ORACLE

Oracle® Healthcare Translational Research

User's Guide

Release 3.3

E91297-01

October 2018

ORACLE®

Contents

1   What can I do with Oracle Healthcare Translational Research?

2   Modify criteria for queries using the Cohort Query

3   View and maintain cohorts

4   View data of a single patient or subject

5 Search genomic results with the Genomic Query

6 Manage queries, gene sets and jobs in My Workspace

Preface

This is your Oracle Healthcare Translational Research User Guide. This book will show you how to use the product and make the best out of OHTR.

Documentation Accessibility

For information about Oracle's commitment to accessibility, visit the Oracle Accessibility Program website at http://www.oracle.com/pls/topic/lookup?ctx=acc&id=docacc.

Access to Oracle Support

Oracle customers that have purchased support have access to electronic support through My Oracle Support. For information, visit http://www.oracle.com/pls/topic/lookup?ctx=acc&id=info or visit http://www.oracle.com/pls/topic/lookup?ctx=acc&id=trs if you are hearing impaired.

Finding more information

Oracle Help Center

The latest user documentation for Oracle Health Sciences products is available at https://docs.oracle.com/en/industries/health-sciences/.

My Oracle Support

The latest release notes, patches and white papers are on My Oracle Support (MOS) at https://support.oracle.com. For help with using MOS, see https://docs.oracle.com/cd/E74665_01/MOSHP/toc.htm.

Related documents

For more information, please access:

Oracle Business Intelligence Enterprise Edition Documentation

The OBIEE documentation is available at https://docs.oracle.com/middleware/12212/biee/index.html.

Oracle Healthcare Foundation

The OHF documentation is available at https://docs.oracle.com/en/industries/health-sciences/oracle-healthcare-foundation/index.html

1

What can I do with Oracle Healthcare Translational Research?

Oracle Healthcare Translational Research enables integrating, controlling, and analyzing the clinical and omics data required to support the complete biomarker life cycle. This includes data acquisition, discovery, and research as well as clinical use of patient and specimen information and cohort maintenance in clinical trials.

Use Oracle Healthcare Translational Research (OHTR) to:

- Select cohorts of patients or subjects using 10 clinical dimensions and 4 types of omics data type criteria.

- Save cohort queries and lists and share them with colleagues.

- View complete clinical and genomic information for a single patient or subject with the integrated Dalliance Genome browser.

- Use the built-in Visquick genome browser to view different types of genomic results in a circular plot.

- Search genomic data based on different genomic features like genes, genomic location, gene components, variants and so on.

- View information on patient or subject specimens linked to genes and variants.

- View and export specific clinical information for a selected cohort.

- View clinical activities on a visual timeline.

- View built-in cohort-driven reports on:

  - Demographic statistics

  - Clinical statistics

  - Genomic variants, copy number variation, gene expression, and structural variants

  - Genes compared to a sample matrix

  - Variants compared to a sample matrix

- Export various types of genomic data in VCF, SEG, GCT, and RES formats in synchronous and asynchronous mode.

- Schedule jobs for genomic data export.

- Drill in from one screen to another to further drill down and generate various reports.

2

Modify criteria for queries using the Cohort Query

In this chapter you can find out how to:

Before you begin building your query

Before you start creating your query, let's go through these tips. Keep this section in mind, you will probably come back to it from time to time.

- Prepare the query definition in a structured format (line-by-line) before creating it in OHTR. Visualizing your definition helps you organize it in a simpler way.

- When you want to search for a criterion, but can't exactly remember the name or the code, just click **Search** without selecting any conditions, or adding a keyword. The entire list of codes and names appears.

- Every criterion you want to add in your query can be included as an inclusion or exclusion. When you select either option, it applies to all of the elements in that criterion.

- By default, most dialog boxes for choosing criteria use the **Equals** operator and list the first 100 types only. If there are more than 100 types for one criterion that you want to include in your query, you can search for them using other operators such as **Starts with**, **Contains**, and so on.

Build a cohort query

You can query on any combination of the following data:

- **Patient Information**
  - Demographics
  - Consent

- **Clinical Data**
    - Diagnosis
    - Clinical Encounter
    - Procedure
    - Medication
    - Patient History
    - Test or Observation
    - Specimen
    - Study
    - Relative Time Events
- **Genomic Data**
    - Sequence Variants
    - Copy Number Variation
    - Microarray Expression
    - RNA-seq Expression
    - Metadata Filters

Patient Information

Demographics

1. On the Home page, on the left, click the **Cohort Query** icon( ).
2. In the **Cohort Query** screen, on the left, click the **Patient Information** category, then click **Demographics**.
3. Either enter the Patient ID or click the search icon ( ).

   To search for a **Patient ID**:

   a. Select a condition and enter part of a patient ID.

   b. Click **Search**.

   c. Select the patient ID that you want to use from the left side and move it to the right side using the arrow (>).

   d. Click **Submit.**
4. Set any other criteria that you want to use in your query:

   - **Gender**: Select the gender from the drop-down list.
   - **Marital Status**: Select the status from the drop-down list.
   - **Age**: Select a condition, and then add the age in years, as a single value or a range.
   - **Date of Birth**: Select condition, and then enter the date.
   - **Deceased**: Select this check box, if needed.
   - **Ethnicity**: Select the appropriate ethnicity.

- **Race**: Select the appropriate race.

- **Location**: Enter either a city, state, zip code, county or country.

- **Data Source**: Enter or search for a Data Source.

5. Under **Insert as** select:

    - **Inclusion** to include the criteria in the query.

    - **Exclusion** to exclude the criteria from the query.

6. Click **Submit**.

Consent

1. On the left, click the **Patient Information** category, then click **Consent**.

2. Enter a **Consent Type** or click the search icon (![search icon]).

    To search for a **Consent Type**:

    Note: Follow the same steps for the Consent Status.

    ---

    a. In the **Select Consent Type** window, select a condition for **Consent Type Name** or **Consent Type Code** and enter part of the consent name or type.

    b. Click **Search**.

    c. Select the consent type or code that you want to use from the left side and move it to the right side using the arrow (![arrow]).

    d. Click **Submit**.

3. Set any other criteria that you want to use in your query:

    - **Consent Status**: Enter a consent status, or follow steps from 2a to 2d, to search for a consent status.

    - **Consent Start Date**: Select a condition, and then add the date.

    - **Consent End Date**: Select a condition, and then add the date.

    - **Data Source**: Enter or search for a Data Source.

4. Under **Insert As** select:

    - **Inclusion** to include the criteria in the query.

    - **Exclusion** to exclude the criteria from the query.

5. Click **Submit**.

Back to "Build a cohort query" on page 2-1.

Have you checked "Before you begin building your query" on page 2-1 yet?

Clinical Data

Diagnosis

1. On the left, click the **Clinical Data** category, then click **Diagnosis**.

2. Either enter the diagnosis or click the search icon (![search icon]).

To search for a **Diagnosis**:

a. Select one of the **Search Modes**.

- **Classic Search**. Select it to use more search conditions for Diagnosis Name, Code, and Code System.

- **Hierarchy Drill-in Search**. Select it to search all of the codes in a hierarchy that are associated with a single Code System, and have only one parent code.

Note: This feature does not support mixed hierarchy, meaning different levels of codes belonging to different Code Systems or multi-parent hierarchy, meaning one code having multiple parent codes.

---

b. Select a condition for one or all: **Diagnosis Name**, **Diagnosis Code** or **Code System**.

c. Click **Search**.

d. Select the diagnosis codes that you want to use from the left side and move them to the right side using the arrow (>).

e. Click **Submit**.

3. Set any other criteria that you want to use in your query:

■ **Diagnosis Status**: Enter or search for a Diagnosis Status.

■ **Onset Date**: Select a condition, and enter a date.

■ **Reported Date**: Select a condition, and enter a date.

■ **End Date**: Select a condition, and enter a date.

■ **Age At First Onset**: Select a condition, and then add the age in years, as a single value or a range.

■ **Anatomical Site**: Enter or search for an anatomical site.

See To search for an Anatomical Site: for more details.

■ **Data Source**: Enter or search for a source of data.

4. Under **Insert as** select:

– **Inclusion** to include the criteria in the query

– **Exclusion** to exclude the criteria from the query

5. Click **Submit.**

Back to "Build a cohort query" on page 2-1.

Have you checked "Before you begin building your query" on page 2-1 yet?

Clinical Encounter

1. On the left, click the **Clinical Data** category, then click **Clinical Encounter**.

2. Either enter the **Encounter Type** or click the search icon ( ).

To search for an **Encounter Type**:

a. In the **Select Encounter Type** window, select a condition for **Encounter Name** or **Encounter Code** and enter part of the encounter name or code.

    b. Select the consent encounter codes and names that you want to use from the left side and move them into the right side using the arrow (>).

    c. Click **Submit**.

3. Set any other criteria that you want to use in your query: **Location**, **Time** or **Data Source**.

- **Location**: Enter or search for a location.

- **Time**: Select a condition, and then enter a time in days or in the form of a date.

- **Data Source**: Enter or search for a source of data.

4. Under **Insert as** select:

- **Inclusion** to include the criteria in the query.

- **Exclusion** to exclude the criteria from the query.

5. Click **Submit**.

Back to "Build a cohort query" on page 2-1.

Have you checked "Before you begin building your query" on page 2-1 yet?

Procedure

1. On the left, click the **Clinical Data** category, then click **Procedure**.

2. Either enter the **Procedure** and **Procedure Type** or click the search icon ( ).

To search for a **Procedure** and **Procedure Type**:

    a. In the new window, select a condition for **Procedure Name**, **Procedure Code** or **Code System** and enter part of the procedure name, code or code system.

    b. Click **Search**.

    c. Select the procedure code, name and code system that you want to use from the left side and move them into the right side using the arrow (>).

    d. Click **Submit**.

3. Set any other criteria that you want to use in your query: **Procedure Start Date**, **Procedure End Date**, **Anatomical Site**, **Procedure Outcome** or **Data Source**.

- **Procedure Start Date**: Select a condition, and enter the date.

- **Procedure End Date**: Select a condition, and enter the date.

- **Anatomical Site**: Enter or search for an anatomical site.

    See To search for an Anatomical Site: for more details.

- **Procedure Outcome**: Enter or search for a procedure outcome.

- **Data Source**: Enter or search for a data source.

4. Under **Insert as** select:

- **Inclusion** to include the criteria in the query.

- **Exclusion** to exclude the criteria from the query.

5. Click **Submit**.

Back to "Build a cohort query" on page 2-1.

Have you checked "Before you begin building your query" on page 2-1 yet?

Medication

1. In the **Cohort Query** screen, click the **Clinical Data** category, then click **Medication**.

2. Either enter the medication or click the search icon (▣).

   To search for **Medication**:

   a. Select one of the **Search Modes**:

      - **Classic Search.** More search conditions are available.

      - **Hierarchy Drill-in Search**. Select it to search all of the codes in a hierarchy that are associated with a single Code System, and have only one parent code.

      Note: This feature does not support mixed hierarchy, meaning different levels of codes belonging to different Code Systems or multi-parent hierarchy, meaning one code having multiple parent codes.
      ---

   b. Search for a medication either by **Medication Name**, **Medication Code** or **Code System** and click **Search**.

   c. Select the medication codes that you want to use from the left side and move them into the right side using the arrow (❯).

   d. Click **Submit**.

3. Set any other criteria that you want to use in your query:

   - **Medication Start Date**: Select a condition and enter a date.

   - **Medication End Date**: Select a condition and enter a date.

   - **Dosage**: Select a condition and enter a date.

   - **Dosage Units**: Enter or search for dosage units.

   - **Medication Outcome**: Select a condition and enter a date.

   - **Data Source**: Enter or search for a data source.

4. Under **Insert as** select:

   - **Inclusion** to include the criteria in the query.

   - **Exclusion** to exclude the criteria from the query.

5. Click **Submit**.

Back to "Build a cohort query" on page 2-1.

Have you checked "Before you begin building your query" on page 2-1 yet?

Patient History

1. In the **Cohort Query** screen, click the **Clinical Data** category, then click **Patient History**

2. Either enter the **Patient History** code or click the search icon (▣).

   To search for **Patient History**:

    a. In the Select Patient History window, select a condition for History Name, History Code or History Type and enter part of the History Code or History name.

    b. Select the history code, name and type that you want to use from the left side and move them into the right side using the arrow (>).

    c. Click **Submit**.

3. Set any other criteria that you want to use in your query:

- **Patient History Start Date**: Select a condition and enter a date.

- **Patient History End Date**: Select a condition and enter a date.

- **Amount**: Select a condition and enter a numeric value.

- **Amount Units**: Enter or search for the amount units.

- **Frequency**: Select a condition and enter a numeric value.

- **Frequency Units**: Enter or search for frequency units.

- **Patient History Value**: Enter or search for a patient history value.

- **History Applicable To**: Then select whom is it applicable to.

- **Data Source**: Enter or search for a data source.

4. Under **Insert as** select:

- **Inclusion** to include the criteria in the query.

- **Exclusion** to exclude the criteria from the query.

5. Click **Submit**.

Back to

Have you checked yet?

Test or Observation

1. In the **Cohort Query** screen, click the **Clinical Data** category, then click **Test or Observation**.

2. Either enter the test or observation code or click the search icon ( ).

To search for a **Test or Observation**:

    a. In the **Select Test or Observation** window, select a condition for Test or Observation Name, Code or Type and enter part of either the code, the name or enter the Test or Observation Type.

    b. Select the test or observation code, name and type that you want to use from the left side and move them into the right side using the arrow (>).

    c. Click **Submit**.

