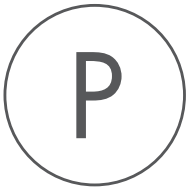

Ingenuity Variant Analysis Plugin

USER MANUAL

User manual for
Ingenuity Variant Analysis 20.0

Windows, macOS and Linux

December 5, 2019

This software is for research purposes only.

QIAGEN Aarhus
Silkeborgvej 2
Prismet
DK-8000 Aarhus C
Denmark

Contents

Chapter 1

Introduction

Ingenuity Variant Analysis plugin provides the ability to carry out an Ingenuity Variant Analysis on variant tracks generated in the Workbench, and to annotate and filter variants based on information present in the Allele Frequency Community. The latest available content from the Ingenuity Knowledge Base and Allele Frequency Community is used in the biological interpretation of input variants from whole genome, whole exome, targeted amplicon, or whole transcriptome sequencing experiments. By using published biological knowledge of disease biology, the Ingenuity Variant Analysis plugin can be used to prioritize your variants. The purpose of the integration is to supplement the abilities of the Workbench with the biological knowledge available in Ingenuity Variant Analysis and the Allele Frequency Community.

The plugin bundles four tools, which can be found in the Ingenuity Variant Analysis folder in the Toolbox:

- Ingenuity Variant Analysis, used to analyze cancer genomes or to carry out stratification analysis
- Ingenuity Variant Analysis for Hereditary Diseases, used to analyze genetic diseases
- Add Information from Allele Frequency Community, used to annotate with information from the Allele Frequency Community database
- Remove Variants found in Allele Frequency Community, used to filter out variants that are present in the Allele Frequency Community database

In addition to the four tools, the plugin comes with three ready-to-use workflows installed under the respective applications in the toolbox (Whole Genome Sequencing or Whole Exome Sequencing).

- Identify and Interpret Causal Variants in a Trio using IVA (WGS)
- Identify and Interpret Causal Variants in a Trio using IVA (WES)
- Identify Somatic Variants from a Single cfDNA Sample using IVA (WES)

Furthermore, the plugin includes the possibility to update a variant track containing the results of an Ingenuity Variant Analysis, if you change the filtering settings inside the Ingenuity Variant Analysis web interface.

Access to Ingenuity Variant Analysis requires a subscription. However a one-month trial period is available and allows the analysis of up to 4 samples. The first step is to register for an Ingenuity Variant Analysis account https://apps.ingenuity.com/isa/account/signup/va?utm_source=ingenuity&utm_medium=banner&utm_campaign=webpage-preview. Upon completion of registration, you will receive an email to activate your new Ingenuity Variant Analysis account. Once you've logged in for the first time and accepted the End User License Agreement, you can use these credentials to allow the plugin to send variant data from CLC Genomics Workbench to Ingenuity Variant Analysis.

If you opt into the Allele Frequency Community, you will get a month of free analysis without a subscription to Ingenuity Variant Analysis. See chapter 8 to change your Allele Frequency Community opt-in status.

Chapter 2

Ingenuity Variant Analysis

To start the tool, go to **Ingenuity Variant Analysis | Ingenuity Variant Analysis**

If you are connected to a server, you will first be asked about where you would like to run the analysis. If you are not connected to a server, the first step is to specify the input for the analysis. The Ingenuity Variant Analysis tool accepts variant tracks ( ) as input. Select the desired variant track or several variant tracks as input, as shown in figure 2.1. All the variant tracks selected in this step are assumed to be "case" samples.

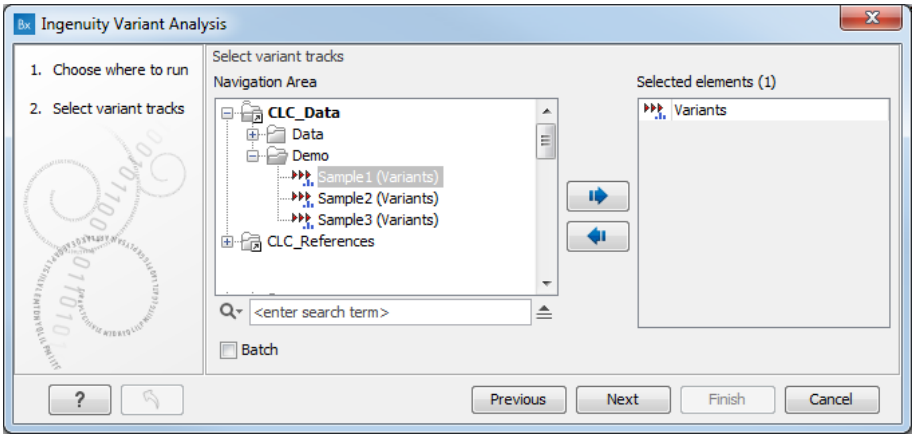

Figure 2.1: *Select the variant track that you would like to analyze.*

In the following dialog, you can set the Variant analysis parameters (figure 2.2).

- **Reference**: Select the human reference sequence that is found under CLC_References in the Navigation Area. Complete human genomes (e.g., hg19 (GRCh37) and hg38 (GRCh38)) and subsets of these (e.g. individual chromosomes) can be used as references.
- **Control tracks**: This is an optional parameter. You may select one or more variant tracks, which will be considered to be "control" samples in the analysis.
- **Analysis pipeline name**: Select the pipeline appropriate for your analysis. The Ingenuity Variant Analysis is performed with predefined settings that differ depending on your choice of analysis pipeline. The following options are available:

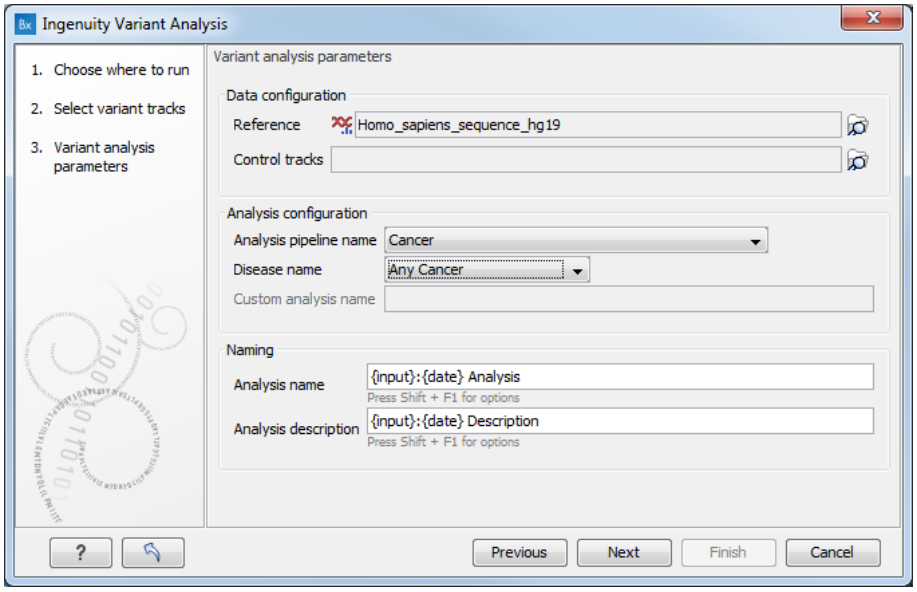

Figure 2.2: *Specify analysis parameters.*

  - Cancer: Useful if you are seeking to identify cancer driver variants. If this option is selected, the type of cancer must be specified in the **Disease name** drop-down menu.
  - Stratification study: Useful if you have two groups of samples and are looking for variants that distinguish the two.
  - Variant Analysis Custom Pipeline: Useful if you have already carried out an Ingenuity Variant Analysis, where you have set up a desired filtering cascade, and want to re-use the same filtering cascade for a new analysis. If this option is selected, the name of the custom analysis must be specified in the **Custom analysis name** field. The name you enter in the **Custom analysis name** field must match the "Name" field of an existing analysis in Ingenuity Variant Analysis, exactly as it appears in the Ingenuity Variant Analysis web interface. Note: please provide unique names to all your custom analyses in Ingenuity Variant Analysis: you might not be able to reuse a filtering cascade if its name was already used previously, even in cases where that previous analysis was already deleted.
  - Upload only: Useful if you just want to upload samples and do not wish to carry out an analysis. Note: in this case, no results will be downloaded.

- **Analysis name**: The name of the analysis. You can enter a name of your own choice by typing in the name, or by using the options that appear when you press Shift + F1. The available shorthand notations are: {input} will be substituted with the name of the input experiment, {date} is substituted with a date stamp, and {pipeline} is substituted by the Analysis pipeline name. The analysis name is the name that is shown on the Ingenuity Variant Analysis page when you choose "My Analyses". The same analysis name can furthermore be used in a Variant Analysis Custom Pipeline, if you specify it in the **Custom analysis name** field (see above).

- **Analysis description**: This will be the description of the analysis in Ingenuity Variant Analysis once created. There are a few shorthand notations available: {input} will be substituted with the name of the input experiment, {date} is substituted with a date stamp.

In the "Account information" dialog, enter your Ingenuity username (email address) and password (figure 2.3). If you do not have an Ingenuity username or password, you must first create an Ingenuity account (see how in the introduction to chapter 1).

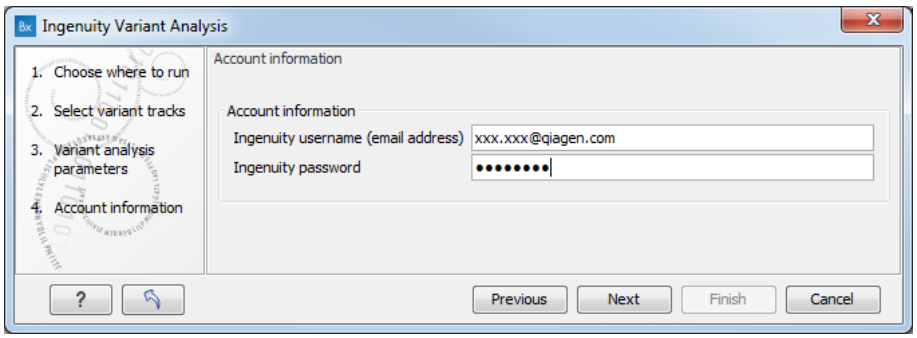

Figure 2.3: *Specify the account information: your Ingenuity username (email address) and password are required.*

The "Result handling" dialog allows you to set the output options (figure 2.4).

