and **Gap cost** at 3. Check **Allow internal matches** and set its Minimal score to 20; Check also **Allow end matches** and set the Minimum score to 15 (figure 2). Click **Finish** to create the adapter trim list.

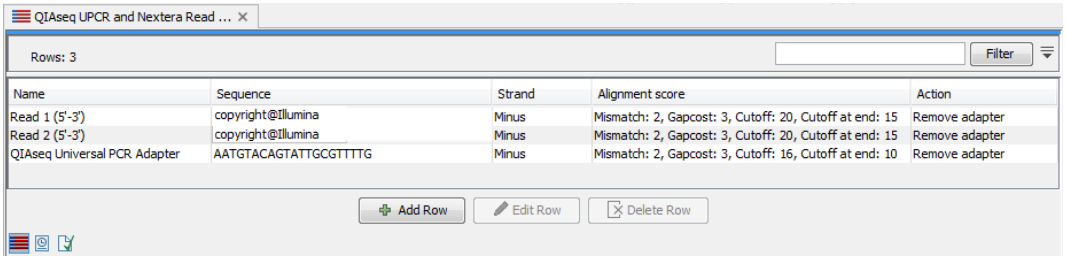


Figure 2: *Creating the adapter list.*

4. Click on the button **Add Row** ( ) again.

5. Name the new adapter Index 2. Copy and paste from the letter the Nextera Transposase Adapter Read 2 sequence. Adjust "Strand" to **Minus** and "Action" to **Remove adapter**. As previously, set **Mismatch cost** at 2 and **Gap cost** at 3. Check **Allow internal matches** and set its Minimal score to 20; Check also **Allow end matches** and set the Minimum score to 15. Click **Finish** to add the adapter to the list.

6. Finally, add the QIAseq Universal PCR Adapter AATGTACAGTATTGCGTTTTG. As previously, adjust "Strand" to **Minus** and "Action" to **Remove adapter**. As previously, set **Mismatch cost** at 2 and **Gap cost** at 3. For the tutorial, we will make the trim more stringent than set by the default parameters: check **Allow internal matches** and set its Minimal score to 16; Check also **Allow end matches** and set the Minimum score to 10. Click **Finish** to add the adapter to the list.

7. Go to **File** in the menu bar and choose to **Save** the generated adapter trim list in the Navigation Area.

Run Detect QIAseq RNAscan Fusions

The Detect QIAseq RNAscan Fusions ready-to-use workflow can be found in the Toolbox here:

Toolbox | **Ready-to-Use Workflows** | **QIAseq Panel Analysis** | **Detect QIAseq RNAscan Fusions**

1. Double-click on the Detect QIAseq RNAscan Fusions ready-to-use workflow to run the analysis.

   If the QIAseq Targeted RNA Panel Analysis plugin has been installed on a CLC Server you are connected to via the workbench, you will be asked where you would like to run the analysis. We recommend that you run the analysis on a CLC Server when possible.

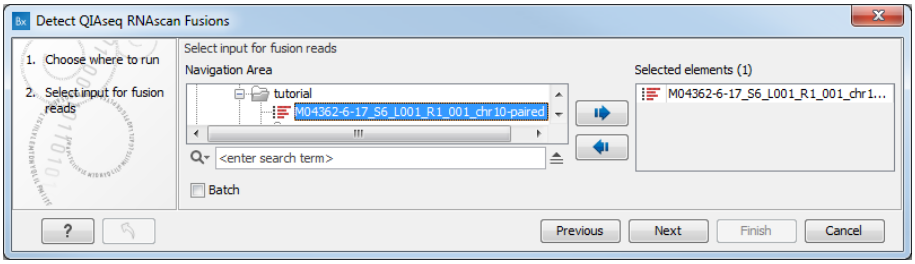


Figure 3: *Select the sequencing reads by double-clicking on the file name or by clicking once on the file name and then on the arrow pointing to the right hand side.*

2. Select the sequencing reads that should be analyzed (figure 3).
3. Select the **Trim adapter list** you created to trim the remaining PCR adapters off your reads (figure 4).