> **Note:** In the **Search Test or Observation** window, the Test or Observation Type by default uses the Equals operator and lists the first 100 types only. If there are more than 100 Test or Observation type values in your data, you can search using other operators like "Starts with", "Contains", and so on.

3. Set any other criteria that you want to use in your query: **Test Date**, **Result (Numeric) Search**, **Operator**, **Units**, **Result String**, or the **Data Source**.

   ■ **Test Date**: Select a condition, and enter a date.

   ■ **Result (Numeric) Search**: For **Based on absolute values**, select an **Operator**, and enter the units. For **Based on Reference Range** select an **Operator**, and an option for the **Compare** field.

   ■ **Result String**: Enter or search for a result string.

   ■ **Data Source**: Enter or search for a data source.

4. Select an **Anatomical Site**.

   See To search for an Anatomical Site: for more details.

5. Under **Insert as** select:

   - **Inclusion** to include the criteria in the query.

   - **Exclusion** to exclude the criteria from the query.

6. Click **Submit**.

Back to "Build a cohort query" on page 2-1.

Have you checked "Before you begin building your query" on page 2-1 yet?

Specimen

1. In the **Cohort Query** screen, click the **Clinical Data** category, then click **Specimen**.

2. Either enter the codes in the mandatory fields (marked with a double asterisk) or click the search icon ( ).

   To search for a **Specimen Type**:

   a. In the **Select Specimen Type** window, select a condition for the specimen type, type code, or code system and enter part of either the code, the name or the code system.

   b. Select the specimen code, name and code system that you want to use from the left side and move them into the right side using the arrow (>).

   c. Click **Submit**.

   Note: Repeat the same steps when searching for a Specimen Number and Specimen Vendor Number.

   ------

3. Set any other criteria that you want to use in your query:

   ■ **Anatomical Site**: Enter or search for an anatomical site.

   See To search for an Anatomical Site: for more details.

   ■ **Specimen Collection Date**: Select a condition and enter a date.

   ■ **Specimen Amount**: Select a condition and a numeric value.

   ■ **Units**: Enter or search for units.

   ■ **Data Source**: Enter or search for a data source.

4. Under **Insert as** select:

- **Inclusion** to include the criteria in the query.

- **Exclusion** to exclude the criteria from the query.

5. Click **Submit**.

Back to "Build a cohort query" on page 2-1.

Have you checked "Before you begin building your query" on page 2-1 yet?

Study

1. In the **Cohort Query** screen, click the **Clinical Data** category, then click **Study**.

2. Either enter the code for a study or click the search icon ( ).

   To search for a **Study**:

   a. In the **Select Study** window, select a condition for **Study Name** or **Study Identifier** and enter part of the name or identifier.

   b. Select the study identifier and study name that you want to use from the left side and move them into the right side using the arrow ( ).

   c. Click **Submit**.

3. Set any other criteria that you want to use in your query:

   ■ **Study Start Date**: Select a condition and enter a date.

   ■ **Study End Date**: Select a condition and enter a date.

   ■ **Data Source**: Enter or search for a data source.

4. Under **Insert as** select:

   - **Inclusion** to include the criteria in the query.

   - **Exclusion** to exclude the criteria from the query.

5. Click **Submit**.

Back to "Build a cohort query" on page 2-1.

Have you checked "Before you begin building your query" on page 2-1 yet?

Relative Time Events

You can search for patients who have a particular diagnosis, procedure, or medication combination in a particular chronological order. For example, you can search for patients who:

■ Received a diagnosis of *X* disease any time after starting medication *A*.

■ Had procedure *K* 30 days or more after receiving a diagnosis of *Y*.

■ Started medication *B* at the same time as having procedure *M*.

1. In the **Cohort Query** screen, click the **Clinical Data** category, then click **Relative Time Events**.

2. Under "Select Patients who have, Step 1" select **Diagnosis**, **Procedure**, **Medication**, **Test or Observation**, or **Genomic Variant.**

3. Specify how the events in **Step 1** and **Step 3** are time-related: "at any time while" or specify a condition.

4. Under Step 3, specify the related event: having Diagnosis, undergoing Procedure, or taking Medication.

5. Under **Insert as** select:

   - **Inclusion** to include the criteria in the query.

   - **Exclusion** to exclude the criteria from the query.

6. Click **Submit**.

Back to "Build a cohort query" on page 2-1.

Have you checked "Before you begin building your query" on page 2-1 yet?

Genomic Data

Sequence Variants

1. In the **Cohort Query** screen, click the **Genomic Data** category, then click **Sequence Variants**.

2. You can search on **Specimen Type** or **Anatomical Site**.

   To search for a Specimen Type:

   a. Either enter part or all of the value for **Specimen Type** or click the search icon ( ).

   b. In the **Select Specimen Type** window, select the condition for **Specimen Type Name**, **Specimen Type Code** or **Code System** and enter part of the code.

   c. Select the specimen code, name and code system that you want to use in your query from the left side and move them into the right side using the arrow ( ).

   d. Click **Submit.**

   To search for an Anatomical Site:

   a. Either enter part or all of a value for **Anatomical Site** or click the search icon ( ).

   b. In the **Select Anatomical Site** window, select one of the **Search Modes**:

      - **Classic Search**. More search conditions are available.

      - **Hierarchy Drill-in Search**. You can drill down to see the patient count for any sub-level.

   c. Search for an anatomical site either by **Anatomical Site Name**, **Anatomical Site Code** or **Code System** and click **Search**.

   d. Select any Anatomical Site codes, names and code systems that you want to use from the left side and move them into the right side using the arrow ( ).

   e. Click **Submit.**

3. You can also search for patients by different genomic criteria. In the **Select Patients based on** section, select one:

- **having Variants in selected Genes**: this enables you to search variants in specific genes.

- **having selected Genomic Variants** enables you to search for variants by their Cosmic or dbSNP identifiers.

- **having Variants within specified Genomic Region** enables you to search variants in specified genomic location.

- **having Zygosity** enables you to search variants with specific zygosity: Any, Homozygous, Heterozygous or Hemizygous.

- **having Genotypes** enables you to search for specific genotypes like AT, or AA, or wildtype (same as reference), and so on.

4. After selecting the genomic criteria, expand the **Variant Location** section and specify a value for the selected option.

   a. Select either **at Genomic Variant** or **at Genomic Position**. Add a value.

   b. Select an **Assembly Version**.

   c. Click V**iew Available Genotypes**.

5. Under **Insert as** select:

   - **Inclusion** to include the criteria in the query.

   - **Exclusion** to exclude the criteria from the query.

   > **Tip:** This applies to all of the criteria specified above.

6. Click **Submit**.

(Optional) Step 2 Additional criteria

To add additional parameters to your query in the **Step 2 tab**:

1. In the **Sequence Variants** window, click **Step 2 (Optional)**.

2. Select any or all of the following parameters:

   - **Variant Attribute**:

     a. Specify **Variant Types** from the options below: Substitution, Insertion, Deletion, Indel, or Complex.

     b. Select a **Variant Impact**: Synonymous, Missense, Nonsense, Unknown.

     c. Select a **Variant Status**: Select Any, Known, or Novel.

     d. **Strand**: Select Any, + for forward, or - for reverse.

   - **Nonsynonymous Substitution Scores**:

     a. Select either **with Polyphen** option or the **with SIFT** option.

     b. Select a condition for the **Version**.

     c. Either enter a **Transcript ID** or search for one by clicking the search icon .

   - **Variant Parameters Depending on the Sequencing File Type**:

     a. Select either **VCF**, **MAF** or **CGI master var**.

     b. Select the condition for all active fields according to your choice in **Step a**.

3. Under **Insert as** select:

- **Inclusion** to include the criteria in the query.

- **Exclusion** to exclude the criteria from the query.

4. Click **Submit**.

See "Metadata Filters".

Back to "Build a cohort query" on page 2-1.

Have you checked "Before you begin building your query" on page 2-1 yet?

Copy Number Variation

1. In the **Cohort Query** screen, click the **Genomic Data** category, then click **Copy Number Variation**.

   To search for a Specimen Type:

   a. Either add the **Specimen Type** or click the search icon ( ).

   b. In the **Select Specimen Type** window, select the condition for **Specimen Type Name**, **Specimen Type Code** or **Code System** and enter part of the code.

   c. Select the specimen code, name and code system that you want to use from the left side and move them into the right side using the arrow (>).

   d. Click **Submit.**

   To search for an Anatomical Site:

   a. Either enter part or all of a value for **Anatomical Site** or click the search icon ( ).

   b. In the **Select Anatomical Site** window, select one of the **Search Modes**:

      - **Classic Search**. More search conditions are available.

      - **Hierarchy Drill-in Search**. You can drill down to see the patient count for any sub-level.

   c. Search for an anatomical site either by **Anatomical Site Name**, **Anatomical Site Code** or **Code System** and click **Search**.

   d. Select any Anatomical Site codes, names and code systems that you want to use from the left side and move them into the right side using the arrow (>).

   e. Click **Submit.**

2. In the **CNV Attributes**, specify the **CNV Result Type** using the two options: numeric or categorized.

   a. If you select **numeric**, specify a time value for the **SNP Log2 Ratio** (Segment Mean).

   b. If you select **categorized**, specify a condition for the **CNV Type**, a value for the **CNV Type Score**, a condition for **Called Ploidy** and a value for the **Ploidy score**.

3. Specify the **CNV Location** using any or all options from the **in Genes from** list: Ad-hoc List, Pathway or Gene Set.

4. Enter or search for a value for any or all of the options you selected above.

5. Select a value for in **Gene Region**.

6. Specify the **Assembly Version**.

7. Specify the **DNA Reference Version**.

   About the DNA Reference Version: it represents the Ensembl reference version for getting gene annotations. This is related to Assembly Version and by default displays the preferred DNA Reference Version, which is set in OHO. The sample value for this is Ensembl Version 70.

8. Select the radio button for **at Genomic Position** and enter a value for it.

9. Under **Insert as** select:

   - **Inclusion** to include the criteria in the query.

   - **Exclusion** to exclude the criteria from the query.

     > **Tip:** This applies to all of the criteria specified above.

10. Click **Submit**.

See Metadata Filters.

Back to "Build a cohort query" on page 2-1.

Have you checked "Before you begin building your query" on page 2-1 yet?

Microarray Expression

1. In the **Cohort Query** screen, click the **Genomic Data** category, then click **Microarray Expression**.

2. To select patients with Microarray Expression Results:

   a. Either add the **Specimen Type** or **Anatomical Site** (both are optional) or click the search icon ( ).

      To search for a Specimen Type:

      a. In the **Select Specimen Type** window, select the condition for **Specimen Type Name**, **Specimen Type Code** or **Code System** and enter part of the code.

      b. Select the specimen code, name and code system that you want to use in your query from the left side and move them into the right side using the arrow ( ).

      c. Click **Submit.**

      To search for an Anatomical Site:

      a. Either enter part or all of a value for **Anatomical Site** or click the search icon ( ).

      b. In the **Select Anatomical Site** window, select one of the **Search Modes**:

         - **Classic Search**. More search conditions are available.

         - **Hierarchy Drill-in Search**. You can drill down to see the patient count for any sub-level.

   c. Search for an anatomical site either by **Anatomical Site Name**, **Anatomical Site Code** or **Code System** and click **Search**.

   d. Select any Anatomical Site codes, names and code systems that you want to use from the left side and move them into the right side using the arrow (>).

   e. Click **Submit.**

3. Select the **Microarray Attributes**: the Array Type, the conditions and values for Intensity, P-value and/ or Call.

   a. **The Array Type**: one-channel or two-channel.

   If you select one-channel, specify the needed conditions and values for **Intensity**, **P-value<** and **Call**.

   If you select two-channel, specify a condition for the **Log2Ratio**.

4. In the **Expression for Genes from** select either: Ad-hoc list, Pathway or Gene Set. Then select the **Assembly Version** and the **DNA Reference Version**.

5. Under **Insert as** select:

   - **Inclusion** to include the criteria in the query.

   - **Exclusion** to exclude the criteria from the query.

6. Click **Submit.**

See Metadata Filters.

RNA-seq Expression

1. In the **Cohort Query** screen, click the **Genomic Data** category, then click **RNA-seq Expression**.

To select patients with **RNA-seq Expression Results**:

   a. Either add the **Specimen Type** or **Anatomical Site** (both are optional) or click the search icon ( ).

   To search for a Specimen Type:

   a. In the **Select Specimen Type** window, select the condition for **Specimen Type Name**, **Specimen Type Code** or **Code System** and enter part of the code.

   b. Select the specimen code, name and code system that you want to use in your query from the left side and move them into the right side using the arrow (>).

   c. Click **Submit.**

   To search for an Anatomical Site:

   a. Either enter part or all of a value for **Anatomical Site** or click the search icon ( ).

   b. In the **Select Anatomical Site** window, select one of the **Search Modes**:

- **Classic Search**. More search conditions are available.

- **Hierarchy Drill-in Search**. You can drill down to see the patient count for any sub-level.

   c. Search for an anatomical site either by **Anatomical Site Name**, **Anatomical Site Code** or **Code System** and click **Search**.

   d. Select any Anatomical Site codes, names and code systems that you want to use from the left side and move them into the right side using the arrow (>).

   e. Click **Submit.**

2. Select a condition for the **RNA-seq Expression Attributes**: RPKM, Raw Counts, Median Length (Normalized) or Strand.

3. Select an option for the **RNA-seq Location**: Ad-hoc list, Pathway and Gene.

4. For each, either enter a value or click the search icon ( ).

   a. Select if the location is available for a certain Transcript ID and select the Assembly Version and the DNA Reference Version.

   b. Select at which Genomic Position the patient is (optional).

5. Under **Insert as** select:

- **Inclusion** to include the criteria in the query.