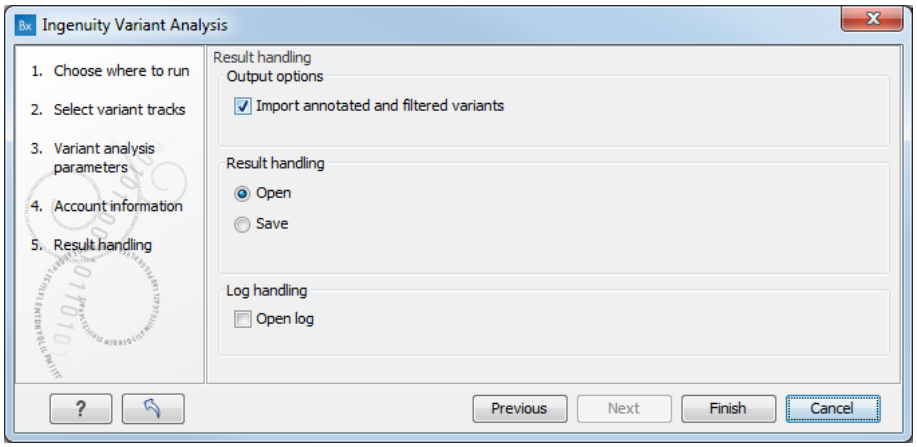

Figure 2.4: *The result handling step in the Ingenuity Variant Analysis wizard.*

If the **Import annotated and filtered variants** option is checked, the tool will produce a variant track as output. If it is unchecked, the analysis will be created, and can be accessed inside the Ingenuity Variant Analysis web interface, but the results will not be imported into the workbench. Note: it is not possible to import results if you have selected the **Upload only** pipeline earlier in the wizard.

If you choose to **Open** the results the two generated outputs will be opened in the View Area without being saved. In this case you will have to manually save the outputs if you would like to keep them. If you choose to **Save** the outputs, click Next to specify where to save the results. Click **Finish** to start the Ingenuity Variant Analysis. Your results will not be opened automatically but will be saved at the destination you have specified.

The outputs and how to manually adjust filter settings are described in section 5.

Chapter 3

Ingenuity Variant Analysis for Hereditary Diseases

To start the tool, go to:

Ingenuity Variant Analysis | Ingenuity Variant Analysis for Hereditary Diseases

If you are connected to a server, you will first be asked about where you would like to run the analysis. If you are not connected to a server, the first step is to specify the input for the analysis. The Ingenuity Variant Analysis for Hereditary Diseases tool accepts a single variant track ( ) as input. Select the proband variant track, i.e., the individual affected by the disease you are studying, as shown in figure 3.1.

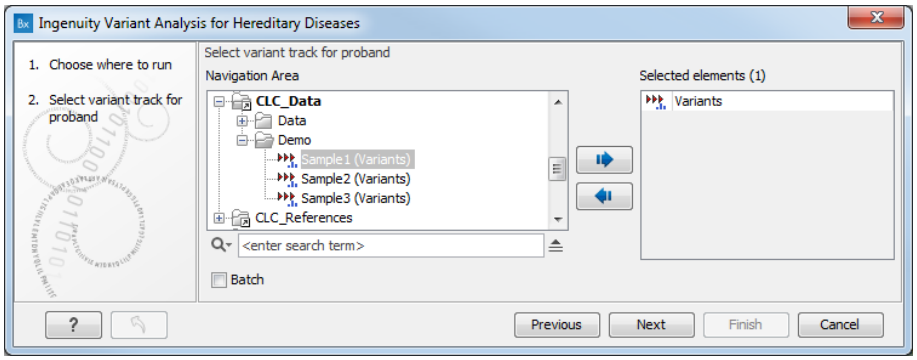

Figure 3.1: *Select first the proband variant track.*

You can then set the analysis parameters (figure 3.2).

- **Reference**: Select the human reference sequence that is found under CLC_References in the Navigation Area. Complete human genomes (e.g., hg19 (GRCh37) and hg38 (GRCh38)) and subsets of these (e.g. individual chromosomes) can be used as references.
- **Upload only**: useful if you just want to upload samples and do not wish to carry out an analysis. Note: in this case, no results will be downloaded.
- **Variant Analysis Custom Pipeline**: useful if you have already carried out an Ingenuity Variant Analysis where you had set up a desired filtering cascade and want to re-use the

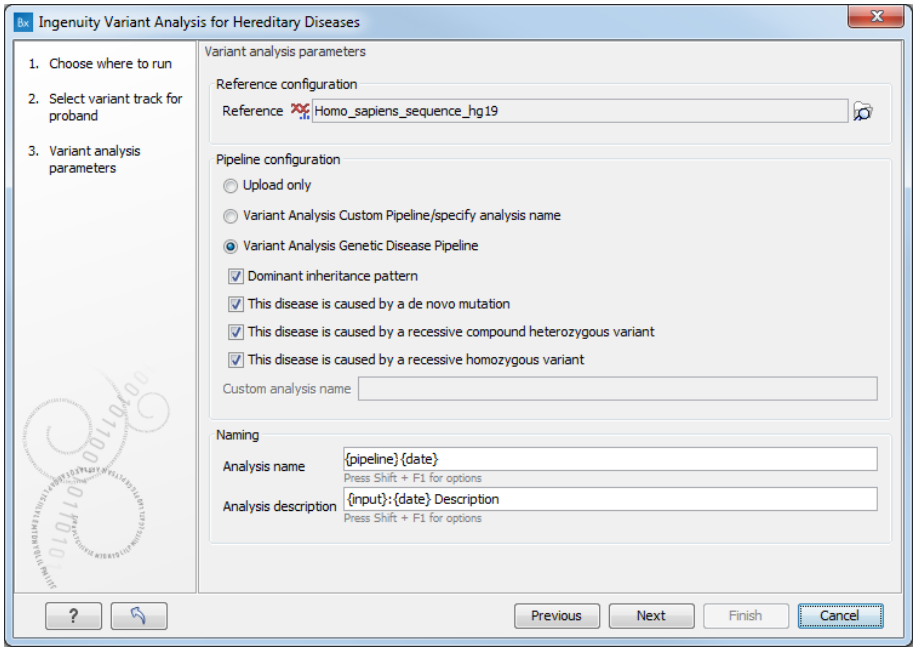

Figure 3.2: *Specify the analysis parameters.*

same filtering cascade for a new analysis. If this option is selected, the name of the custom analysis must be specified in the **Custom analysis name** field. The name you enter in the Custom analysis name field must match the "Name" field of an existing analysis in Ingenuity Variant Analysis exactly as it appears in the Ingenuity Variant Analysis web interface. Note: please provide unique names to all your custom analyses in Ingenuity Variant Analysis: you might not be able to reuse a filtering cascade if its name was already used previously, even in cases where that previous analysis was already deleted.

- **Variant Analysis Genetic Disease Pipeline**: to be used if you are studying genetic disease. Learn more about the four genetic disease pipelines in section 3.1.
  - Dominant inheritance pattern
  - This disease is caused by a de novo mutation.
  - This disease is caused by a recessive compound heterozygous variant.
  - This disease is caused by a recessive homozygous variant.
- **Analysis name**: The name of the analysis. You can enter a name of your own choice by typing in the name, or by using the options that appear when you press Shift + F1. The available shorthand notations are: {input} will be substituted with the name of the input experiment, {date} is substituted with a date stamp, and {pipeline} is substituted by the Analysis pipeline name. The analysis name is the name that is shown on the Ingenuity Variant Analysis page when you choose "My Analyses". The same analysis name can furthermore be used in a Variant Analysis Custom Pipeline, if you specify it in the **Custom analysis name** field (see above).
- **Analysis description**: This will be the description of the analysis in Ingenuity Variant Analysis once created. There are a few shorthand notations available: {input} will be substituted with the name of the input experiment, {date} is substituted with a date stamp.

The "Family information" dialog allows you to specify family data for the analysis (figure 3.3).

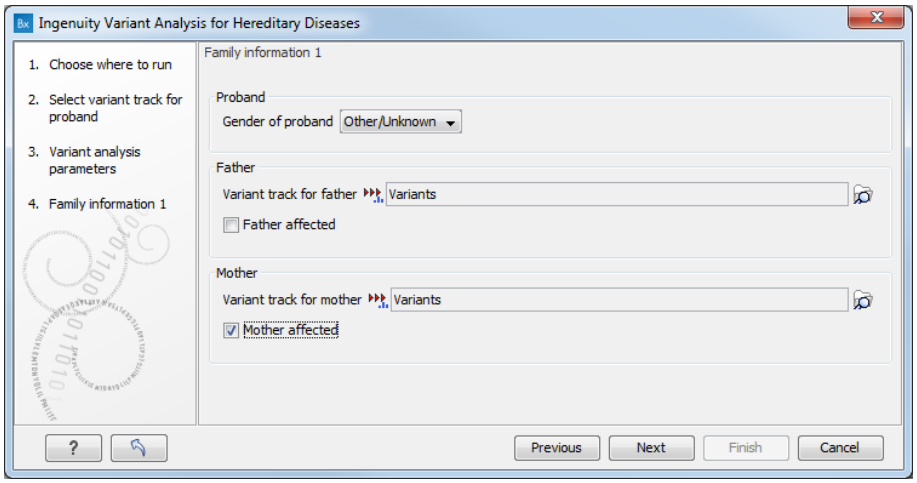

Figure 3.3: *Specify family data for your analysis. Data for at least one parent must be specified at this step*

- **Gender of proband**: Select the gender of the individual affected by the disease.
- **Variant track for father/mother**: Select the variant track for the father or mother, as appropriate. The variant track for at least one parent must be specified.
- **Father/mother affected**: Once you have selected a variant track for a parent, the option to set the disease status of that parent will be enabled. Check this box if the given parent is affected by the same disease as the proband. Uncheck this box if the given parent is not affected.

A second "Family information" dialog gives you the possibility to specify further family data for the analysis (figure 3.4). Similarly to the previous step, you can specify for each sibling a variant track, disease status and gender. However, check the pipeline of interest in section 3.1, as adding non-supported information for a specific pipeline might alter the results.

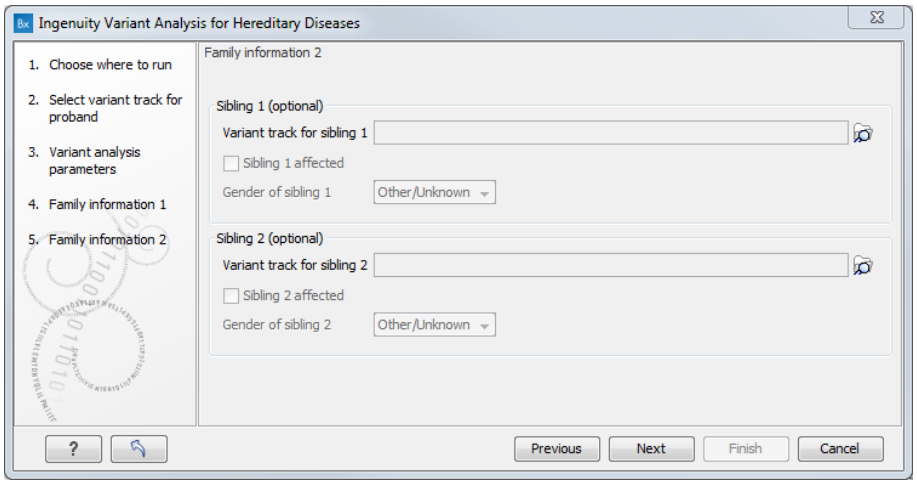

Figure 3.4: *At this step, you have the option to specify variant tracks for siblings of the proband.*

In the next dialog you must specify your Ingenuity username (email address) and password (figure 3.5). If you do not have an Ingenuity username or password, you must first create an Ingenuity account (see the introduction of chapter 1 to learn how).