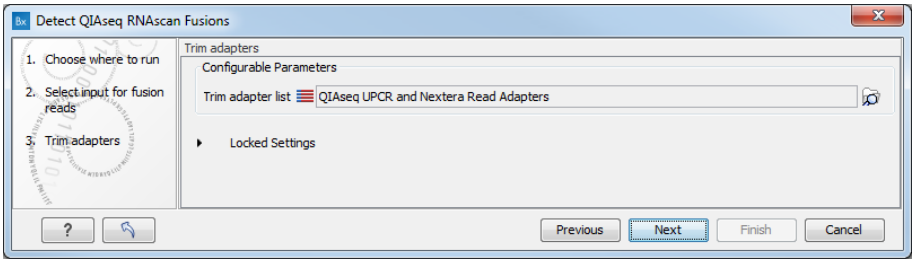


Figure 4: *Select the Trim adapter list you generated ahead of running the workflow.*

4. The tutorial Reference Data Set only contains the primer track specific to the panel you used to generate the sequenced reads (figure 5), so you do not need to select a specific track form a drop down menu.

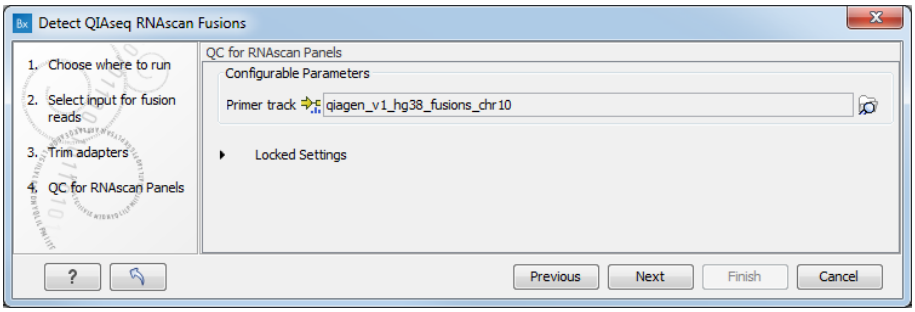
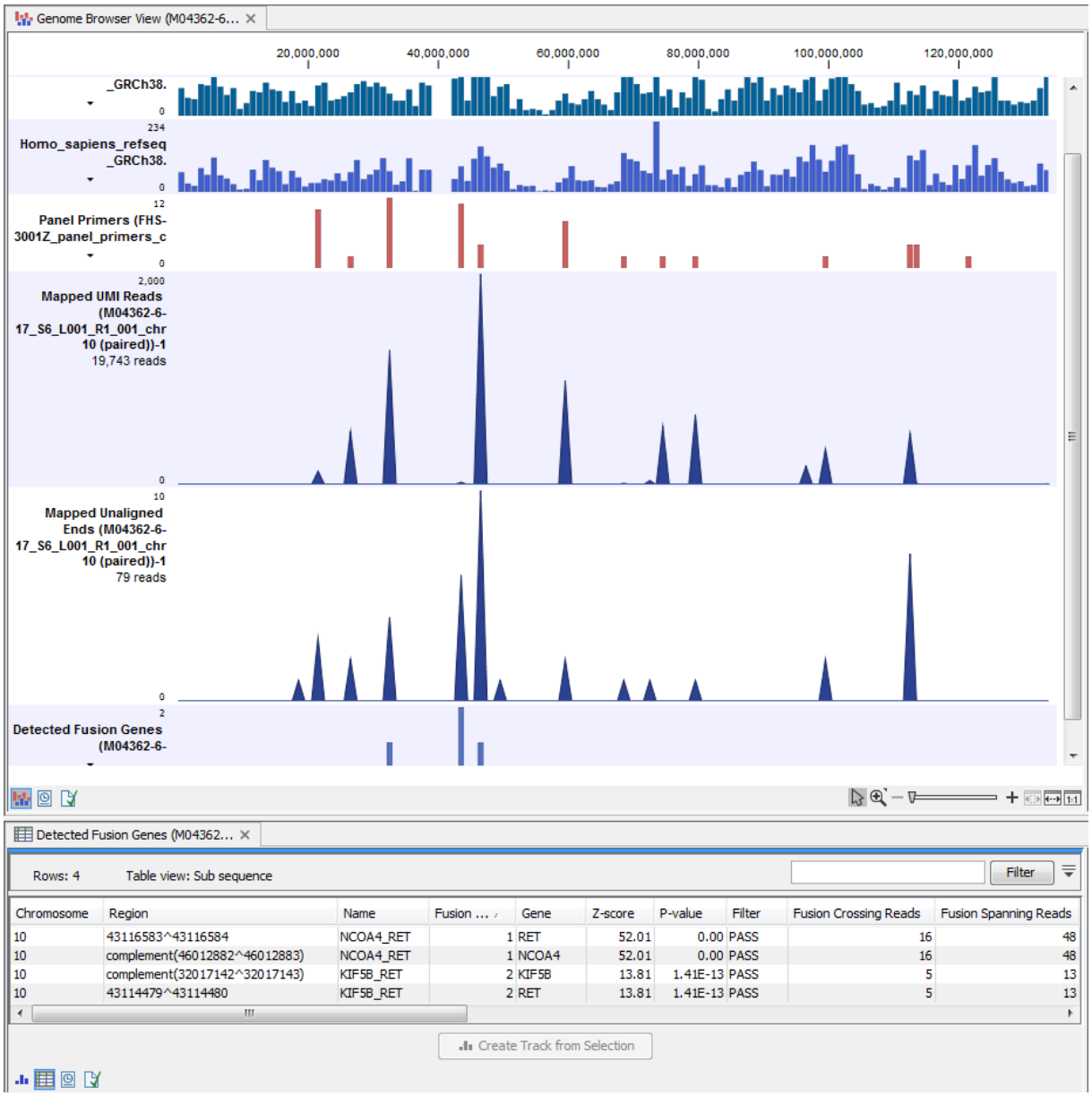