- **Exclusion** to exclude the criteria in the query.

6. Click **Submit**.

See Metadata Filters.

Back to "Build a cohort query" on page 2-1.

Have you checked "Before you begin building your query" on page 2-1 yet?

Metadata Filters

This option is available at the bottom of each of the genomic criteria screens.

To add metadata filters:

1. Expand the **Metada Filters** option.

2. Click **Add Metadata Filter** to open the **Select Metadata Attribute** window.

3. Search for a metadata filter based on **Attribute Name**, **Scope** or **Category** to get a list of attributes associated with metadata.

4. Select the attributes from the list that you want to use and click **Add to Filter**. When you have added all the filters you need, click **Done**.

5. For each attribute name from the table, select a condition in the **Operator** column, then enter a value for it in the **Value** column.

Example 2–1   Metadata Filters

Attribute Name	Operator	Value
ANALYSIS_ID	Starts with	TGTAID120

6. Under **Insert as** select:

- **Inclusion** to include the criteria in the query.

- **Exclusion** to exclude the criteria from the query.

7. Click **Submit**.

Back to "Build a cohort query" on page 2-1.

Have you checked "Before you begin building your query" on page 2-1 yet?

Combine inclusion and exclusion criteria

Figure 2–1   How inclusion and exclusion criteria work

For you to get the best results, you can use a combination of inclusion and exclusion criteria that can give the exact data that you need.

Once you have added several criteria to your query, you can:

- Expand or collapse the display for the Inclusions or Exclusions criteria.

- Select the arrow to the left of the criteria to expand or collapse their details.

Modify details of an existing criteria statement

You can continue to adjust the details of an existing statement using the icons on the right.

- The pencil icon (✎) displays and edits your criteria selection.

- The trash can icon (🗑) removes any particular statement.

- The plus icon (⊕) adds another set of criteria from the same topic. For example, clicking the plus icon from within a Demographics tab prompts you to select additional Demographics criteria.

Understand the logic behind Inclusion and Exclusion criteria

Within an Inclusion statement, distinct criteria statements related to the same data element (Procedure, Medication, Gender, etc.) is considered with an OR criteria.

Example 2–2   Examples of how Inclusion and Exclusion criteria work

INCLUDE: Diagnosis = Diabetes AND (Demographics = Male OR Demographics = Married)

Within an Exclusion statement, distinct criteria statements related to the same topic will be considered with an AND.

EXCLUDE: Diagnosis = Diabetes OR (Demographics = Male AND Demographics = Married)

This function can be used to simplify your query structure, where you can address each data topic and category one at a time rather than having the same topic listed in both the Inclusion and Exclusion sections.

View Patient or Subject Count and data

With this functionality, you can see the number of patients or subjects who meet specified criteria and the total number of patients or subjects in the database. You can also see those patients' or subjects' data.

1. On the Home page, on the left, click the **Cohort Query** icon ().

2. Select at least one row of criteria defined in the **Inclusions** or **Exclusions** options.

3. Click **Run Query**. The patient count is visible in the top right.

4. To view patient data, click **Patient Count**.

View Specimen Count

1. Along the top, click the **Options** drop-down.

2. Select **Show Specimen Count**.

3. The specimen count is displayed. This is the count of all the specimens in the cohort that match the selected criteria.

> **Note:** This count is shown for all specimens associated with a patient or subject, not just the ones specific to the selected criteria.

Save a query

After selecting the criteria for your query:

1. In the **Cohort Query** screen, click **Query Library**.

2. Click **Save Query** in the drop-down menu.

3. In the **Save Query** window, select:

   - **create new** to save it as a new query

   - **search existing** to save the query under an already existing name in the query library.

4. Click **Submit**.

Load a saved query

1. In the **Cohort Query** screen, click **Query Library**.

2. Click **Load Query** in the drop-down menu.

3. In the **Load Query** window, enter all or part of a **Query Name**.

4. Click **Search**.

5. Select the query.

6. Click **Submit**.

Figure 2–2   Warning message for adding or clearing criteria
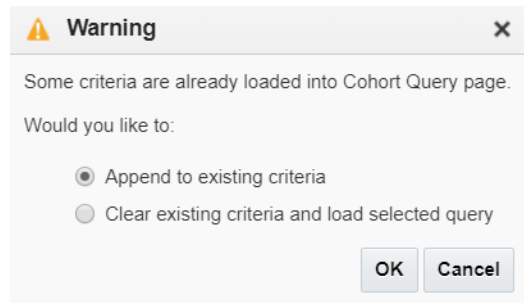
> **Note:**   If an existing query is currently running, you can either add the loaded query to the newly created one or you can clear the existing criteria and use only your loaded query.

3

View and maintain cohorts

In this chapter, you will find out how to:

- View cohort lists
- View cohort timelines
- Align data by Patient or Subject Event
- View cohort reports
- Export genomic data

View cohort lists

Figure 3–1   A cohort list with Demographic and Events data

1. On the Home page, on the left, click the **Cohort Viewer** icon (   ).

2. On the left side, in the **Patients** section, choose your option:

   - **active query**: to view the cohort list for a patient or subject query that is currently running

   - **library query**: to view the cohort list for a patient or subject query from the library

   - **list**: to view the cohort list for a saved patient or subject list

   - **ad-hoc query**: to view lists for specific patients or subjects you want to examine

3. Click **Submit**.

   Use the **Data** pane on the left to select and filter the data you want to view for the cohort list. For details on filtering data, see Filter data in cohort lists.

After running the query, the patient or subject count is displayed above the table, and you can review the data of those patients or subjects.

> **Note:** Patient or subject count can also be viewed in the **Cohort Query** screen as well. For details, see View Patient or Subject Count and data.

Filter data in cohort lists

1. On the left, below the **Patients** section, click **Data**.

2. Select or deselect any of the information that you want included in the cohort lists.

3. Either click the radio button for **Include criteria used in a query** or set your custom criteria.

> **Note:** The **Include criteria used in a query** can only be used for active queries.
>
> This option filters clinical records in cohort lists and cohort timelines only on the basis of the primary coded attributes from the Cohort Query screen. For example, it only uses Diagnosis Codes, Procedure Codes or Medication codes.
>
> If the inclusion criteria of the active query is not based on any data topics supported in the cohort lists, a message is displayed to select at least one supported data element.

To set custom criteria:

a. Select the criteria.

b. Enter the criteria or click the search icon ( ).

c. In the new window, enter all required details and click **Search**.

d. Select the elements that you want to use from the left side and move them to the right side using the arrow (>).

e. Click **Submit**.

4. Click **Submit**.

To remove data, clear the appropriate boxes and click **Submit** again.

To export the list to a Microsoft Excel sheet, click **Export** (Export), along the top of the table. Keep in mind that additional information, such as address and email are only visible in the exported XLS file.

To save a cohort list, click the **Save** button (Save Cohort List). You can rename the list, or save it under an existing name. All cohort lists are saved as **Private**.

To switch between Demographic and Events data, use these buttons above the table:

Figure 3–2 Switching between Demographics and Events data in the Cohort Lists

> **Tip:** To display the date in the dd/mm/yyyy format, use the formatting option in Microsoft Excel.

Display reference range values

Cohort lists can display reference range values, if available, along with numeric results of observation.

1. In the **Cohort Viewer** screen, on the left, click and the **Data** section.

2. Click the radio button for **Test or Observation**.

3. Enter the value you want to search for.

4. Click **Submit**.

   Where available, the reference high and reference low range values are displayed along with numeric results in the cohort list.

View cohort timelines

Figure 3–3   A cohort timeline containing data on diagnoses, procedures and medications of patients
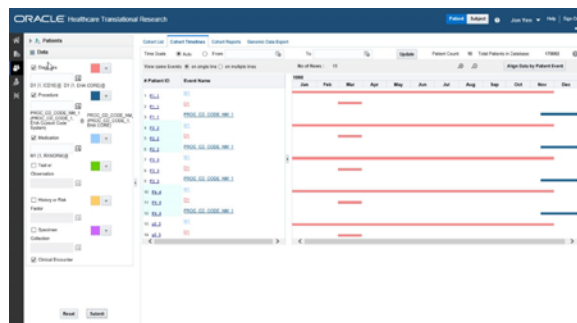
Select patients or subjects

This step is similar to how you select patients or subjects for the **Cohort List**.

1. Along the top, click the **Cohort Timelines** tab.

2. On the left, expand **Patients** or **Subjects** to select the source of your patients' or subjects' IDs from an active query, library query, list or ad-hoc query.

3. Click **Make initial pool**.

   All patients or subjects from the selected query are added to the initial pool.

4. Select up to 20 patients or subjects and click the arrow (>) to move them from the initial pool to the display list.

   The data displayed in the cohort timeline is of these 20 patients or subjects.

   > **WARNING:** Don't click Submit yet! Wait until the next section.

Filter data in cohort timelines

Tip: Each data topic has a distinct color associated with it, so that you can visually separate data for the same patient or subject. **You can change the assigned colors by using the drop-down arrow next to each color.**

1. On the left, click the **Data** section.

2. Change the colors of the data criteria, if needed.

3. Either click the radio button for **Include criteria used in a query** or set your custom criteria.

Note: The **Include criteria used in a query** can only be used for active queries.

This option filters clinical records in cohort lists and cohort timelines only on the basis of the primary coded attributes from the Cohort Query screen. For example, it only uses Diagnosis Codes, Procedure Codes or Medication codes.

If the inclusion criteria of the active query is not based on any data topics supported in the cohort lists, a message is displayed to select at least one supported data element.

To set custom criteria:

a. Select the criteria.

b. Click the search icon ( ).

c. In the newly opened window, enter all required details and click **Search**.

d. Select the elements that you want to use from the left side and move them to the right side using the arrow (>).

e. Click **Submit**.

4. Now that you have your data set, click **Submit** again.

Display patient or subject data

The subject and patient data display is divided into two sections.

■ The left side has attributes about the selected patients/ subjects: **Patient/ Subject ID** and **Event Name**.

■ To see demographic data for a Patient or Subject, hover over their ID.

■ The right side shows the colored layers of the selected clinical data, in a timeline view.

■ To see more details of a patient or subject and their medical records, click on an **Event** link from the left side table. A pop-up appears with more data on the event.

Figure 3–4 Event data shown in a pop-up window
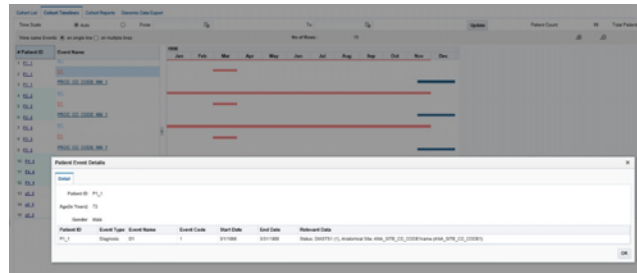
- To navigate to the Single Patient Viewer, and see details of a specific subject or patient, click the Subject or Patient ID.

- To view data for a specific date range, along the top, use the **Time Scale** options:

  - To view data for certain date range, enter the **From** and **to** dates and click **Update**.

  - To view data without a specific time reference, select the **Auto** option.

- Since the **Test or Observation** event has only one date event, the user interface uses dummy **End Date** values that are 1 day later than the corresponding **Start Date** values to render the graph. The **Detail** table also displays these dummy **End Date** values.

- Events are not displayed on the left-hand table or on the timeline in the following conditions:

  - When an event's **Start Date** is null.

  - When an event has only one **Date** attribute (for example, **Observation** has only **Observation Date**) and the attribute's value is null.

- If an event's **End Date** is null, then the **System Date** is used for displaying the event.

Select the Timeline Mode

To view events in different ways, go to **View same Events** along the top and select one of the radio buttons:

On single line

This is the default mode of displaying data in a timeline.

In this mode, all instances of a repeating event are displayed on the same line and separate events are displayed on separate lines.

On multiple lines

In this mode, all of instances of events are displayed on different lines, even if the events are repeating.

Align data by Patient or Subject Event

1. Click the **Align Data by Patient or Subject Event** button (Align Data by Patient Event).

2. In the new window, select a **Patient** or **Subject ID** from the drop-down list.

3. Click the search icon (![search icon]) next to the **Align by** field, and select the option you want to use.

> **Note:** When you select a particular diagnosis or medication code to align patients, all event codes at different hierarchical nodes are shown for the selected event code.

4. Click **Submit**.

Figure 3–5   How data is aligned by Patient Event
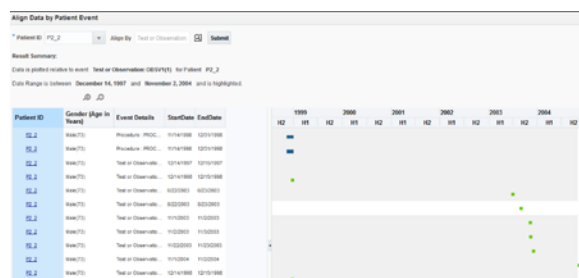
View cohort reports

The **Cohort Reports** tab lets you view various cohort reports.

> **Note:** In the Active Query mode, if no filters have been selected in Cohort Query, then the reports display data for only the first 10,000 patients or subjects. This limit is configurable and can be changed using the DEFAULT_ACTIVE_QUERY_LIMIT property in the TRC.properties file.
>
> However, irrespective of the patient or subject selection option, if the count exceeds the value specified in GENOMIC.REPORTS_MAX_PATIENT_COUNT of TRC.properties file, a warning message is displayed. You can either continue to plot with a large number of patients or subjects (might impact performance) or change the selected cohort to a smaller number.