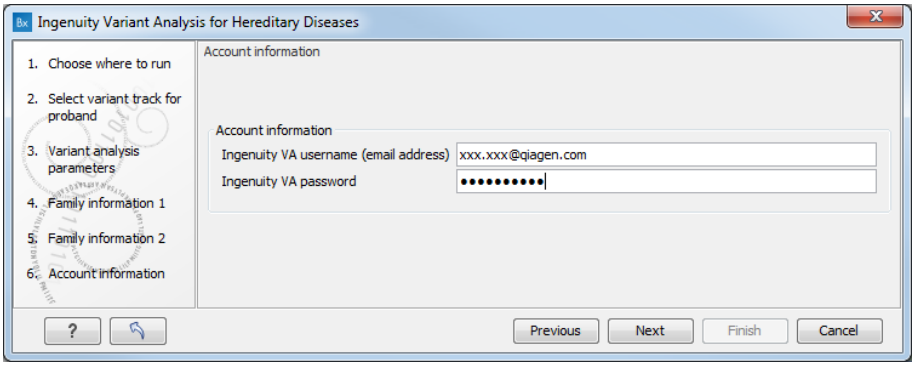

Figure 3.5: *Specify the account information: your Ingenuity username (email address) and password are required at this step.*

In the "Result handling" wizard step (figure 3.6) you can set the output options.

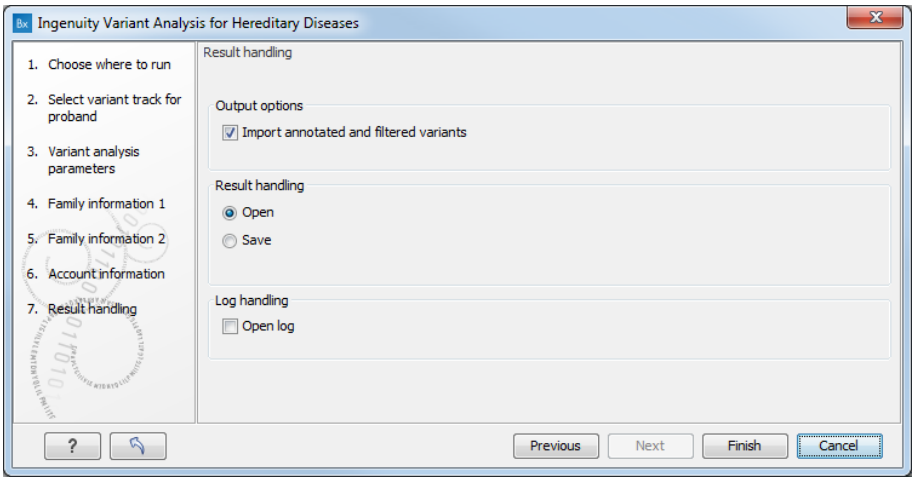

Figure 3.6: *The result handling step.*

If the **Import annotated and filtered variants** option is checked, the tool will produce a variant track as output. If it is unchecked, the analysis will be created, and can be accessed inside the Ingenuity Variant Analysis web interface, but the results will not be imported into the workbench. Note: it is not possible to import results if you have selected the **Upload only** pipeline earlier in the wizard.

If you choose to open the results the two generated outputs will be opened in the View Area without being saved. In this case you will have to manually save the outputs if you would like to keep them. If you choose to save the outputs, click on the button labeled **Next** to specify where to save the results and click on the button labeled **Finish** to start the Ingenuity Variant Analysis. Your results will not be opened automatically but will be saved at the destination you have specified.

The outputs and how to manually adjust filter settings are described in section 5.

3.1 Genetic disease pipelines

The following pipelines can be used to identify variants associated with genetic disease. Each pipeline will run an analysis and specifically return variants with the indicated mode of inheritance.

- Dominant inheritance pattern
- This disease is caused by a de novo mutation.
- This disease is caused by a recessive compound heterozygous variant.
- This disease is caused by a recessive homozygous variant.

Note that the genetic disease pipelines listed above correspond to the analyses that can be run from the Ingenuity Variant Analysis web interface, with the exception that the constituent filters are static when running them from the workbench (running them from the IVA web interface means that filters that do not let any variants pass through are automatically removed).

The user is encouraged to review and customize the best practice filter cascade settings as they see fit through the IVA web interface after the workbench analysis has completed. The filter cascade that corresponds to an analysis can be easily found by right-clicking the open analysis result variant track, and selecting "Launch Ingenuity Variant Analysis (in browser)". When filter settings have been changed, the altered results can be imported in the workbench by right-clicking the original analysis result variant track in the workbench, and selecting "Variant Analysis Update". This which will generate a new variant track marked with (IVA update) (see section 5).

3.2 Pipeline description and supported input

For each genetic disease pipeline, only certain sample input combinations are supported, and running an analysis with an unsupported sample input combination may output wrong results. Make sure to review filter cascade settings in Ingenuity Variant Analysis if you are running an unsupported analysis. For example, if all four analyses are run on a sample trio with one affected parent, then only the results from the dominant and recessive homozygous analyses are supported.

Dominant inheritance pattern Best practice filter cascade to identify variants present in affected parents and transmitted to the proband and affected siblings in a dominant inheritance pattern. When one parent is set to affected, variants present in the unaffected parent will be excluded. Setting both parents to unaffected will be interpreted as if the affected status of the parents is unknown, and transmitted heterozygous variants from both parents will then be reported. Note that de novo variants are only reported by the dedicated analysis described in the paragraph below.

Supported input:

- Proband affected + Mother unaffected + Father unaffected
- Proband affected + Mother unaffected + Father affected
- Proband affected + Mother affected + Father unaffected

This disease is caused by a de novo mutation Best practice filter cascade for identification of disease associated variants in the proband that are not present in any parent.

Supported input:

- Proband affected + Mother unaffected + Father unaffected

This disease is caused by a recessive compound heterozygous variant Best practice filter cascade to identify recessive heterozygous variants present in the unaffected parents and transmitted to the proband in a compound heterozygous pattern, thus impacting both alleles of a gene associated with genetic disease.

Supported input:

- Proband affected + Mother unaffected + Father unaffected

This disease is caused by a recessive homozygous variant Best practice filter cascade for identification of recessive variants present in both parents and transmitted to the proband, resulting in a homozygous variant that impacts both alleles of a gene associated with genetic disease.

Supported input:

- Proband affected + Mother unaffected + Father unaffected
- Proband affected + Mother unaffected + Father affected
- Proband affected + Mother affected + Father unaffected

Chapter 4

Workflows

Installing the Ingenuity Variant Analysis plugin will also add three ready-to-use workflows in the Toolbox, under the Whole Genome Sequencing and Whole Exome Sequencing folders (figure 4.1). If you are working with Targeted Amplicon Sequencing data, use the workflows from the Whole Exome Sequencing folder.

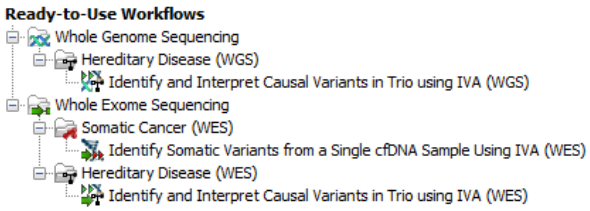

Figure 4.1: *Accessing the user settings inside the Ingenuity Variant Analysis web interface*

The concept of the pre-installed ready-to-use workflows is that read data are used as input in one end of the workflow and in the other end of the workflow you get - among other output files - a track based genome browser view and a table with all the identified variants subjected to the Ingenuity Variant Analysis.

Once you have selected the workflow in the folder relevant to your input data, you can read the steps you need to take to start the workflow. For more information on the specific tools used in this workflow, see the CLC Genomics Workbench manual chapter on Workflows.

4.1 Identify and Interpret Causal Variants in a Trio using IVA (WGS)

To run this workflow, go to:

Toolbox | **Ready-to-Use Workflows** | **Whole Genome Sequencing** ( ) | **Hereditary Disease** ( ) | **Identify and Annotate Variants in a Trio using IVA (WGS)** ( )

1. Double-click on the **Identify and Annotate Variants in a Trio using IVA (WGS)** tool to start the analysis. If you are connected to a server, you will first be asked where you would like to run the analysis.
2. Select the **sequencing reads** for the proband, father and mother respectively (figure 4.2). You can do that by double-clicking on the reads file name or clicking once on the file and

then clicking on the arrow pointing to the right side in the middle of the wizard. Click **Next** between each family member.

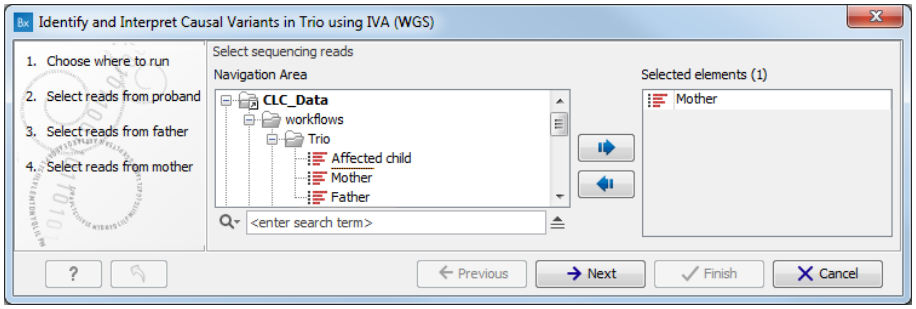

Figure 4.2: *Specify the sequencing reads for each family member successively.*

3. Specify the parameters for the **Fixed Ploidy Variant Detection** tool for the proband, father and mother successively (figure 4.3).

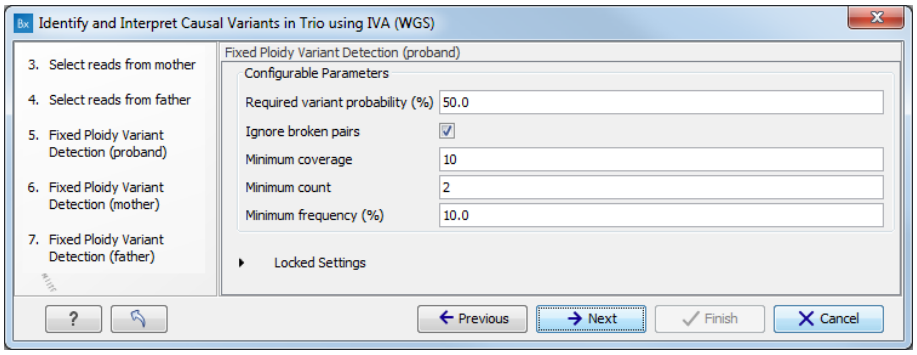

Figure 4.3: *Specifying the parameters for the Fixed Ploidy Variant Detection tool.*

The parameters that can be set are:

- **Required variant probability** is the minimum probability value of the 'variant site' required for the variant to be called. Note that it is not the minimum value of the probability of the individual variant. For the Fixed Ploidy Variant detector, if a variant site - and not the variant itself - passes the variant probability threshold, then the variant with the highest probability at that site will be reported even if the probability of that particular variant might be less than the threshold. For example if the required variant probability is set to 0.9 then the individual probability of the variant called might be less than 0.9 as long as the probability of the entire variant site is greater than 0.9.
- **Ignore broken pairs:** When ticked, reads from broken pairs are ignored. Broken pairs may arise for a number of reasons, one being erroneous mapping of the reads. In general, variants based on broken pair reads are likely to be less reliable, so ignoring them may reduce the number of spurious variants called. However, broken pairs may also arise for biological reasons (e.g. due to structural variants) and if they are ignored some true variants may go undetected. Please note that ignored broken pair reads will not be considered for any non-specific match filters.
- **Minimum coverage:** Only variants in regions covered by at least this many reads are called.