Figure 5: *Select the primer file you imported ahead of running the workflow.*

5. Finally, in the last wizard step, choose to **Save** the results of the workflow and specify a location in the Navigation Area before clicking **Finish**.

Output from the Detect QIAseq RNAscan Fusions workflow

The Detect QIAseq RNAscan Fusions workflow produces a series of output. The easiest way to review the results is to open the Genome Browser View (figure 6).



Chromosome	Region	Name	Fusion ...	Gene	Z-score	P-value	Filter	Fusion Crossing Reads	Fusion Spanning Reads
10	43116583^43116584	NCOA4_RET	1	RET	52.01	0.00	PASS	16	48
10	complement(46012882^46012883)	NCOA4_RET	1	NCOA4	52.01	0.00	PASS	16	48
10	complement(32017142^32017143)	KIF5B_RET	2	KIF5B	13.81	1.41E-13	PASS	5	13
10	43114479^43114480	KIF5B_RET	2	RET	13.81	1.41E-13	PASS	5	13

Figure 6: *A Genome Browser View showing the NCOA4_RET fusion.*

Before checking the detected fusions, users are usually recommended to review the QC for RNAscan Panels Report, and in particular, to verify that no DNA contamination was found in the sample. Also the four reference genes included in the RNAscan panels should have been detected and their expression to that of the target genes should be reviewed (Target gene vs reference gene coverage ratio [%]). In the case of this tutorial, the data was limited to chromosome 10 only and the QC for RNAscan Panels Report is therefore incomplete.

Now open the Genome Browser View, and double-click on the Detected Fusion Genes track name (to the left of the Genome Browser View) while pressing the Ctrl button. The fusion track will open as a table in split view, below the Genome Browser View. You can sort and filter the table using one of the metrics. Fusions that get the filter annotation PASS have not been caught by any filter and are therefore the most trustworthy fusions. Sorting the table using the "Fusion

Rank" column will move the most trustworthy fusions to the top of the table. Finally, clicking on a fusion event in the table will zoom in to its location in the UMI read mapping, allowing you to review the UMI reads supporting the detected fusion. Note that each fusion event is represented by two rows in the table, allowing you to visualize successively both breakpoints of each fusion. You can estimate the evidence for the fusion by studying the Mapped UMI Reads that aligns to one side of a fusion, and the Mapped Unaligned Ends that, after re-mapping, align to the other side of the fusion.

Note that the Mapped UMI track displays the mapping of UMI reads to the transcript sequences. Paired reads still in proper pairs are blue. Forward reads are green, reverse reads are red. Unaligned ends of mapped reads are shown in a lighter shade of the respective read color. Multi-match reads, i.e., reads that could have been mapped to another position within the mapping criteria used, are in yellow.