View demographic reports

1. Along the top, click the **Cohort Reports** tab.

2. On the left, click the **Patients** or **Subjects** section, and select a source for your reports, if needed.

3. Click **Demographic Reports**.

4. Select the check boxes for the data you want displayed in your list.

5. Click **Submit**.

Figure 3–6   Demographic reports in OHTR
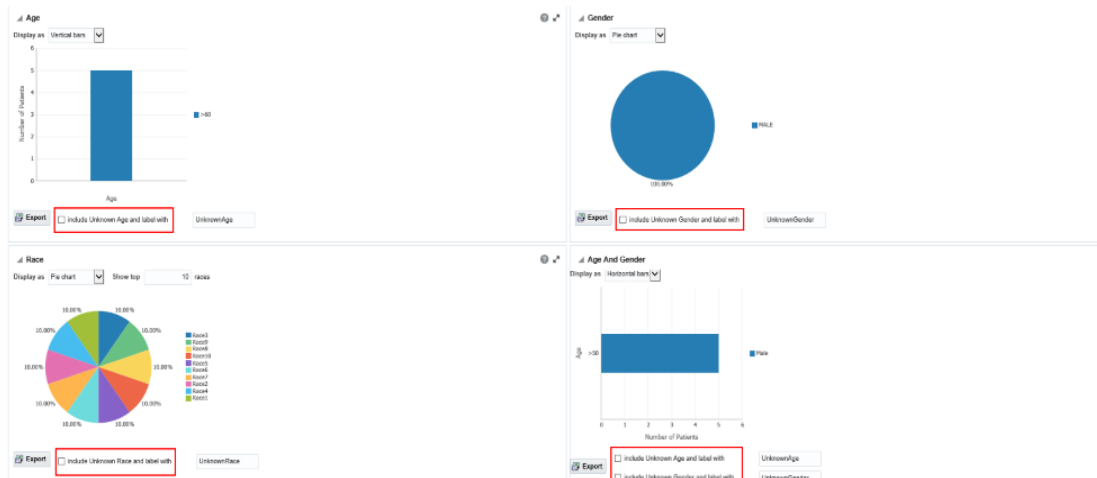
Handle unknown data

If any missing or unknown data is present, a check box is provided at the bottom of each graph to include missing or unknown data.

For the Age, Gender and Race graphs, if the check box is selected, the graph is refreshed with the unknown data. In this case, unknown data refers to age, gender or race details that are not defined for a particular patient.

For the Age and Gender graph, the unknown age for different genders is combined into one group and the known age is in another group. Both groups are shown in the graph.

View clinical reports

1. Along the top, click the **Cohort Reports** tab.

2. On the left, click the **Patients** or **Subjects** section, and select a source for your reports, if needed.

3. Click **Clinical Reports**.

4. Select the **Specimen Available** check box.

5. Click **Submit**.

Figure 3–7   A clinical report in OHTR
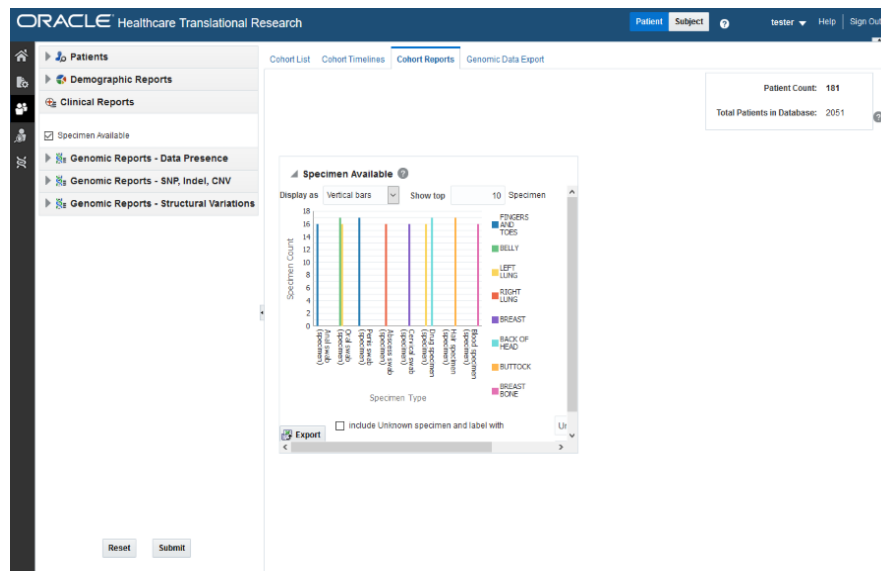
Handle Unknown Data

If any missing or unknown data is present, a check box is provided at the bottom of each graph to include missing or unknown data.

If **Include Unknown Specimen and Label With** or **Include Unknown Anatomical site and Label With** are selected, the graph is refreshed with the unknown data. In this case, unknown data refers to the specimen and anatomical site data that is not defined for a particular patient.

View genomic reports

View genomic data presence reports

1. Along the top, click the **Cohort Reports** tab.

2. On the left, click the **Patients** or **Subjects** section, and select a source for your reports, if needed.

    > **Note:**   There is no upper limit on the number of patients or subjects to be selected, however the performance slows down as more and more patients are selected.

3. Click **Genomic Reports - Data Presence**.

4. Select the **Genomic Data Presence** check box.

    An optional **Specimen Type** input field appears to allow for filtering the initial cohort on specimen type.

5. Click **Submit**.

    > **Tip:**   To change the display type after generating the report, select the display option from the **Display as drop-down** list.

To view a breakup of present genomic data by data type, click **View distribution of genomic data**.

To export the graphical plot type as a PNG image, and the table type as an Excel spreadsheet, click the **Export** button(![Export]) for every graphic.

Figure 3–8   A Genomic Data Presence report in OHTR
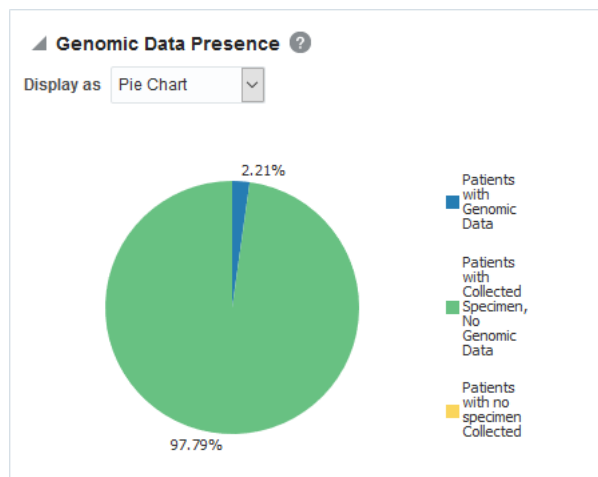
Figure 3–9   Distribution of Genomic Data in OHTR
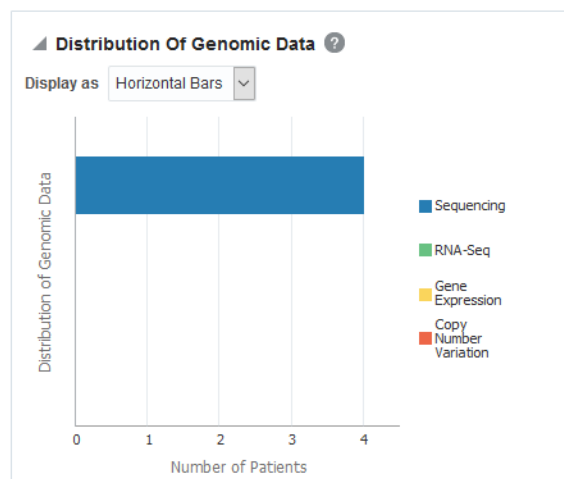
View Genomic Reports - SNP, Indel, CNV

> **Note:** Genomic Reports SNP, Indel and CNV display the available SNP Indel genomic reports based on the selected cohort of patients or subjects.

Mutated Gene Frequency and Gene Expression

To view these reports:

1. Along the top, click the **Cohort Reports** tab.

2. On the left, click the **Patients** or **Subjects** section, and select a source for your reports, if needed.

> **Note:** There is no upper limit on the number of patients or subjects to be selected. However the performance may slow down as the number of patients or subjects increases.

3. Click **Genomic Report - SNP, Indel, CNV**.

4. Select the radio button for **Gene Level Reports**.

5. Select **Mutated Gene Frequency and Gene Expression**.

   Selecting **Include All Specimens** includes specimens without genomic data. Not selecting this option generates results with genomic data only.

6. Enter or search for at least one gene using one of the three search fields: Ad-hoc List, Pathway or Gene Set.

7. Expand the **Additional Parameters** section to add a Specimen Type or an Anatomical Site.

8. Click **Submit**.

You can display the results as horizontal bars, vertical bars, or as a table. To export the bar results into PDF or the table results into Excel, click the Export button (Export).

Figure 3–10   A genomic report on Mutated Gene Frequency and Gene Expression
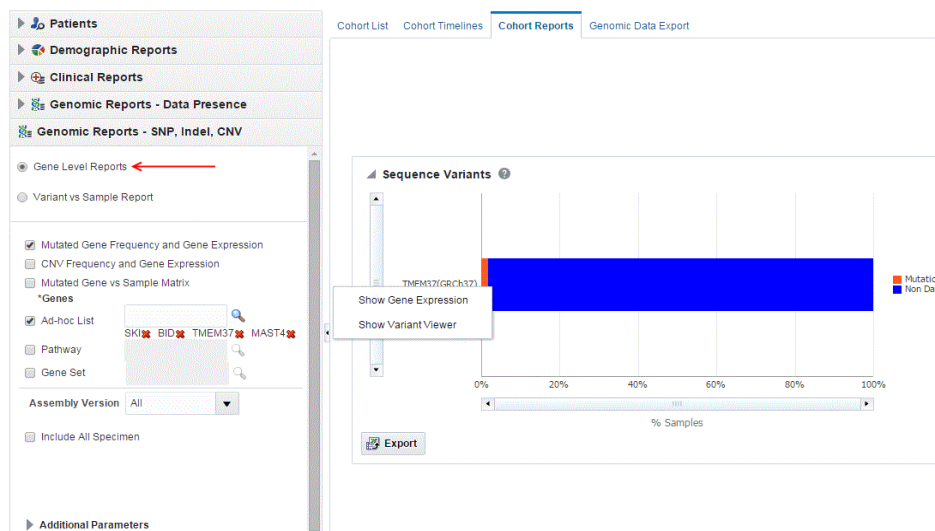
About the Mutated Gene Frequency and Gene Expression

Each bar represents a specimen, which is grouped patient-wise and in the order of the specimen collection date. The Sequence Variants and CNV information for each specimen is displayed. The percentage of specimens with mutation on the gene is also mentioned. Export functionality is provided at each specimen level.

There is a limit to the number of specimens that can be seen in the gene matrix plot. The default value of this parameter is 1000. If the number of specimens in the cohort used for this analysis is greater than the specified value, then the following warning message is displayed:

Figure 3–11   Warning message for reaching a limit on specimens

If you continue from this warning message, the report displays the summary statistics instead of the matrix plot. You may have to decrease the cohort size based on some criteria like specimen type or anatomical site or otherwise use cohort query to view the matrix plot. Alternatively, you can also change the default value of MAX_SPEC_REPORT parameter to a desired value and rebuild the plot. However, rebuilding maybe affect the performance of the plot generation.

Copy Number Variation Frequency and Gene Expression

To view these reports:

1. Along the top, click the **Cohort Reports** tab.

2. On the left, click the **Patients** or **Subjects** section, and select a source for your reports, if needed.

> **Note:**   There is no upper limit on the number of patients or subjects to be selected. However the performance may slow down as the number of patients or subjects increases.

3. Click **Genomic Reports - SNP, Indel, CNV**.

4. Select **Gene Level Reports**.

5. Select **CNV Frequency and Gene Expression**.

    Selecting **Include All Specimens** includes specimens without genomic data. Not selecting this option generates results with genomic data only.

6. Enter or search for at least one gene using one of the three search fields: Ad-hoc List, Pathway or Gene Set.

7. Expand the **Additional Parameters** section to add a Specimen Type or an Anatomical Site.

8. Click **Submit**.

> **Note:** When opening the Excel file, you may receive a warning from Excel stating that the file is in a different format than specified by the file extension. This warning can be safely ignored. For more information, refer to `http://docs.oracle.com/cd/E23943_01/web.1111/b31973/af_table.htm#autoId34`.

Figure 3–12   A genomic report on Copy Number Variation Frequency and Gene Expression in OHTR
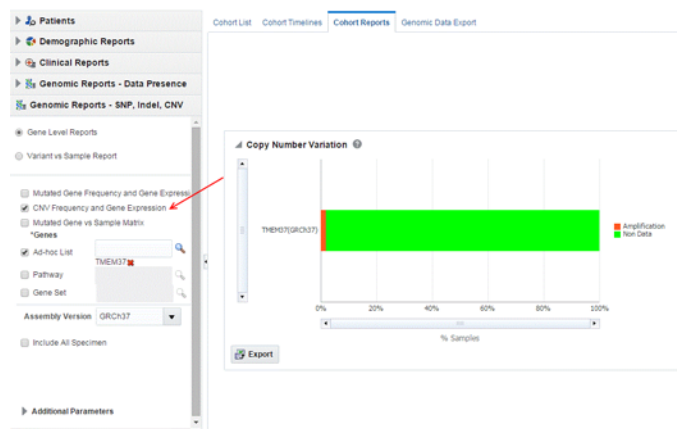
To export the bar results into or table results into Excel by clicking **Export** ( ).

Select **Show Gene Expression** to display the Gene Expression details for **Single Channel** or **Dual Channel**. In the **Single Channel Expression**, you must provide hybridization details.