- **Minimum count:** Only variants that are present in at least this many reads are called.
- **Minimum frequency:** Only variants that are present at least at the specified frequency (calculated as 'count'/'coverage') are called.

4. In the Ingenuity Variant Analysis for Hereditary Diseases window, you need to specify your login information to Ingenuity Variant Analysis (figure 4.4).

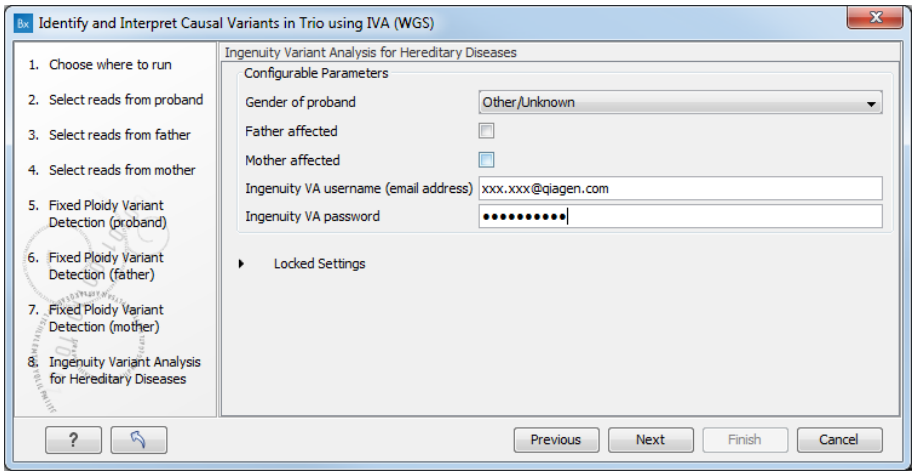

Figure 4.4: *Specify login information to Ingenuity Variant Analysis*

It is also possible to configure the following parameters:

- Gender of proband: you can choose between male, female, ambiguous (for babies born with sex chromosomes anomalies or sexual organs that are not yet fully developed) and unknown.
- Check if another family member is affected: the mother or the father.
- Ingenuity VA username: usually the email address you used to sign in the Ingenuity Variant Analysis.
- Ingenuity VA password: the password you chose when you signed in on the Ingenuity website.

5. On the last wizard window, pressing the button **Preview All Parameters** allows you to preview all parameters. At this step you can only view the parameters, it is not possible to make any changes. Choose to save the results and click on the button labeled **Finish**.

The following outputs are generated:

- **3 Reads Track**, one for each family member
- **3 Filtered Variant Track**, one for each family member
- A **Genome Browser View**
- A **URL file**
- A **Recessive-Compound Heterozygous** track
- A **Recessive-Homozygous** track

- A **De novo** track
- A **Dominant** track

4.2 Identify and Interpret Causal Variants in a Trio using IVA (WES)

To run this workflow, go to:

Toolbox | **Ready-to-Use Workflows** | **Whole Exome Sequencing** () | **Hereditary Disease** () | **Identify and Annotate Variants in a Trio using IVA (WES)** ()

1. Double-click on the **Identify and Annotate Variants in a Trio using IVA (WES)** tool to start the analysis. If you are connected to a server, you will first be asked where you would like to run the analysis.
2. Select the **sequencing reads** for the proband (figure 4.5). You can do that by double-clicking on the reads file name or clicking once on the file and then clicking on the arrow pointing to the right side in the middle of the wizard. Click **Next**.

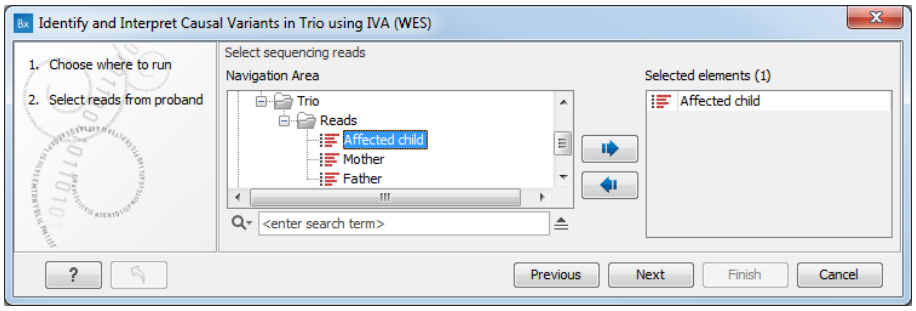

Figure 4.5: *Specify the sequencing reads for each family member successively.*

3. Specify a **target region** file (figure 4.6). This is a file that depends on the technology you used for sequencing.

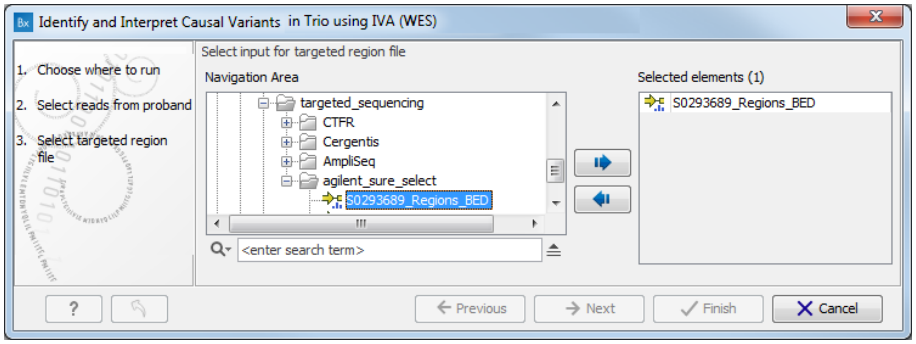

Figure 4.6: *Specify a target region file*

4. Select now the reads for the mother and father respectively. Click Next between each family member.
5. Specify the parameters for the **Fixed Ploidy Variant Detection** tool for the proband, mother and father successively (figure 4.7).

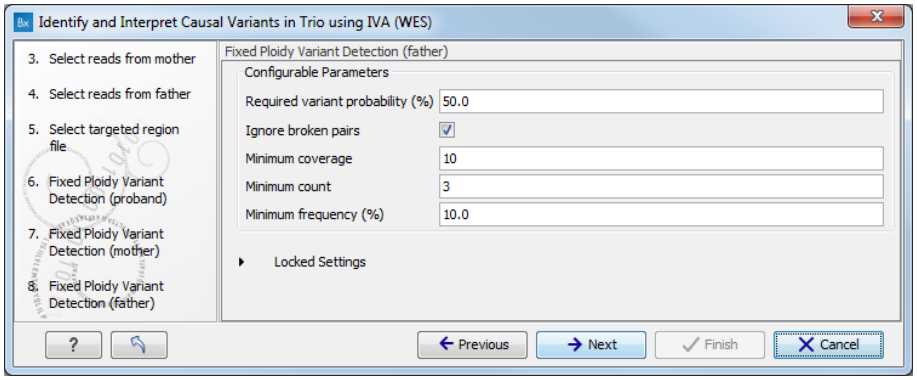

Figure 4.7: *Specifying the parameters for the Fixed Ploidy Variant Detection tool.*

The parameters that can be set are:

- **Required variant probability** is the minimum probability value of the 'variant site' required for the variant to be called. Note that it is not the minimum value of the probability of the individual variant. For the Fixed Ploidy Variant detector, if a variant site - and not the variant itself - passes the variant probability threshold, then the variant with the highest probability at that site will be reported even if the probability of that particular variant might be less than the threshold. For example if the required variant probability is set to 0.9 then the individual probability of the variant called might be less than 0.9 as long as the probability of the entire variant site is greater than 0.9.
- **Ignore broken pairs:** When ticked, reads from broken pairs are ignored. Broken pairs may arise for a number of reasons, one being erroneous mapping of the reads. In general, variants based on broken pair reads are likely to be less reliable, so ignoring them may reduce the number of spurious variants called. However, broken pairs may also arise for biological reasons (e.g. due to structural variants) and if they are ignored some true variants may go undetected. Please note that ignored broken pair reads will not be considered for any non-specific match filters.
- **Minimum coverage:** Only variants in regions covered by at least this many reads are called.
- **Minimum count:** Only variants that are present in at least this many reads are called.
- **Minimum frequency:** Only variants that are present at least at the specified frequency (calculated as 'count'/'coverage') are called.

6. In the Ingenuity Variant Analysis for Hereditary Diseases window, you need to specify your login information to Ingenuity Variant Analysis (figure 4.8).

   It is also possible to configure the following parameters:

   - Gender of proband: you can choose between male, female, ambiguous (for babies born with sex chromosomes anomalies or sexual organs that are not yet fully developed) and unknown.
   - Check if another family member is affected: the mother or the father.
   - Ingenuity VA username: usually the email address you used to sign in the Ingenuity Variant Analysis.

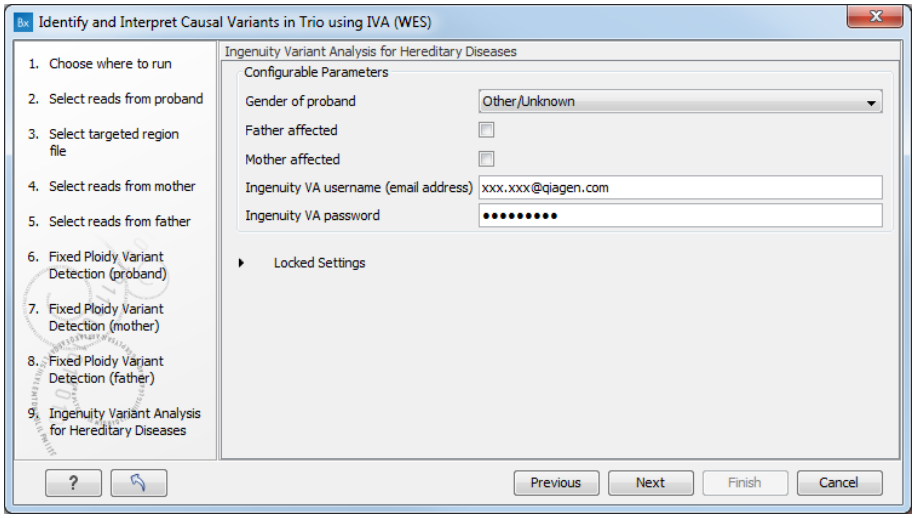

Figure 4.8: *Specify login information to Ingenuity Variant Analysis*

- Ingenuity VA password: the password you chose when you signed in on the Ingenuity website.

7. On the last wizard window, pressing the button **Preview All Parameters** allows you to preview all parameters. At this step you can only view the parameters, it is not possible to make any changes. Choose to save the results and click on the button labeled **Finish**.

The following outputs are generated:

- **3 Reads Track**, one for each family member
- **3 Filtered Variant Track**, one for each family member
- **3 Coverage Report (Target Region Coverage Report)**, one for each family member
- **3 Per-region Statistics Track (Target Region Coverage)**, one for each family member
- A **Genome Browser View**
- A **URL file**
- A **Recessive-Compound Heterozygous** track
- A **Recessive-Homozygous** track
- A **De novo** track
- A **Dominant** track

4.3 Identify Somatic Variants from a Single cfDNA Sample using IVA (WES)

To run this workflow, go to:

Toolbox | **Ready-to-Use Workflows** | **Whole Exome Sequencing** ( ) | **Somatic Cancer** ( ) | **Identify Somatic Variants from a Single cfDNA Sample using IVA (WES)**

1. Double-click on the **Identify Somatic Variants from a Single cfDNA Sample using IVA (WES)** tool to start the analysis. If you are connected to a server, you will first be asked where you would like to run the analysis.

2. Select the **sequencing reads** (figure 4.9). You can do that by double-clicking on the reads file name or clicking once on the file and then clicking on the arrow pointing to the right side in the middle of the wizard. Click **Next**.

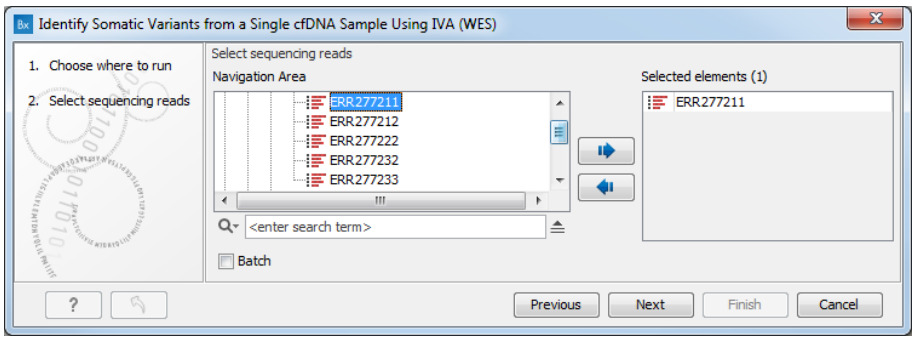

Figure 4.9: *Specify the sequencing reads.*

3. Specify the parameters for the **Low Frequency Variant Detection** tool (figure 4.10).

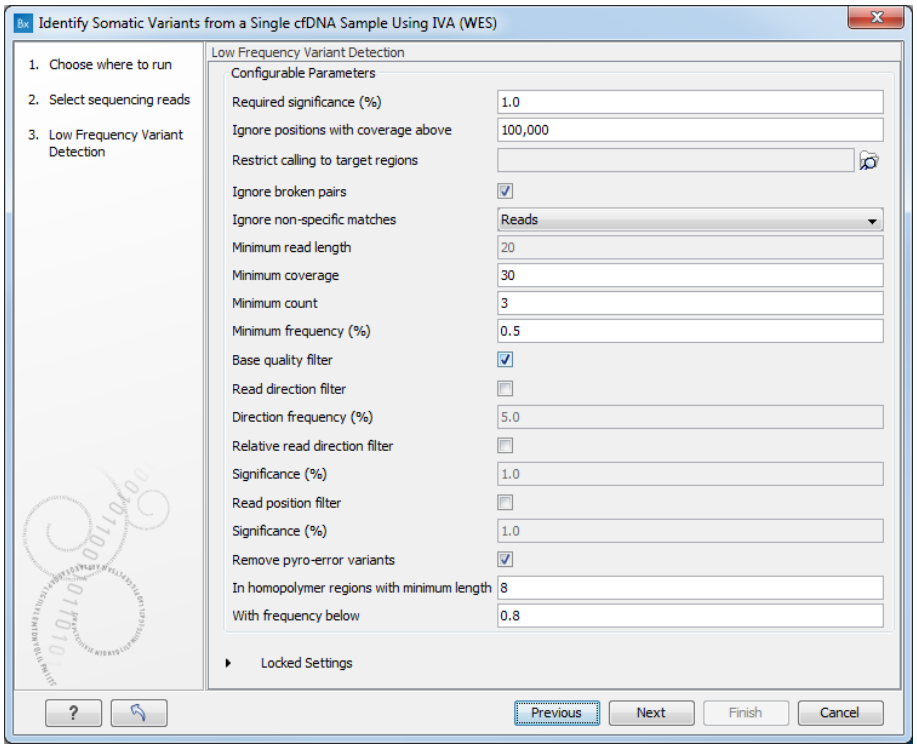

Figure 4.10: *Specifying the parameters for the Low Frequency Variant Detection tool.*

The parameters that can be set are:

- **Required significance (%)**: this parameter determines the cut-off value for the statistical test for the variant not being due to sequencing errors. Only variants that are at

least this significant will be called. The lower you set this cut-off, the fewer variants will be called.

- **Ignore positions with coverage above**: All positions with coverage above this value will be ignored when inspecting the read mapping for variants. The option is highly useful in cases where you have a read mapping which has areas of extremely high coverage as are areas around centromeres in whole genome sequencing applications for example.
- **Restrict calling to target regions**: Only positions in the regions specified by the target region file will be inspected for variants.
- **Ignore broken pairs** When ticked, reads from broken pairs are ignored. Broken pairs may arise for a number of reasons, one being erroneous mapping of the reads. In general, variants based on broken pair reads are likely to be less reliable, so ignoring them may reduce the number of spurious variants called. However, broken pairs may also arise for biological reasons (e.g. due to structural variants) and if they are ignored some true variants may go undetected. Please note that ignored broken pair reads will not be considered for any non-specific match filters.
- **Ignore non-specific matches** You can choose to ignore non-specific matches between reads, regions or to not ignore them at all.
- **Minimum read length** Only variants in reads longer than this size are called.
- **Minimum coverage** Only variants in regions covered by at least this many reads are called.
- **Minimum count** Only variants that are present in at least this many reads are called.
- **Minimum frequency** Only variants that are present at least at the specified frequency (calculated as 'count'/'coverage') are called.
- **Base quality filter:** The base quality filter can be used to ignore the reads whose nucleotide at the potential variant position is of dubious quality. This is assessed by considering the quality of the nucleotides in the read in the region around the nucleotide position.
- **Read direction filter:** The read direction filter removes variants that are almost exclusively present in either forward or reverse reads. For many sequencing protocols such variants are most likely to be the result of amplification induced errors. Note, however, that the filter is **NOT suitable for amplicon data**, as for this you will not expect coverage of both forward and reverse reads. The filter has a single parameter:
  - **Direction frequency:** Variants that are not supported by at least this frequency of reads from each direction are removed.
- **Relative read direction filter:** The relative read direction filter attempts to do the same thing as the 'Read direction filter', but does this in a statistical, rather than absolute, sense: it tests whether the distribution among forward and reverse reads of the variant carrying reads is different from that of the total set of reads covering the site. The statistical, rather than absolute, approach makes the filter less stringent. The filter has one parameter:
  - **Significance:** Variants whose read direction distribution is significantly different from the expected with a test at this level, are removed. The lower you set the significance cut-off, the fewer variants will be filtered out.

- **Read position filter:** The read position filter is a filter that attempts to remove systematic errors in a similar fashion as the 'Read direction filter', *but* that is also **suitable for hybridization-based data**. It removes variants that are located differently in the reads carrying it than would be expected given the general location of the reads covering the variant site. This is done by categorizing each sequenced nucleotide (or gap) according to the mapping direction of the read and also where in the read the nucleotide is found; each read is divided in five parts along its length and the part number of the nucleotide is recorded. This gives a total of ten categories for each sequenced nucleotide and a given site will have a distribution between these ten categories for the reads covering the site. If a variant is present in the site, you would expect the variant nucleotides to follow the same distribution. The read position filter carries out a test for whether the read position distribution of the variant carrying reads is different from that of the total set of reads covering the site. The filter has one parameter:
  - **Significance:** Variants whose read position distribution is significantly different from the expected with a test at this level, are removed. The lower you set the significance cut-off, the fewer variants will be filtered out.
- **Remove pyro-error variants:** This filter can be used to remove insertions and deletions in the reads that are likely to be due to pyro-like errors in homopolymer regions. There are two types of such errors: They may occur either at (1) the immediate ends of homopolymer regions or (2) as an 'overspill' a few nucleotides downstream of a homopolymer region. In case (1) the exact numbers of the same number of nucleotide is uncertain and a sequence like "AAAAAAAA" is sometimes reported as "AAAAAAAA**A**". In case (2) a sequence like "CGAAAAAGTCG" may sometimes get an 'overspill' insertion of an A between the T and C so that the reported sequence is C "CGAAAAAGT**A**CG". Note that the removal is done in the reads as a very first step, before calling the initial 1 bp variants.

  There are two parameters that must be specified for this filter:
  - **In homopolymer regions with minimum length**: Only insertion or deletion variants in homopolymer regions of at least this length will be removed.
  - **With frequency below:** Only insertion or deletion variants whose frequency (ignoring all non-reference and non-homopolymer variant reads) is lower than this threshold will be removed.

4. Specify a **target region** file (figure 4.11). This is a file that depends on the technology you used for sequencing.

5. Specify a variant track for the **Identify Known Mutations from Sample Mappings** tool (figure 4.12).

   A Clinvar variant track is selected by default, but can be changed to the variant track of your choice if needed.

6. In the Ingenuity Variant Analysis dialog, you can specify the **Disease name**, or leave it to "Any cancer". Enter your login information to Ingenuity Variant Analysis (figure 4.13).

7. On the last wizard window, pressing the button **Preview All Parameters** allows you to preview all parameters. At this step you can only view the parameters, it is not possible to make any changes. Choose to save the results and click on the button labeled **Finish**.