Click **Show CNV Viewer** to display necessary information about all CNV variants that are present in that gene in the samples belonging to the selected cohort of subjects or patients.

Figure 3–13   How to switch between views in a CNV report
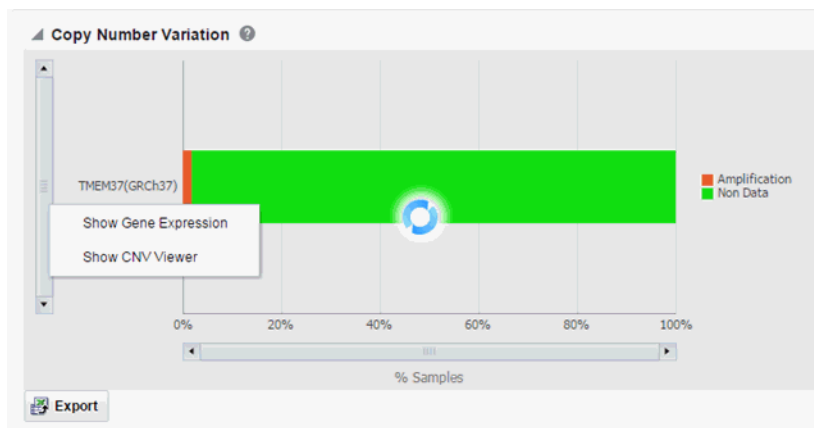
Figure 3–14   CNV Sample Details for a selected gene
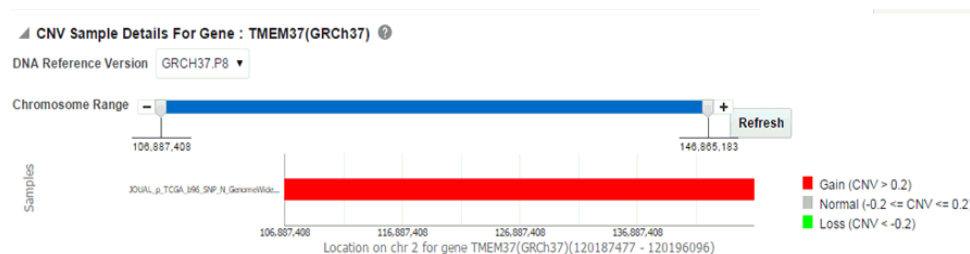
Mutated Gene vs Sample Matrix

To view this plot:

1. Along the top, click the **Cohort Reports** tab.

2. On the left, click the **Patients** or **Subjects** section, and select a source for your reports, if needed.

    > **Note:**   There is no upper limit on the number of patients or subjects to be selected. However the performance may slow down as the number of patients or subjects increases.

3. Click **Genomic Reports - SNP, Indel, CNV**.

4. Select **Gene Level Reports**.

5. Select **Mutated Gene vs Sample Matrix**.

6. Enter or search for at least one gene using one of the three search fields: Ad-hoc List, Pathway or Gene Set.

    Selecting **Include All Specimens** includes specimens without genomic data. Not selecting this option generates results with genomic data only.

7. Expand the **Additional Parameters** section to add a Specimen Type or an Anatomical Site.

8. Click **Submit**.

    > **Note:**   When opening the Excel file, you may receive a warning from Excel stating that the file is in a different format than specified by the file extension. This warning can be safely ignored. For more information, refer to http://docs.oracle.com/cd/E23943_01/web.1111/b31973/af_table.htm#autoId34.

Figure 3–15   A Mutated Gene Vs. Sample Matrix Report in OHTR

Figure 3–16   A Mutated Gene Vs. Sample Matrix Report ready to be exported
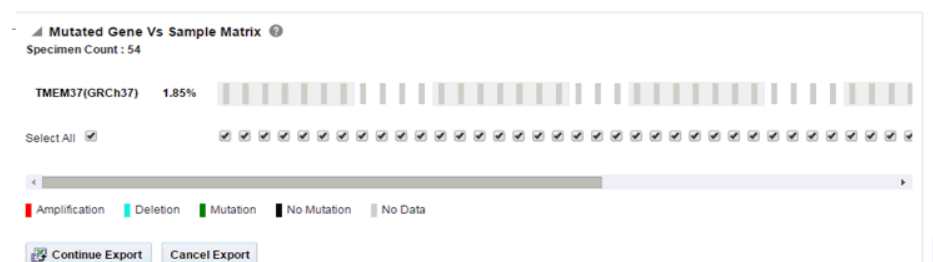
Figure 3–17   A Summary Report for Mutated Gene Vs Sample Matrix Reports
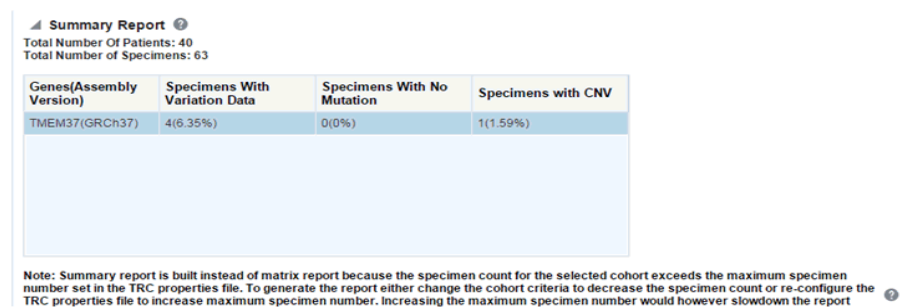
Variant vs Sample Reports

To generate the plot:

1.   Along the top, click the **Cohort Reports** tab.

2.   On the left, click the **Patients** or **Subjects** section, and select a source for your reports, if needed.

> **Note:**   There is no upper limit on the number of patients or subjects to be selected. However the performance may slow down as the number of patients or subjects increases.

3.   Click **Genomic Reports - SNP, Indel, CNV**, and select **Variant vs Sample Report**.

4.   Enter or search for a **Variant ID**.

Selecting **Include All Specimens** includes specimens without genomic data. Not selecting this option generates results with genomic data only.

5. Expand the **Additional Parameters** section to add a Specimen Type or an Anatomical Site.

6. Click **Submit**.

Figure 3–18   A Genomic Variant Vs Sample Report in OHTR
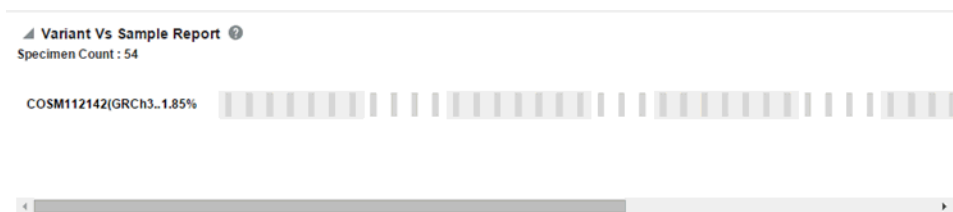
Figure 3–19   A Genomic Variant Vs Sample Report ready to be exported
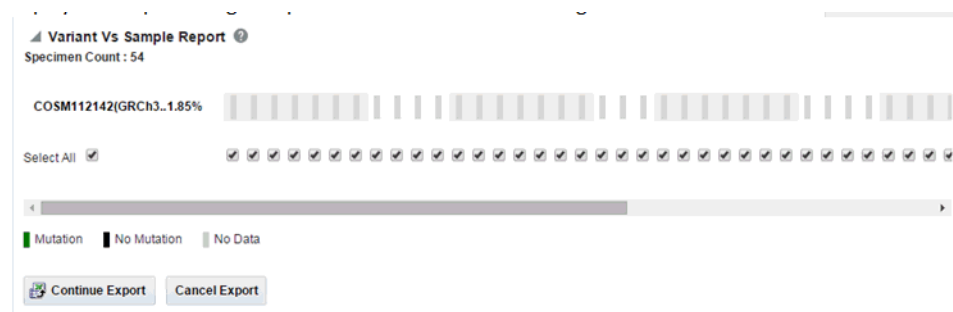
> **Note:** When opening the Excel file, you may receive a warning from Excel stating that the file is in a different format than specified by the file extension. This warning can be safely ignored. For more information, refer to http://docs.oracle.com/cd/E23943_01/web.1111/b31973/af_table.htm#autoId34.

Structural Variations

Structural Variations in Genes

To view the structural variations in genes:

1. Along the top, the click **Cohort Reports** tab.

2. On the left, click the **Patients** or **Subjects** section, and select a source for your reports, if needed.

> **Note:** There is no upper limit on the number of patients or subjects to be selected. However the performance may slow down as the number of patients or subjects increases.

3. Click **Genomic Reports - Structural Variations** and select **Structural Variations in Genes**.

4. Expand the **Additional Parameters** section to add a Specimen Type or an Anatomical Site.

5. Click **Submit**. By default, the top 10 genes are shown and you can increase or decrease the number of displayed genes.

    To export the bar results in PDF or table results in Excel, click **Export** (![Export]).

Figure 3–20   A Structural Variations in Genes Report in OHTR
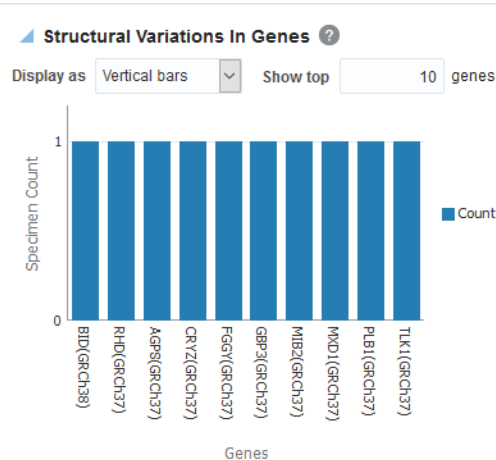
> **Note:** When opening the Excel file, you may receive a warning from Excel stating that the file is in a different format than specified by the file extension. This warning can be safely ignored. For more information, refer to http://docs.oracle.com/cd/E23943_01/web.1111/b31973/af_table.htm#autoId34.

Structural Variations in Gene Pairs

To generate a Structural Variations in Gene Pairs report in Subject context:

1. Along the top, click the **Cohort Reports** tab.

2. On the left, click the **Patients** or **Subjects** section, and select a source for your reports, if needed.

> **Note:** There is no upper limit on the number of patients or subjects to be selected. However the performance may slow down as the number of patients or subjects increases.

3. Click **Genomic Reports - Structural Variations** and select **Structural Variations in Gene Pairs**.

4. Expand the **Additional Parameters** section to add a Specimen Type or an Anatomical Site.

5. Click **Submit**. By default, top 10 gene pairs are shown and you can change the default number of bars shown.

   To export the bar results in PDF or the table results in Excel, click **Export** (⬆ Export).

   > **Note:** When opening the Excel file, you may receive a warning from Excel stating that the file is in a different format than specified by the file extension. This warning can be safely ignored. For more information, refer to `http://docs.oracle.com/cd/E23943_01/web.1111/b31973/af_table.htm#autoId34`.

Export genomic data

The **Genomic Data Export** page is used to export the genomic data for filtered patients or subjects based on **Study**, **Specimen type** and **Anatomical Site** in a specific file format.

> **Note:** Currently, exporting variation data from sequencing platform, single and double channel gene expression and Copy Number Variation data in VCF, SEG, RES and GCT file formats are supported. These formats are supported by the IGV browser.

Select results to export

Currently only one version of data can be exported at a time. To select results:

1. On the left, click the **Patients** or **Subjects** section, and select a source for your results.

2. Select the **Assembly Version** from the drop-down list.

3. Enter a **Specimen Type** or search for one using the search icon (🔍)

4. Enter an **Anatomical Site** or search for one using the search icon (🔍).

Set the location

You can export genomic data using three different methods:

- Click the radio button for **In Genes from** and select one or all options:

  - **Pathway** so you can select one or more pathways.

  - **Gene Set**, so you can use the user-defined collection of genes.

  - **Ad-hoc List** so you can select one or more genes.

- Click the radio button for **At Genomic Position** and enter a specific chromosome region.

  The following chromosome region formats are supported:

  ```
  CHR15:10000-200000: Considers region between 10000 to 200000 in chromosome 15.
  CHR15:1,200,000+5000 - Considers 5000 bases upstream from 1,200,000 position
  in chromosome 15.
  CHR15 - Considers whole of the chromosome 15.
  ```

```
CHR15:1000 - Considers 1000th nucleotide position of chromosome 15.
```

■ Click the radio button for **All Data** to export the complete genomic data of patients or subjects.

Select file type

Select one or all of the available file types:

■ Mutation - VCF

■ Copy Number Variation - SEG

■ Microarray Expression - RES

■ Microarray Expression Dual Channel - GCT

> **Note:** For the **At Genomic Position** location selection, the Gene Expression - RES and Gene Expression Dual Channel options are disabled.

About the file types

Mutation - VCF

This option exports the sequencing variation data for the selected patients or subjects for either the selected genes, pathway, gene set or for a given chromosome region as selected in the previous option. VCF supports multiple specimens' data in a single file.

The metadata header gives the following information that differs based on the search criteria:

1. ##fileformat=VCFv4.1

2. ##fileDate: Date and time of the VCF file generated.