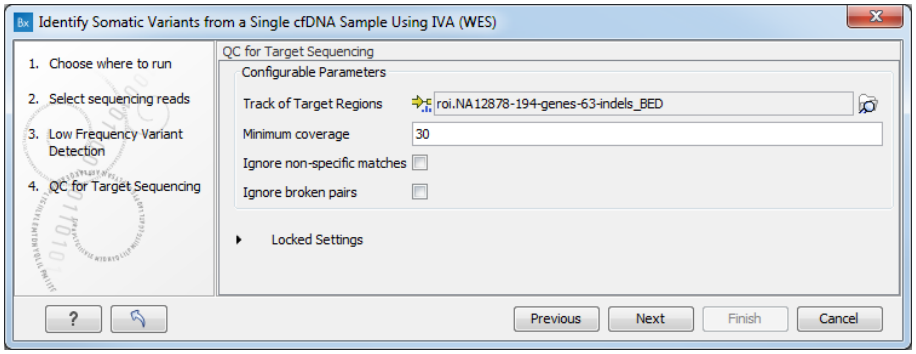

Figure 4.11: *Specify a target region file*

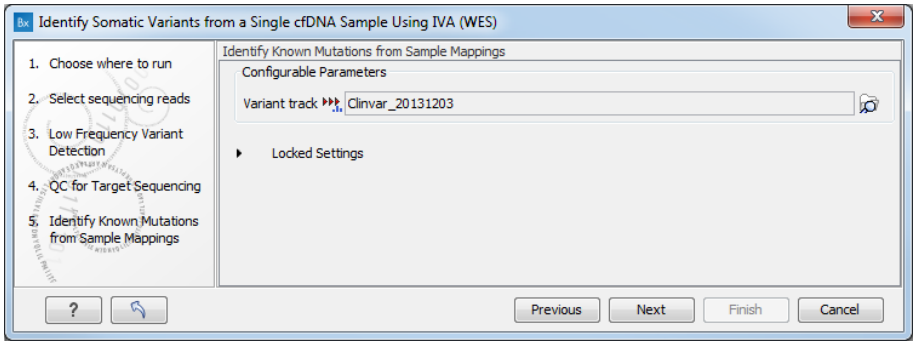

Figure 4.12: *Specifying the parameters for the Low Frequency Variant Detection tool.*

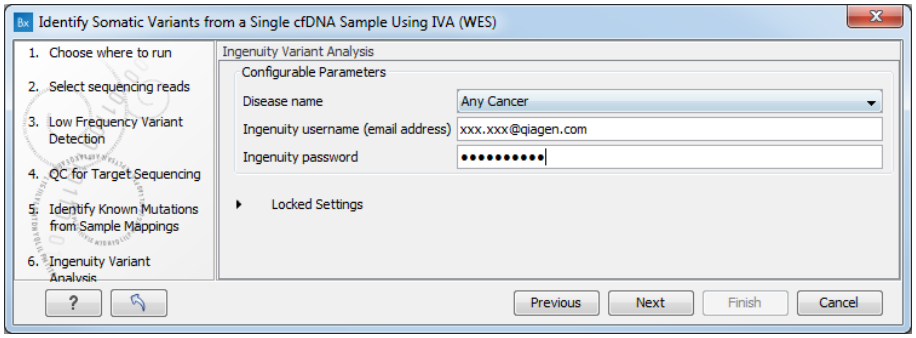

Figure 4.13: *Specify login information to Ingenuity Variant Analysis*

The following outputs are generated:

- A **Reads Track**, the read mapping
- A **Coverage Report (Target Region Coverage Report)**
- A **Per-region Statistics Track (Target Region Coverage)**
- An **Annotated Variant Track**, the Identified Known Variants track which contains the known variants provided by the user.
- An **Filtered Variant Track** for the Somatic Driver Variant, called the Identified Variants
- An **Amino Acid Track**

- An **Imported Variant Track** for the Somatic Driver Variant
- A **URL file** for the Somatic Driver Variant
- A **Genome Browser View**

Chapter 5

Analysis using the plugin and the IVA web interface

When the analysis is complete, you will get different kinds of output:

- A variant track with the annotated and filtered variants (figure 5.1). This track can be opened in a Genome Browser View by double-clicking on the name of the variant track in the **Navigation Area**.

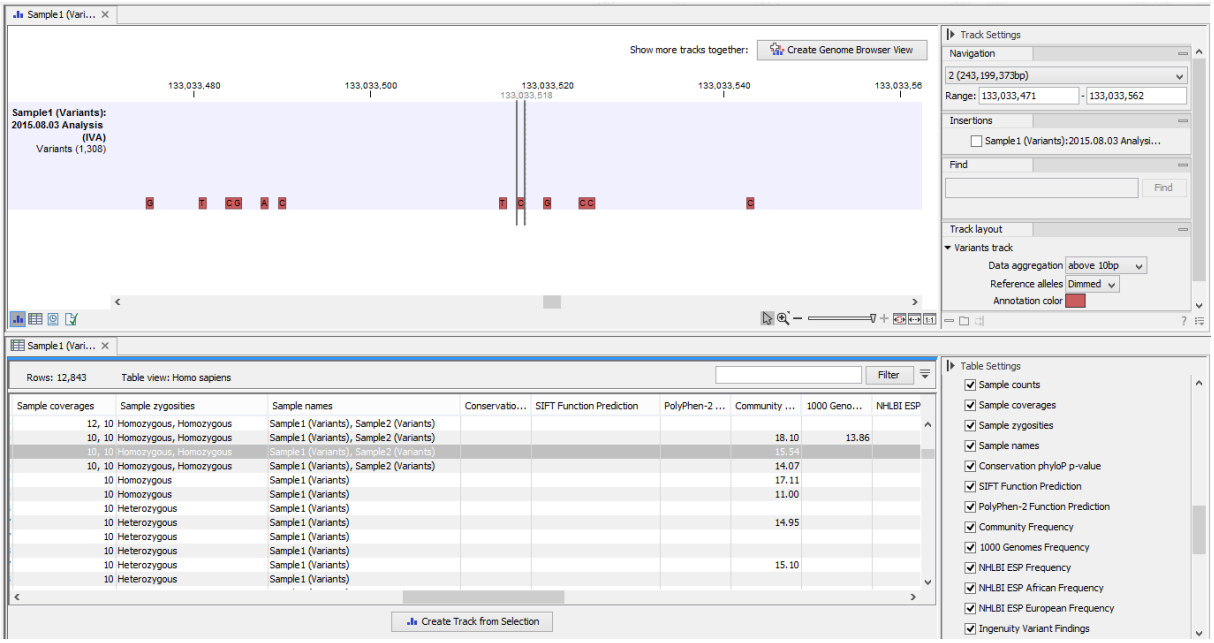

Figure 5.1: *The result of the Ingenuity Variant Analysis opened in the Genome Browser View in CLC Genomics Workbench. The variant track is shown in split view with the variant table.*

- A document providing a link to the Ingenuity Variant Analysis page (see figure 5.2).

  Copying this link and pasting it into an internet browser will take you to the Ingenuity Variant Analysis page, where you can narrow down your analysis further by applying different filters or by adjusting the predefined filter settings.

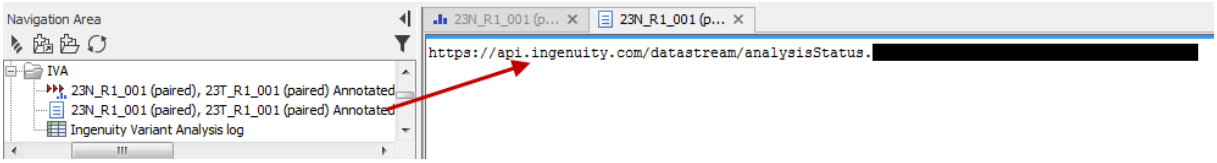

Figure 5.2: *Copy this link into an internet browser to see the result of the Ingenuity Variant Analysis.*

- A log file, if you ticked the **Open log** box.

There are two different approaches to how you can handle the identified variants:

- Open the variants in the Genome Browser View in the Workbench. The identified variants can be viewed in track format in the Genome Browser View by double-clicking on the name of the variant track in the **Navigation Area**. The button labeled **Create Genome Browser View** in the upper right corner of the **View Area** can be used to create a list of tracks in the same view, which allows comparison of the identified variants with other tracks, such as the reference sequence, the CDS, read mappings, or other variant tracks.

- View the variants on the Ingenuity Variant Analysis web page. This option allows adjustment of the predefined filter settings. The variants in Ingenuity Variant Analysis can be accessed in two different ways:

  – Use the link provided in one of the output files in the Workbench. Copy the link and paste it into an internet browser. This will send you directly to the variant analysis on the Ingenuity Variant Analysis web page. An example is shown in figure 5.3.

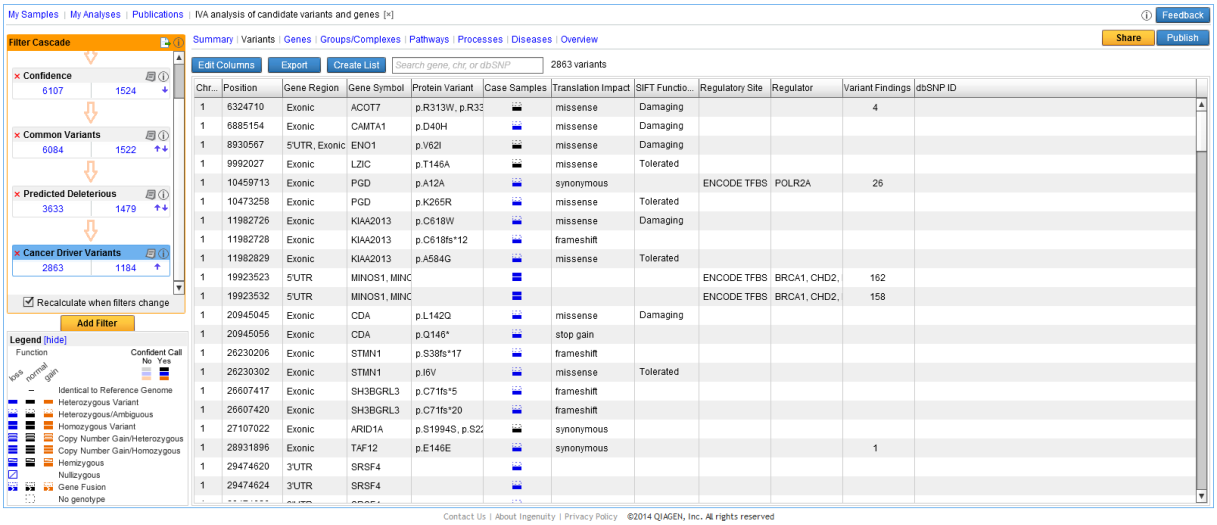

Figure 5.3: *Copy the link provided in one of the output files and paste it into an internet browser to go directly to the specific variant analysis on the Ingenuity Variant Analysis page.*

  – Open the variant track that was produced as one of the outputs. Right-click on the variant track in the Genome Browser View and select **Launch Ingenuity Variant Analysis** (figure 5.4). This will also send you directly to the Ingenuity Variant Analysis web page.

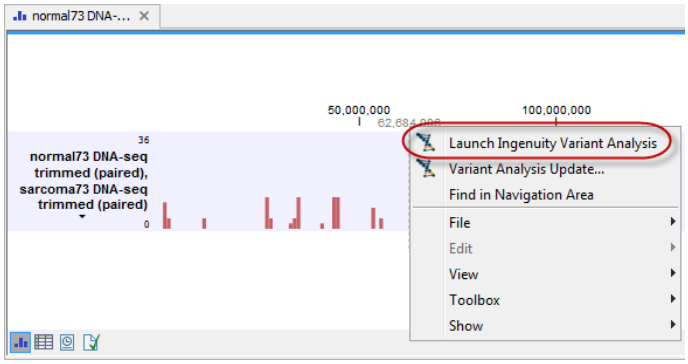

Figure 5.4: *Right-clicking on the variant track in the Genome Browser View and selecting Launch Ingenuity Variant Analysis will send you directly to the specific variant analysis on the Ingenuity Variant Analysis page.*

Ingenuity Variant Analysis enables you to apply a number of different filters. The example in figure 5.5 shows a filter cascade with the default filters. Note that in the case of the Hereditary Disease pipeline, four custom pipelines (one for each inheritance mode) with a fixed set of filters are available.