3. ##source=Oracle Healthcare Omics (OHO, formerly known as Omics Data Bank)

4. ##Total Number of patients included in this VCF file

5. ##Total Number of samples included in this VCF file

6. 7.##INFO=<ID=NS, Number=1, Type=Integer, Description=Number of Samples With Data>

7. ##FORMAT=<ID=GT,Number=1,Type=String,Description=Genotype>

8. ##FORMAT=<ID=GQ, Number=1, Type=Integer, Description=Genotype Quality>

9. ##FORMAT=<ID=GQVAF, Number=2, Type=Integer, Description=Genotype_ quality_X>

10. ##FORMAT=<ID=DP, Number=1, Type=Integer, Description=Read Depth>

11. ##FORMAT=<ID=AD,Number=.,Type=Integer,Description=Allelic depths for the ref and alt alleles in the order listed >

12. ##FORMAT=<ID=HQ, Number=2, Type=Integer, Description=Haplotype Quality>

13. ##FORMAT=<ID=BQ,Number=.,Type=Integer,Description=Average base quality >

14. ##FORMAT=<ID=MQ,Number=.,Type=Integer,Description=Average mapping quality >

15. ##FORMAT=<ID=SS,Number=1,Type=Integer,Description=Variant status relative to non-adjacent Normal,0=wildtype,1=germline,2=somatic,3=LOH,4=post-transcriptional modification,5=unknown>

16. ##FORMAT=<ID=SSC,Number=1,Type=Integer,Description=Somatic Score>

The following data types are imported to VCF file:

Data Type	Description
CHROM	chromosome
POS	position of the variation
ID	dbSNP ID or COSMIC ID associated with a variant
REF	reference allele
ALT	variant alleles
QUAL	not populated. Will have '.' specified in this column.
FILTER	is populated as PASS
INFO	Not populated. Will have '.' specified in this column.
FORMAT:GT	genotypic data for each specimen.
FORMAT:GQ	genotype quality. If not value available in DB, then '.' is specified in the file.
FORMAT:GQX	mapped to GENOTYPE_QUALITY_X column
FORMAT:DP	this stores the TotalReadCount for a specific variant
FORMAT:AD	this stores the reference read count and Allele read count for a specific variant.
FORMAT:HQ	not populated as of now. Will have '.' specified in this column.
FORMAT:FT	this stores GENOTYPE_FILTER column value
FORMAT:BQ	stores the RMS base quality
FORMAT:MQ	stores the RMS mapping quality
FORMAT:SS	stores the somatic status
FORMAT:SSC	stores the somatic status score value
Flex field format	If any custom formats are available, they are also included in the export.

1000 Genomes VCF 4.1 conventions are followed while exporting variation data, however certain data types, which are non-standard, like BQ and MQ, may differ in convention for some customers since there is no standard way to represent them.

Handle Non-variant and No-call Data

If NON_VARIANT and (or) NOCALL records exist for any given position, the zygosity is checked to determine if the information format from these tables is used.

> **Note:** For *het-ref* or *half* zygosity values, these other format fields are compared with the existing SEQUENCING information. This information is then used with zygosity to create the format string.

The NON_VARIANT data allows for GQ, GQX, MQ, BQ and the first reference read count of AD. The NOCALL data allows for all format fields to be compared. Both NON_VARIANT and NOCALL do not support exporting flex fields. The GT value of the format string reflects the stored zygosity as follows:

Zygosity	FORMAT string GT:GQ:GQX:BQ:MQ:AD:DP
het-ref	1/0:99:98:38:45:20:10,10
Half	1/.:99:98:34,34:45,45:20:10,5
Het-alt	1/2:99:98:43,44:56,67:20:0,10,10
Hom	1/1:99:98:34,34:45,45:20:0,19

If there are no result records for any specimen, the export displays "." with no other information for the format.

Handle Duplicate Genetic Information Exports

There could be cases where users reload genetic information multiple times for the same specimen. This may create ambiguous values for the different fields that exist in the VCF export file. The export code deals with such ambiguous numerical values that represent the quality (that is, GQ, GQX, AD, BQ, MQ). This code now computes minimum values and ensure that the value of least confidence is reported. There could be more complex cases, for instance, if there are 2 different alleles for the same position belonging to the same specimen, or variants with same position for same specimen with different zygosity. The export code uses MIN functions on all values including all the text fields. This allows for VCF export to create a valid file that can be loaded into genome browsers.

Alternatively, you can choose not to consider data from a specific specimen or a specific file using following methods:

■ Using DELETE_FLG - A user may load results for a specimen more than once that can completely contradict previous results. Users can set the DELETE_FLG as 'Y' on W_EHA_RSLT_SPECIMEN and (or) W_EHA_SPEC_PATIENT or W_EHA_ SPEC_SUBJECT to have previous loads excluded, and then reload the correct result files. When the user now exports the data, only the latest loaded specimen data is considered for export.

■ Using FILE_URI - Oracle recommends using this method since you need not reload the data again as opposed to the above method. When there are multiple files loaded with contradicting data for the same specimen, user can set some files as obsolete by changing the W_EHA_FILE_LOAD.FILE_WID column. For example, if you have loaded the same specimen data 3 times and would like to consider the latest file loaded for export, then you must first identify the latest FILE_WID from W_EHA_FILE_LOAD table. Then change the FILE_WID of two old files in W_EHA_FILE_LOAD table to the latest FILE_WID. Now, all the three records belonging to the three file loads contain same FILE_WID, which represents the latest file load and only the latest file export data is exported.

Represent AD Values

Allele depth values represented under the AD data type are in the order of the alleles represented in the GT. Refer to the following table with examples:

ALT	FORMAT	SAMPLE1	
G,C,T	GT:AD	1/2:0,4,6	0 represents reference_read_count 4 represents allele_read_count of 'G' 6 represents allele_read_count of 'C'
G,T	GT:AD	2/2:0,4	0 represents reference_read_count 4 represents allele_read_count of 'T'
G,T	GT:AD	1/0:10,5	10 represents reference_read_count 5 represents allele_read_count of 'G'

Copy Number Variation - SEG

The copy number variation data is exported in SEG format. Currently, CNV data from any array based system like Affymetrix Genome Wide SNP 6 array whose data is in SEG format while loading in OHO is supported. The main requirement for exporting CNV data is to have the SEG_MEAN value in the CNV table of OHO.

For exporting data that is not loaded from SEG files, for example, data from CGI CNV files or any other source of CNV data, users have to create their own loader. The loader is expected to calculate the SEG_MEAN value since this value is most important for export.

1. ID: specimen ID of the reported CNV segment

2. chrom: chromosome name

3. loc.start: start position of the CNV segment

4. loc.end: end position of the CNV segment

5. num.mark: for array based CNV data, this stores the number of probes details

6. seg.mean: this stores the segment mean value from SEG_MEAN column in CNV table.

Gene Expression - RES

RES is one of the gene expression formats supported by IGV browser. Currently, only microarray gene expression data is exported to this format. Following data types are imported to RES format:

1. Description: hugo name of a specific probe

2. Accession: probe ID

3. Intensity: intensity value of the associated probe

4. Call: call of the associated probe

Gene Expression Dual Channel - GCT

GCT is one of the gene expression formats supported by IGV browser. Currently, only AgilentG4502A platform microarray gene expression data is exported to this format. Following data types are imported to GCT format:

1. Description: Gene symbol of a specific probe

2. Accession: probe ID

3. Intensity: intensity value of the associated probe

4. Call: call of the associated probe

> **Note:** The GCT file takes its gene symbol for the probe from the 2-channel composite element of ADF file. This is input into the ADF composite table in OHO. This value may not match with HUGO name in certain cases as OHTR associates 2-channel records in the result table that has partial (which includes a flanking region set by the user) genomic coordinate. The coordinate overlaps between composite elements and gene segments in the reference. This may also result in some cases in more than one unique gene in the reference mapping to a gene composite element.

Specify the export options

Export data in three ways:

- **Select the option to download last loaded file(s)**
- **Immediately**, which is the default option
- **Schedule**

After selecting the **Export** option, click **Submit** to finalize the genomic data export. The data is generated and a separate link is provided in the bottom panel for each result type.

About Export Options

The **Immediately** option gives you the file link on the same screen and you can click to download it immediately. The link provided has a specific naming convention: *<file type>_OHO_<date:MM-DD-YYYY>_<time:HH24*-MI-SS>.<file_type_extension>*. For example, RES_OHO_09-14-2014_04-26.res. A short description of the file stating data type and advice on the expected count of features is displayed below the created link.

The **Schedule** option runs the process as a job. You can track the status of the job from the **Home**, **Jobs** tab. This option is best suited for exporting large data sets like **All Data** or whole chromosome variants. For the scheduling option, you must provide a job name and description.

There is a possibility of replicate and duplicate data in the database. This could be due to loading multiple files belonging to the same specimen_number. This can happen if the same library is sequenced multiple times or the data is reanalyzed. For example, the reads were realigned using the new reference version, so new VCF or gVCF files are created for the same sample. In this scenario, you can use the option to export VCF data only from last loaded files. For example, if variation data has been loaded for a specimen in Jan 2015, Mar 2015 and July 2015, then using this option you can export data from the file loaded in July 2015 and it would not consider variants from the file loaded in Jan and Mar 2015.

Note: The **Schedule Jobs** option uses an asynchronous approach to store the file in DBFS. As an alternative to downloading the file using the link in the **Oracle Healthcare Translational Research Jobs** page, there are other ways to access DBFS. From a Linux OS, you can mount DBFS using *dbfs_client* application and then browse the directories. Windows OS does not support the FUSE interface and cannot mount DBFS directly. However, there is a *dbfs_client* application for Windows that can execute commands to access DBFS. The Windows version of *dbfs_client* lets you use the command line to execute normal directory commands. You can list the DBFS directories as well as copy data from DBFS to the local drive. The *dbfs_client* application is part of the standard Oracle client software.

For more information about using *dbfs_client*, see http://docs.oracle.com/cd/E11882_01/appdev.112/e18294/adlob_client.htm#ADLOB0006.

4

View data of a single patient or subject

In this chapter you will find out how to:

- View records
- View circular genomic data

View records

To view a patient record:

1. On the Home page, on the left, click the **Single Patient Viewer** icon ( ).

2. Enter the patient ID.

3. Select the **Source**:

   - **All records** is the default option. This allows you to search through all available patients in the CDM database.

   - **Cohort**

4. Select the clinical and genomic data you want to view for the patient.

5. Click **Submit**.

   > **Note:**   Attributes displayed in the S**ingle Patient Viewer** align with the criteria described in the **Cohort Query** screen, where you select criteria for patients and subjects.

To sort patient or subject data, either click the phenotype column to reorder them, or enter a value in the blank field, at the top of any column, then hit **Enter**.

To save patient or subject data in an Excel file, click the **Export button** (Export) next to the category title. You can export either all data or just the selected criteria.

Variants can only be exported separately. To do so, follow these steps below:

1. In the Single Patient Viewer screen, on the left, select the Variants checkbox.

2. Click **Submit**. The Variants hierarchy graph is displayed on the right side of the screen.

3. Click on any nodes in the graph to display the tables with Variants data and select it. These tables show 500 records only.

4. Along the top, click **Export** to export the variants. Variants are exported in an XLS file for the selected node.

Remember that variant export is not supported for the top two level nodes in the Variant hierarchy graph.

If the number of variants that you want to export is less or equal than the pre-configured limit of 50000, then all variants for the selected node are exported. If the number of variants that you want to export exceeds the limit, variants are then exported only for the selected chromosome within the node.

> **Note:** If the number of variants that you want to export exceeds the limit of 50000, contact your database administrator. The limit can be modified within the database. For more details, see the *Oracle Healthcare Translational Research Administrator's Guide* on the Oracle Help Center.

Navigate through selected patients or subjects

To navigate through patients' or subjects' data in the **Single Patient Viewer,** keep in mind three things:

1. If the selected patient ID generates a patient or subject count larger than 1, you can navigate through patients' or subjects' history using the **Previous** and **Next** buttons found next to the **Patient ID** field, on the left. This is also available for patient or subject counts generated in the **Cohort Query** or **Cohort Viewer** screens.

2. Every patient has an ordinal number (his position within the selection) listed under the **Patient** field. If the selected source is a cohort, then the total number of available patients or subjects from that cohort is also displayed.

3. If you went from the **Cohort Timeline** screen to the **Single Patient Viewer** screen, then the total number of available patients or subjects in the cohort is the count of patients or subjects selected in the initial pool, in the cohort timeline.

View genomic data

Genomic data is displayed in four sections as follows:

Specimen with Genomic Results

Table 4–1   Specimens with Genomic Results

Column Heading	Definition	Sample Value or Values
Specimen Id	Specimen belonging to the selected patient or subject	HG00096
Specimen Vendor Id	Specimen Vendor for that specimen	Vcf
Version Label	Represents Assembly Version (DNA reference version against which this data was loaded)	GRCh37(V68)
Sequence Variants Results	Whether the sample has sequence variants results	Yes / No
Copy Number Variation Results	Whether the sample has copy number variants results	Yes / No
Single Channel Microarray Results	Whether the sample has single channel results	Yes / No
Dual Channel Microarray Results	Whether the sample has dual channel results	Yes / No
Rna-Seq Expression Results	Whether the sample has rna sequencing results	Yes / No

Derived Files

Table 4–2   Derived Files

Column Heading	Definition	Sample Value or Values
Alternate Filepath	The FTP path of the file	
File Type, Version	Type of file and Version	Variant Call Format, 4.1
Result Type	Type of result data in the file	Sequencing, Copy Number Variation, Gene Expression (2-channel or single channel)
Alignment Version (DNA Reference Version)	Represents Assembly Version (DNA reference version against which this data was loaded)	GRCh37(v68)
Total Number of Specimen in File	Total number of specimen that the file contains where not all specimen belong to the selected patient	Numeric, positive integer

File Lineage

> **Note:** If you have appropriate permissions, and if files are present in the middle tier accessible location, the path listed in the **File Name** fields for Genomic data are enabled to let you click and download the files directly from OHTR.