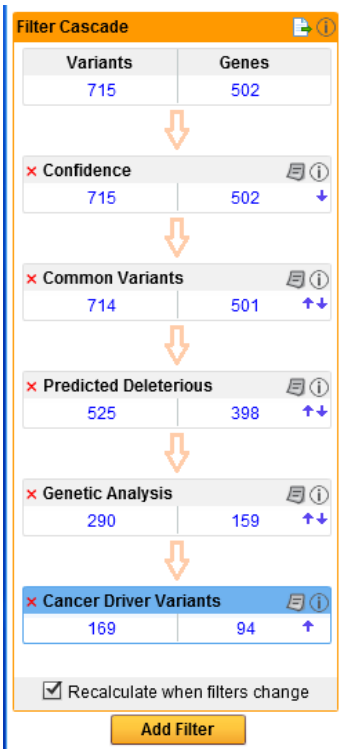

Figure 5.5: *An example of an Ingenuity Variant Analysis filter cascade that narrows down the initial number of variants to focus on a limited number of specific variants that are left after applying a number of different filters.*

After running the initial analysis from the Workbench, it is possible to customize the filter cascade in the Ingenuity Variant Analysis web interface: add more filters with the button labeled **Add Filter** found at the bottom of the filter cascade, and modify or delete existing filters with the paper icon found in the right hand side of the individual filters in the filter cascade (figure 5.6). Click on the

information icon next to the paper icon to get more information about Ingenuity Variant Analysis.

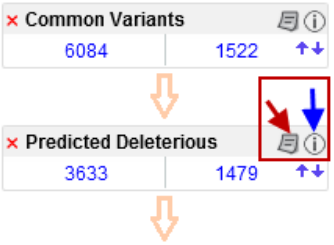

Figure 5.6: *Click on the red cross if you would like to delete the filter, click on the paper icon (red arrow) to see the filter details, or if you would like to adjust them. Click on the information icon (blue arrow) if you would like to learn more about Ingenuity Variant Analysis.*

When you have modified the filters on the Ingenuity Variant Analysis web page, you can either choose to use the options provided on the Ingenuity Variant Analysis web page to go into detail with the individual variants, or you can go back to the workbench and visualize the variants in the Genome Browser View.

The modified variant track can be imported into the Workbench by right-clicking on the original Ingenuity Variant Analysis variant track output that was generated with the default filter settings (see figure 5.7). Choose **Variant Analysis Update** and save the updated variant track in the **Navigation Area**. The updated variant track will be saved with the name extension "(IVA update)", which means that the original variant track will not be overwritten by the updated variant track.

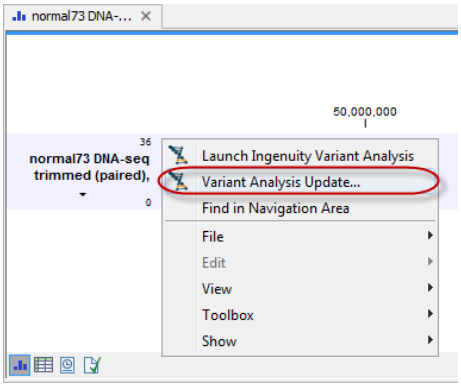

Figure 5.7: *If you have made changes to the used filters, you can import the updated variant track into the Workbench by right-clicking on the original variant track and choosing "Variant Analysis Update".*

When you have imported the updated variant track, we recommend that you open the updated variant track in split view with the table view. After running the variant analysis, the variant table will contain additional columns holding Ingenuity Variant Analysis-specific information. The type of analysis performed and which filters were used will determine which of these columns (see figure 5.8) will be added to your results.

Please visit our website https://www.qiagenbioinformatics.com/products/ingenuity-variant-analysis/ for more information about the wide range of options available on the Ingenuity Variant Analysis web page. If you would like to learn more about Ingenuity Variant Analysis annotations, please see http://ingenuity.force.com/variants/VariantTutorials.

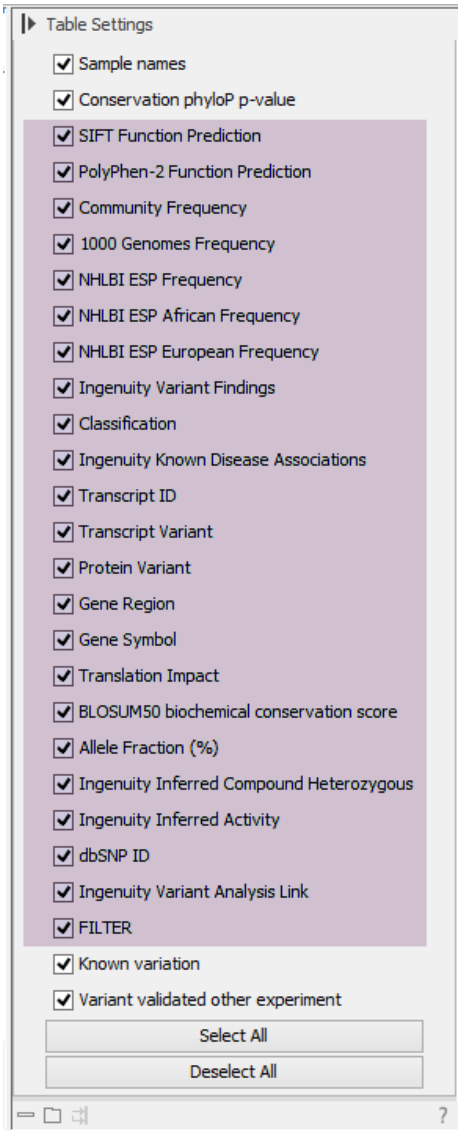

Figure 5.8: *You can see which columns have been added in the Table view of the variants.*

Chapter 6

Add Information from Allele Frequency Community

The Add Information from Allele Frequency Community tool allows you to add AFC Frequency annotations from the Allele Frequency Community to variant tracks. To be able to obtain AFC Frequency annotations from the Allele Frequency Community, your Ingenuity user account must be opted in to the Allele Frequency Community. Section 8 describes how to change your Allele Frequency Community opt-in status.

To start the tool, go to

Ingenuity Variant Analysis | Add Information from Allele Frequency Community

If you are connected to a server, you will first be asked where you would like to run the analysis. If you are not connected to a server, the first step is to specify the input for the analysis. The Add Information from Allele Frequency Community tool accepts a single variant track ( ) as input (figure 6.1).

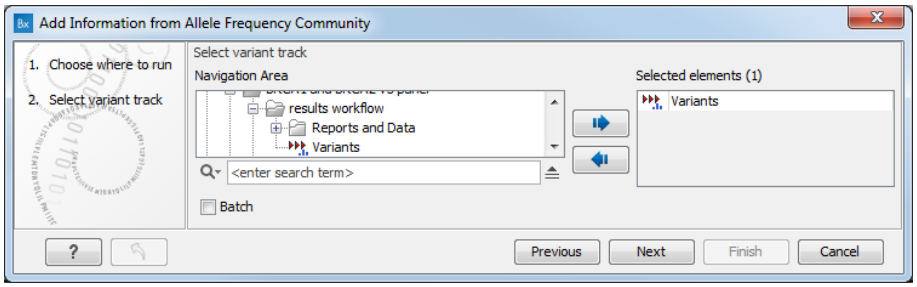

Figure 6.1: *Select the variant track that you would like to analyze.*

In the next dialog, you can specify your Ingenuity username, password and the reference sequence (figure 6.2).

- **Ingenuity username**: email address used to log in to Ingenuity Variant Analysis
- **Ingenuity password**: password corresponding to your Ingenuity username
- **Reference**: Select the human reference sequence that is found under CLC_References in the Navigation Area. Complete human genomes (e.g., hg19 (GRCh37) and hg38 (GRCh38))

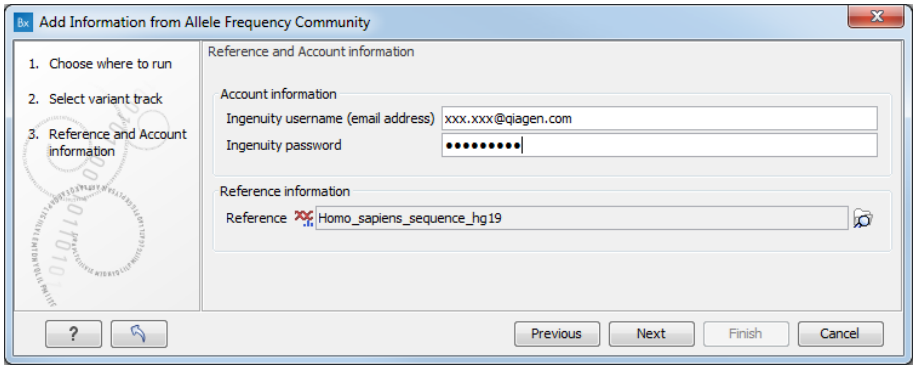

Figure 6.2: *Specify analysis parameters.*

and subsets of these (e.g. individual chromosomes) can be used as references.

Finally, you can choose your output options. If you choose to open the results the two generated outputs will be opened in the View Area without being saved. In this case you will have to manually save the outputs if you would like to keep them. If you choose to save the outputs, they will not be opened automatically but will be saved at the destination you have specified.

When the analysis is finished, the resulting track will contain an additional column named "AFC Frequency", containing the observed ethnic group specific frequencies (in percent) of the variants in the Allele Frequency Community. If the "AFC Frequency" column is empty for a variant, it indicates that the variant was not found in the Allele Frequency Community.

Chapter 7

Remove Variants Found in Allele Frequency Community

The Remove Variants found in Allele Frequency Community tool allows you to add AFC Frequency annotations from the Allele Frequency Community to variant tracks, and to filter the variants based on those annotations. To be able to obtain AFC Frequency annotations from the Allele Frequency Community, your Ingenuity user account must be opted in to the Allele Frequency Community. Section 8 describes how to change your Allele Frequency Community opt-in status.

To start the tool, go to:

Ingenuity Variant Analysis | Remove Variants Found in Allele Frequency Community

If you are connected to a server, you will first be asked about where you would like to run the analysis. If you are not connected to a server, the first step is to specify the input for the analysis. The Remove Variants found in Allele Frequency Community tool accepts a single variant track ( ) as input as shown in figure 7.1.

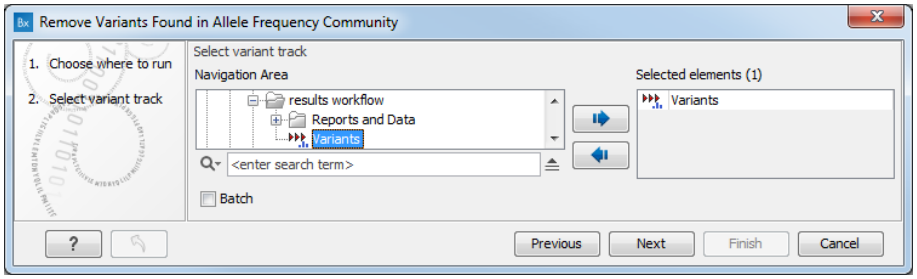

Figure 7.1: *Select a variant track.*

In the following wizard, specify your Ingenuity username, password and the reference sequence as described below (figure 7.2).