Table 4–3   File Lineage

Column Heading	Definition	Sample Value or Values
Parent Filename	Parent Filename including path of the genomic file stored including path	C:/John_specimen01.BAM
Alternate Filepath	The FTP path of the file	
Alignment Version (DNA Reference Version)	Represents Assembly Version (DNA reference version against which this data was loaded)	GrCH37
Derived Child Files: File - Specimen Id, Vendor Id	Information about Derived Results files that have their lineage based on the particular Low Level file.	C:/John_specimen01.vcf - HG00096_1,HarvardLab1

Variants Found

Selecting this option displays different variants available for a patient. These variants are grouped and displayed in a hierarchical structure with the count of the variants displayed for each type of variants.

> **Note:** Variants can only be exported after they've been submitted.

Variant details are displayed in a table as follows:

Table 4–4   Variant Details

Column Heading	Definition	Sample Value or Values
Chromosome	Chromosome location of the variant	1
Position	The position of the variant within the chromosome	Numeric value
Reference Allele	The reference allele of the variant	C
Alternate Allele	The alternate allele of the variant	G
Gene	The gene containing the variant	BID

Table 4–4   (Cont.) Variant Details

Column Heading	Definition	Sample Value or Values
Transcript	The transcript name	ENSTXXX
Variant Name	The reference id of the variant	rs111
Variant Type	The type of variant	Substitution
Variant Status	Status of the variant	Known
Protein Name		
SIFT Impact	The SIFT impact of the variant	Intolerant
PolyPhen Impact	The polyphen Impact of the variant	Damaging
Drug	The related drug	Clofazimine
Associated Disease	The disease associated to the variant	Anaemia
Histology		
Site		
Specimen Id	The specimen containing the variant	HG00096
Alignment Version	Alignment version of a variant	GRCh37

Select the Assembly Version

By default, the last loaded assembly is displayed but this selection can be modified. The Genomic data is filtered out with this assembly version instead of the DNA reference version. One Assembly version can belong to multiple DNA reference versions.

Use the Dalliance Browser

To view variants, chromosomes or gene ranges:

1. On the left, in the patient details screen, scroll down to **Variants**. The variant detail table hyperlinks for variant reference IDs, gene names and chromosome positions.

2. Click any of those hyperlinks to navigate to the Dalliance browser where you can see the particular VCF tracks.

   If a file is loaded multiple times for a particular specimen, then the following warning message is displayed. It lists all the multiple files names.

Figure 4–1   Warning message for a file loaded multiple times in Single Patient Viewer
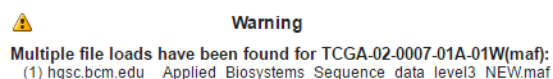
3. Scroll down to see some configurations that are required to plot the gene track in the Dalliance browser.

You can create your own DAS server and the corresponding entries should be added in the TRC.properties file.

The following is an example of the required entries, to look up the Authority, University of California Santa Cruz (UCSC) name for the alignment and the public URLs for the Sequence and Genes tracks (reference tracks):

```
DALLIANCE.AUTHORITY_37=GRCh
```

```
DALLIANCE.UCSC_NAME_37=hg19
DALLIANCE.SEQ_URL_HG18=http://www.derkholm.net:8080/das/hg18comp/
DALLIANCE.SEQ_URL_HG19=http://www.derkholm.net:8080/das/hg19comp/
DALLIANCE.GENES_URL_HG18=http://www. derkholm.net:8080/das/hsa_54_36p/
DALLIANCE.GENES_URL_HG19=http://www. derkholm.net:8080/das/hsa_59_37d/
```

While integrating Dalliance with OHTR, the server `http://www. derkholm.net:8080/das/` has been used. However, this has been shut down.

These can be manually customized in TRC.properties. To add any new alignments, add all the above code to TRC.properties. Make sure that you research the reference for where they can be found. It is possible to use downloaded files instead of public DAS server(s) for reference, but the client should host these files on a web server.

About the Dalliance Browser

The Dalliance browser is a third party tool that displays a graphical representation of the variant, chromosome or gene range. The variant detail table comprises of hyperlinks for variant reference ID, gene name and chromosome position.

View circular genomic data

1. On the left, click the **Circular Genomic Viewer** tab. Select a **Patient ID** or **Subject ID**.

   If an ID is selected in the **View Record** tab, it's inherited in the **Circular Genomic Viewer** tab.

2. To filter your view, you can select a **Specimen Type** or an **Anatomical Site**. The DNA Reference Version selected is used to filter out the results and determine the cytoband to be used while rendering the circular genomic plot for any of the five data types.

3. Select an **Assembly Version**.

4. Click **Submit**.

5. Select only one specimen of different result types to plot the graph. By default, the cytoband of chromosomes is also plotted which is the outer most ring of the circular plot.

6. Click **Plot Graph**.

Note: For the circular viewer to run properly in IE9, turn off compatibility mode.


About the Circular Genomic Viewer

Circular genomic data viewer provides an interface for you to visualize the genomic data which includes variation, micro array expression, copy number variation, dual channel expression and RNA sequencing. The system uses the VisQuick tool, which is a Javascript library built specifically for genomic data visualization.

Select data to plot

Figure 4–2   Selecting data to plot for the Circular Genomic Viewer
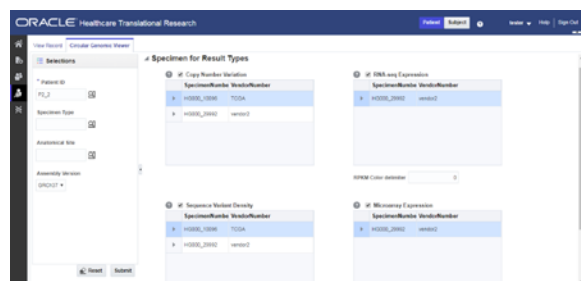
Microarray Expression

The microarray expression panel displays all the specimens of the result type expression. You can select only one specimen at a time.

For the selected specimen from the panel, you can view the list of hybridizations available for that specimen. You can select a maximum of two hybridizations from the multiple choice box, for which the plot is rendered.

The color delimiter value is used to render color to the data points in the plot based on the intensity value. If the intensity value is above the value defined in color delimiter then the data points have green color Otherwise, the points have red color.

Sequencing-Variant Density

This option enables you to plot the density of variants for every 100-kb region for a specimen. The sequence variant density panel displays all the specimens of the result type sequence variants. You can select only one specimen at a time.

RNA-Seq Expression

This panel displays all the specimens of result type RNA Sequencing.You can select only one specimen at a time.

The color delimiter value is used to render color to the data points in the plot based on the RPKM value. If the RPKM value is above the value defined in color delimiter then the data points have green color. Otherwise, the points have red color.

You can use the check boxes provided next to each of the result types to determine if the specimen selected is plotted or not. You can select one or all of the result types for plotting of the graph. Once the specimens are selected, the circular genomic plot is displayed. The outermost circle in the plot is the Cytoband. The supported Cytoband versions are: hg-18 and hg-19.

> **Note:**   To display the cytoband in the Circular Genomic plot we need an entry mapping each DNA Reference Version to the Cytoband in the TRC_LOOKUP_CODE table in the APP_schema.

For each reference version we must insert a row with the following values:

- CODE_TYPE: TRC_REFVERSION_CYTOBAND
- CODE:cytobandHG19 (for HG19), cytobandHG18 (for HG18), or cytobandHG38 (for HG38)
- CODE_NAME: Name of the loaded DNA reference version (for example, V69)

You can hover over the plot to get details such as:

- Microarray Expression: Chromosome, Start Position, End Position, Value and Gene
- Copy Number Variation: Chromosome, Start Position, End Position and Value
- Sequence Variants: Chromosome, start position, end position, and value. The density value is calculated by: Total number of variants for 100kb region / 100000
- Dual Channel Microarray Expression: Chromosome, Start Position, End Position, Value and Gene
- Rna Seq Expression: Chromosome, Start Position, End Position, Value and Gene

Copy Number Variation

The panel displays all the specimens of the result type copy number variation. You can select only one specimen at a time.

The value for copy number variation depends on the CNV result type selected. If the selected specimen contains data from Genome_Wide_SNP_6 array, then the value for CNV is taken from segment mean stored in the database. If the selected specimen has data from complete genomics, then the value is calculated based on the called Ploidy value stored in the database. The value to be plotted is calculated using the following formula for CNV data from complete genomics:

log2(called_ploidy/expected_ploidy)

where,

expected_ploidy is 2 for chr1-22

expected_ploidy is 2 for chrX for females.

expected_ploidy is 1 for chrX outside the pseudo-autosomal region in males

expected_ploidy is 2 for chrX inside pseudo-autosomal region in males

The pseudo autosomal regions on chrX for 'NCBI build 37' as reported by Complete Genomic are 60000 - 2,699,519 and 154,931,043 - 155,260,559.

The pseudo autosomal regions on chrX for 'NCBI build 36' as reported by Complete Genomic are 0 -2,709,519 and 154,584,237 - 154,913,753.

For called_ploidy zero, the log2 should be infinity, in which cases the final value is taken as -2.

Dual Channel Microarray Expression

This panel displays all the specimens of result type Dual Channel Microarray expression. You can select only one specimen at a time.

The color delimiter value is used to render color to the data points in the plot based on the Log2Ratio value. If the Log2Ratio value is above the value defined in color delimiter, then the data points have green color. Otherwise, the points have red color.

Here is how the graph looks.

Figure 4–3   How genomic data is visualized in the Circular Genomic Viewer

5
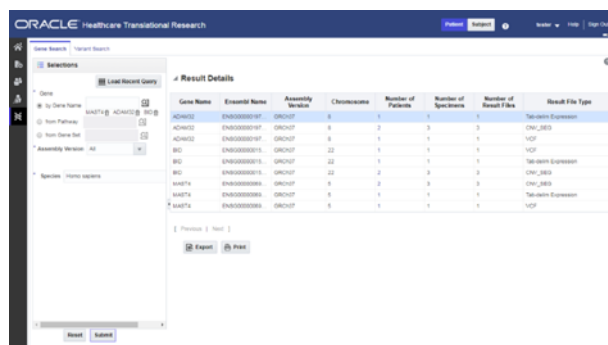
Search genomic results with the Genomic Query

In this chapter you will find out how to:

- Search for genes
- Search for variants

Search for genes

Figure 5–1   How to search for genes

1. On the Home page, on the left, click the **Genomic Query** icon ().

   Make sure that the **Gene Search** tab is open.

2. Either load a recently used query by clicking **Load Recent Query** or search for the gene by gene name, pathway and gene set.

3. Specify the **Assembly Version**, **Species** and **DNA Reference version**.

4. Click **Submit**.

   The results are displayed on the right. You can see the following:

   - **Result Summary**: displays how many genes were searched for, how many have results found and how many do not have any results present in OHO for the criteria selected, what types of results are found.

- **Result Details**: tabular list of metadata for each gene that is found to have genomic results in OHO.

To print the results, click **Print**.

To export the resulting table to a spreadsheet, click **Export**.

Click the **Number of Patients** column to navigate to different screens such as:

- View Record
- Circular Genomic Viewer
- Cohort List
- Cohort Timelines
- Cohort Reports
- Genomic Data Export

> **Note:** If the pathway, gene set or genes are selected from the ad-hoc genes selection, the result details table is paginated based on gene selection. Each page displays the result for a 25 genes selection.

Search for variants

You can specify one or more variant identifiers to be searched across results in Oracle Healthcare Omics (OHO).

1. In the upper left corner, click the **Variant Search** tab.

2. Either load a recently used query by clicking **Load Recent Query** or search for the variant by:

    a. Variant ID

    b. Variant Attributes

    - **Variant status**: select one or all options (**All**, **Novel** and **Known**)

    - **Variant type**: select one or all options (**all**, **complex**, **indel**, **substitution**, **insertion**, **deletion**).

3. Enter the **Species** and **Assembly Version**.

4. Click **Submit**.

    The results are displayed on the right. You can see the following:

    - **Result Summary**: displays how many variants were searched for, how many have results found and, if applicable, how many do not have any results present in Oracle Healthcare Omics for the selected criteria, what types of result files are found for the variants, if any.

    - **Result Details**: displays a tabular list of metadata for each variant that is found to have genomic results in Oracle Healthcare Omics. The columns list includes options such as

Perform a search by Variant ID

1. Either use a recent query by clicking **Load Recent Query** or search for a variant **by Variant ID**.

> **Tip:** You can specify Genomic Position by typing in chr#:from-to location.

```
Example: to find variants in base pair region 1-200 on chromosome
7, you would need to type in the chr7:1-200. If you want to search
for any variants on chromosome 7, you can enter chr7.
```

Additional options that can be specified include known or novel status on the variant, variant type such as insertion, deletion, substitutions, indel, complex. The strand is the directionality of the variant and it can be either plus (+) or minus (-).

2. Select the **Species**.

3. If in **Subject** context, enter a **Study** or search for one.

4. Select the **Assembly Version**.

5. Click **Submit**.

Perform a search by Variant Attributes

Figure 5–2   Search for variants by attributes
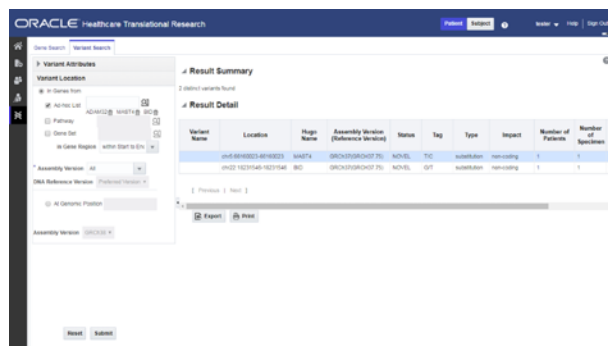
Figure 5–3   Select a variant location
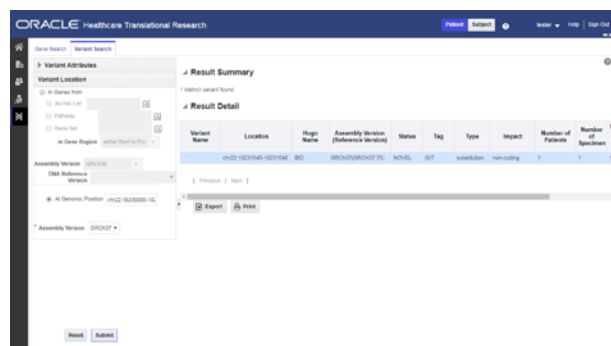
1. Either use a recent query by clicking **Load Recent Query** or search for the variant by **Variant attributes**.