- **Ingenuity username**: email address used to log in to Ingenuity Variant Analysis.
- **Ingenuity password**: password corresponding to your Ingenuity username
- **Reference**: Select the human reference sequence that is found under CLC_References in the Navigation Area. Complete human genomes (e.g., hg19 (GRCh37) and hg38 (GRCh38))

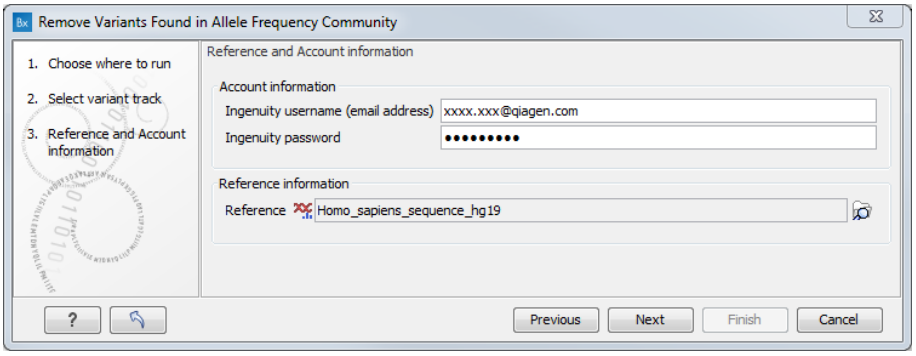

Figure 7.2: *Specify analysis parameters.*

and subsets of these (e.g. individual chromosomes) can be used as references.

Click **Next** to go to the next wizard step. Here you can specify the cutoff for filtering by entering the desired value in the **Maximum frequency** field (figure 7.3). Only variants whose Allele Frequency Community frequency is equal to or lower than the specified value will be considered.

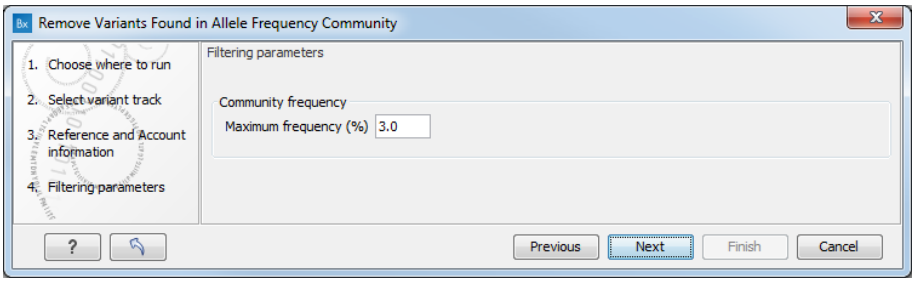

Figure 7.3: *Specify the filter cutofff.*

Click **Next** to go to the final wizard step where you can set the output options.

If you choose to open the results the two generated outputs will be opened in the View Area without being saved. In this case you will have to manually save the outputs if you would like to keep them. If you choose to save the outputs, they will not be opened automatically but will be saved at the destination you have specified.

When the analysis is finished, the resulting track will contain an additional column named "AFC Frequency", containing the observed frequencies (in percent) of the variants in the Allele Frequency Community. Furthermore, the number of variants will have been reduced according to the cutoff parameter that you have specified.

Note that if the "AFC Frequency" column is empty for a variant, it indicates that the variant was not found in the Allele Frequency Community. This can happen when the variant is a "new" one in the Allele Frequency Community database, but also for reference allele from an heterozygous pair whose non-reference allele was kept by the Remove Variants found in Allele Frequency Community tool as no reference variants are ever found in the Allele Frequency Community database. To filter out reference alleles, click on the "Filter" button in the variant table, select "Reference allele" in the drop-down menu, and keep only the alleles that contain "No". After this step, the only variants without an annotation in the "Community frequency" column will be the ones considered as "new" variants, i.e., not previously found in the Allele Frequency Community database.

Chapter 8

Changing Allele Frequency Community opt-in settings

In order to gain access to AFC Frequency annotations from the Allele Frequency Community, your Ingenuity user account must be opted in to the Allele Frequency Community. To change your Allele Frequency Community opt-in settings, carry out the following steps:

1. Log in to the Ingenuity Variant Analysis web interface: go to http://www.qiagenbioinformatics.com/products/ingenuity-variant-analysis/.
2. After logging in, go to Settings (figure 8.1)

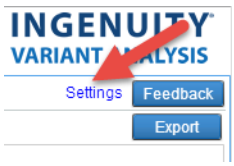

Figure 8.1: *Accessing the user settings inside the Ingenuity Variant Analysis web interface*

3. Change your Allele Frequency Community opt-in status using the checkbox (figure 8.2)

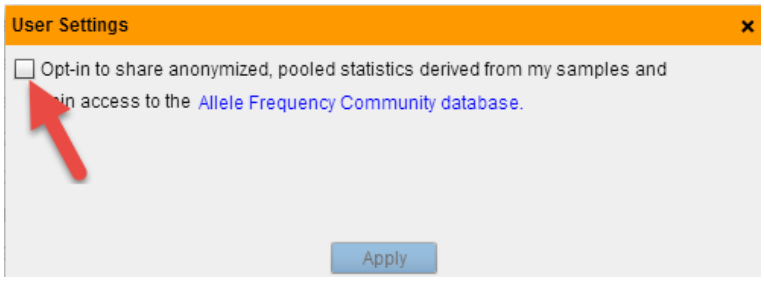

Figure 8.2: *Changing the Allele Frequency Community opt-in status using the checkbox inside user settings*

Chapter 9

Install and uninstall plugins

Ingenuity Variant Analysis is installed as a plugin.

Note: In order to install plugins and modules, the Workbench must be run in administrator mode. On Linux and Mac, it means you must be logged in as an administrator. On Windows, you can do this by right-clicking the program shortcut and choosing "Run as Administrator".

Plugins are installed and uninstalled using the plugin manager.

Help in the Menu Bar | Plugins... ( ) or Plugins ( ) in the Toolbar

The plugin manager has two tabs at the top:

- **Manage Plugins.** This is an overview of plugins that are installed.
- **Download Plugins.** This is an overview of available plugins on QIAGEN Aarhus server.

9.1 Install

To install a plugin, click the **Download Plugins** tab. This will display an overview of the plugins that are available for download and installation (see figure 9.1).

Select Ingenuity Variant Analysis to display additional information about the plugin on the right side of the dialog. Click **Download and Install** to add the plugin functionalities to your workbench.

Accepting the license agreement

The End User License Agreement (EULA) must be read and accepted as part of the installation process. Please read the EULA text carefully, and if you agree to it, check the box next to the text **I accept these terms**. If further information is requested from you, please fill this in before clicking on the **Finish** button.

If Ingenuity Variant Analysis is not shown on the server but you have the installer file on your computer (for example if you have downloaded it from our website), you can install the plugin by clicking the **Install from File** button at the bottom of the dialog and specifying the plugin *.cpa file saved on your computer.

When you close the dialog, you will be asked whether you wish to restart the workbench. The plugin will not be ready for use until you have restarted.

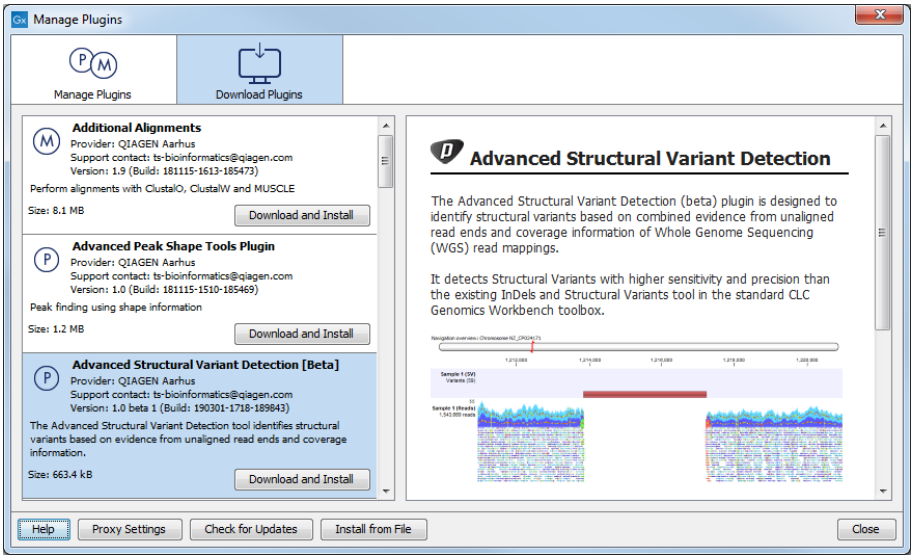

Figure 9.1: *The plugins that are available for download.*

9.2 Uninstall

Plugins are uninstalled using the plugin manager:

Help in the Menu Bar | Plugins... ( ) or Plugins ( ) in the Toolbar

This will open the dialog shown in figure 9.2.

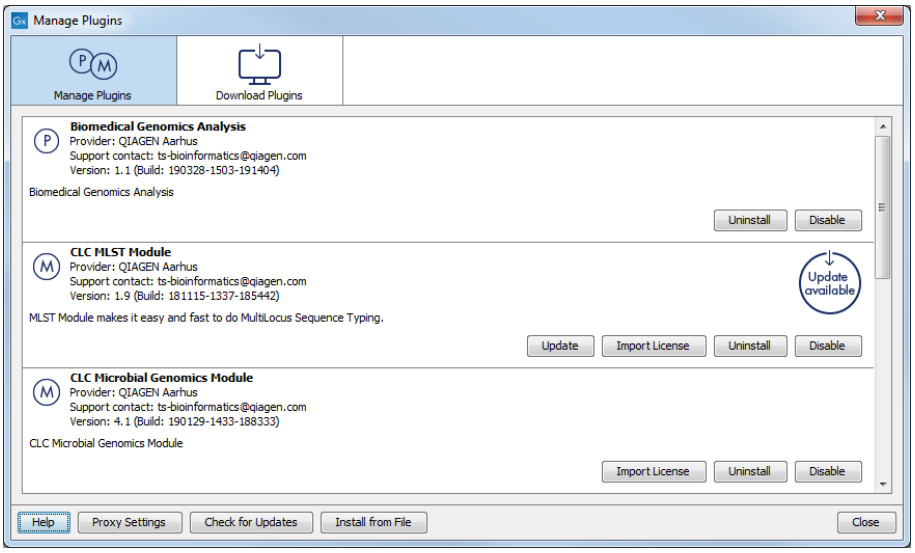

Figure 9.2: *The plugin manager with plugins installed.*

The installed plugins are shown in the **Manage plugins** tab of the plugin manager. To uninstall, select Ingenuity Variant Analysis and click **Uninstall**.

If you do not wish to completely uninstall the plugin, but you do not want it to be used next time you start the Workbench, click the **Disable** button.

When you close the dialog, you will be asked whether you wish to restart the workbench. The plugin will not be uninstalled until the workbench is restarted.