2. Select one or all options for a **Variant status**.

3. Select one or all options for a **Variant type**.

4. Select one or all options for a **Strand**.

5. If in **Subject** context, enter a **Study** or search for one.

6. Select the **Species**.

7. Expand the **Variant location** section and choose one of these options:

   - **in Genes from**: Ad-hoc List, Pathway or Gene Set. Also specify the **Gene Region** and the **DNA Reference Version**.

   - **At Genomic Position**

     Note: For both variant location options you need to specify the **Assembly Version**.

     ---

8. Click **Submit**.

6

Manage queries, gene sets and jobs in My Workspace

In this chapter you will learn how to:

- What can I do in My Workspace?

- Manage queries

- Manage cohort lists

- Manage gene sets

- Administrate user groups

- Schedule and monitor genomic data export jobs

- Switch between Patient and Subject context

> **Note:** If you encounter issues resulting in Internet Explorer 11 incompatibility, see the *Oracle® Healthcare Translational Research Installation Guide* for upgrade instructions.

What can I do in My Workspace?

Figure 6–1 OHTR landing page and your workspace

The My Workspace page is the landing page for OHTR. Depending on the license purchased, you may have access to most or all of the sections on this web page. With a standalone OHTR license, you can only view the Cohort Queries and Cohort Lists details.
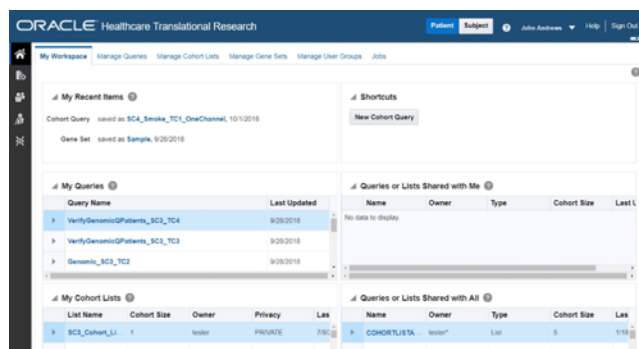
Besides managing queries, cohort lists, gene sets, user groups, and jobs, in here you can also get a constant overview of your work in the product. Your workspace features lists of your most recent items, shortcuts, queries and cohort lists.

To understand what you can see in every list, check the details below.

My recent items

This lists the most recent saved cohort query, cohort list, and gene set, which provides the quickest access to your recent work.

My queries

This lists the most recent saved patient cohort queries, sorted by the last updated date. Clicking the selected query will load it in the Cohort Query window.

Queries or lists shared with me

This lists all queries and lists that are shared with you. Clicking the selected query will load it in the Cohort Query window. Clicking the selected list will load it in the Cohort Viewer/Cohort List window.

My cohort lists

This lists the most recently saved cohort lists sorted by the last updated date. Clicking the selected list will load it in the Cohort Viewer/Cohort List window.

Queries or lists shared with all

This lists all public queries and lists. Clicking the selected query will load it in the Cohort Query window. Clicking the selected list will load it in the Cohort Viewer/Cohort List window.

My gene sets

The My Gene Sets section lists the most recently updated gene sets. Clicking the selected gene sets will load it in the Manage Gene Sets/Manage window.

Manage queries

What can I do here? In the **Manage Queries** tab, you can:

- Search through saved queries.
- View a list of queries, either created by or shared with you.
- Modify queries.
- Set or modify restrictions to available queries.

For information on modifying criteria for queries, inclusions and exclusions, see Chapter 2, "Modify criteria for queries using the Cohort Query".

Search for queries

1. On the **Home page**, click the **Manage Queries** tab.

2. In the **Search My Queries** section, select a condition and criteria for your search.

Example 6–1   Select the conditions and enter the criteria to search for queries

```
Query contains "cohort"
Privacy equals "shared"
Creation Date "On or Before" 21.05.2016
Last Update equals 23.07.2016
```

3. Click **Search**.

   To clear the search and enter new criteria, click **Reset**.

   To view a query, select it from the table and click **Load**.

Modify queries

To change a query's name or description:

1. Select it from the **My Queries** list.

2. Click **Edit** and modify as needed.

3. Click **Submit**.

To change the privacy status of a query:

1. Select it from the **My Queries** list.

2. Change the access:

   ■ Click **Private** to prevent anyone else from seeing or using the query.

   ■ Click **Shared with All** to allow anyone to use the query.

   ■ Click **Shared** to share a private query with a user or a group of users.

      If you chose **Shared**, follow the steps below.

      a. Click **Add**.

      b. Select a condition and enter part of a user name or a group name.

      c. Click **Search**.

      d. Select the user name or user group you want to include and click **Add**.

      e. Click **Done**.

      To share the query with multiple user groups, repeat the steps above before clicking **Done**.

Manage cohort lists

What can I do here? In the **Manage Cohort Lists** tab you can:

■ Search through available saved cohort lists.

■ View a list of cohorts, either created by or shared with you.

■ Modify cohort lists.

■ Set or modify restrictions to available queries.

Search and view your cohort lists

You can only search for cohort lists that you have access to.

1. On the **Home page**, click the **Manage Cohort Lists** tab.

2. In the **Search my Cohort Lists** section, select the conditions and enter the criteria to search for your saved cohort list.

Example 6–2   How to enter criteria to search for your saved cohort list

```
List Name equals "cohort"
Privacy equals Private
Creation Date "On or Before" 21.05.2016
Last Update equals 23.07.2016
```

3. Click **Search**.

   To clear the search and enter new search criteria, click **Reset**.

   To view a cohort list, select it from the **My Cohort Lists** and click **Load**.

Modify cohort lists

You can modify your cohort lists in multiple ways, from editing their names to changing their privacy status.

To change a cohort list's name and description:

1. Select it from the **My Cohort Lists** section.

2. Click **Edit**.

3. Click **Submit**.

To change the privacy status of a cohort list:

1. Select it from the **My Cohort Lists**.

2. Change the access:

- Click **Private** to prevent anyone else from accessing the list.

- Click **Shared with All** to allow everyone to access the list.

- Click **Shared** to share a private list.

   If you chose **Shared**, follow the steps below.

   a. Click **Add**.

   b. Select a condition and enter part of a user name or a group name.

   c. Click **Search**.

   d. Select the user name or user group you want to include and click **Add**.

   e. Click **Done**.

   To share the cohort list with multiple users or user groups, repeat the steps above before clicking **Done**.

Manage gene sets

What are gene sets and what can I do in here? Gene set refers to groups or lists of genes. This set can consist of a couple of genes or hundreds of genes. This allows you

to group genes into convenient *collections* for reuse. You can group a few genes for quick search retrieval or for use in a cohort query. In this tab, you can create, edit, and manage gene sets.

A few things to keep in mind about gene sets in OHTR:

- There are no restrictions on genes that can be included in a gene set. You can mix genes from multiple species or assembly versions.

- The same gene can be part of many different gene sets.

- Gene sets are private and cannot be shared among users.

- Gene set names are not case-sensitive.

- When genes are compared using the **Contains** or **Starts With** option, the limit is 512 characters.

Create a new gene set

1. Along the top, click **Manage Gene Sets** and go to the **Create New or Edit** section.

2. Specify the genes you want to include in the gene set. This can be done in three ways:

   - **Type in Gene Names** - Search for genes using their Ensembl or HUGO names. You can enter multiple gene names separated by space, comma or semicolon.

   - **Add from existing gene set** - Search for genes based on an existing gene set.

   - **Upload from a file** - Select a text file (CSV, TSV, TXT) from your computer where the genes are delimited by comma, space or tab. Click **Choose File** to browse for the file.

   **Note:** The file size limit is less than 5MB when genes are matched using Equals.

3. Select a value for **Species**.

4. Select an **Assembly Version**.

5. Click **Submit**. The matching genes appear in the **Available Genes** pane.

6. Use the arrow (>) to move available genes to the **Final Gene Set** panel.

   Use the double arrow (») to move all available genes to the **Final Gene Set** panel.

   To remove genes from the **Final Gene Set** panel, use the **Remove** or **Remove All** buttons.

7. In the **Save Gene Set** section, enter the name and description of the gene set.

8. Click **Submit**.

Edit a gene set

To edit a gene set:

1. Along the top, click the **Manage Gene Sets** tab, and go to the **Create New or Edit** section.

2. Search for the gene set you want to edit using the **Add from existing Gene Set** option.

3. Select a value for **Species**.

4. Select an **Assembly Version**.

5. Click **Submit**. The gene set appears in the **Available Genes** pane.

6. Use the arrow(>) to move available genes to the **Final Gene Set** panel.

   Use the double arrow (») to move all available genes to the **Final Gene Set** panel.

   To remove genes, use the **Remove** or **Remove All** buttons.

7. Click **Submit**.

> **Note:** When you search genes using the **Contains** or **Starts With** option, the limit is **512** characters

---

Manage a gene set

To manage a gene set:

1. Click the **Manage Gene Sets** tab and go to the **Manage** section.

2. Enter the gene set you want to modify.

3. Click **Submit**.

   To print the data, click **Print** on the top right.

   To export the data in an.xls file, click **Export**.

   To remove a gene set, click **Delete Gene Set** at the bottom.

> **Note:** When you delete a gene set, its record is still available in the database but it can only be restored by a system administrator.

Administrate user groups

Administrating user groups can be done directly from the OHTR UI.

To simplify sharing, create user groups and assign users to them. Instead of sharing queries or lists with each user individually, set up a list of users to share multiple items to multiple users at the same time.

> **Note:** Only the user that is the owner of a list can modify its members. However, any user can share any given lists with other groups and see members in other lists.

Create a user group

1. Navigate to **Manage User Groups**, **User groups**.

2. Click **Create**.

3. In the **Create User Group** window, enter a name and description.

4. Click **Submit**.

Assign users to a group

Immediately after you create a group

1. Navigate to **Manage User Groups**, **Group has members**.

2. Click **Add**.

3. In the **Add Users** window, enter the user names or search for them.

4. Select the names you want to add.

5. Click **Add**.

6. Click **Done**.

Search for a group you created earlier

1. Navigate to **Manage User Groups**, **Search User Groups**.

2. Enter the details of the group you want to find.

3. Click **Search**.

4. Navigate to **Manage User Groups**, **Group has members**.

5. Click **Add**.

6. Enter the user name you want to add and select them.

7. Click **Add**.

8. Click **Done**.

Remove users from a group

1. Navigate to **Manage User Groups**, **Search User Groups**.

2. Enter the details of the group you want to find.

3. Click **Search**. The user group and its details are displayed.

4. Navigate to **Manage User Groups**, **Group has members**.

5. Select the user you want to remove.

6. Click **Remove**.

Manage user groups

You can edit, duplicate or delete a user group using the buttons at the top of the **User Groups** section.

1. Go to **Manage User Groups**, **Search User Groups**.

2. Enter the details of the group you want to find.

3. Click **Search**.

4. Follow one of these steps:

   ■ Click **Edit** to change the name and description of the user group.

   ■ Click **Duplicate** to create a copy of a user group.

5. Click **Submit**.

Schedule and monitor genomic data export jobs

Currently, jobs are scheduled only for **Genomic Data Export**. To schedule a job, navigate to the **Genomic Data Export** screen in the **Cohort Viewer** tab.

> **Note:**   Any job executed prior to applying OHTR 3.1 will have its status changed back to '**Scheduled**'.

1. Select the patient IDs source from one of these options:
   - **active query**
   - **library query**
   - **list**
   - **ad-hoc**
   - **Omics query**
2. Select the **DNA reference version**.
3. Select the location for the gene:
   - **Ad-hoc List**
   - **Pathway**
   - **Gene Set**
4. Select one of the following file formats to export:
   - **Mutation-VCF**
   - **Copy Number Variation-SEG**
   - **Microarray Expression-RES**
   - **Microarray Expression Dual Channel-GCT**
5. Select the **Schedule** export option.
6. Click **Submit**.

To view job information and progress, go to the Home page, and click the **Jobs** tab, along the top. All the jobs created by the OHTR UI are displayed. All users who created jobs can see job lists.

Manage jobs

To search for a job:

1. On the Home page, along the top, click **Jobs**.
2. Scroll down, after the table, and either enter part of the job's name or click the search icon ( ).
3. In the **Search User Job** window, select the appropriate conditions and enter part of a **User Job Run ID** or a **User Job Name**.
4. Click **Search**.
5. Select the job you want to include in the list.
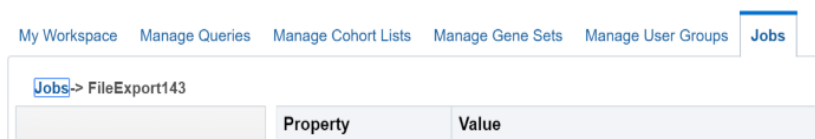6. Click **Submit**.

To run that job, on the **Jobs** tab, click **Submit** again.

To view details of a particular job:

1. On the **Jobs** tab, in the first table column, click a job's name to get to the details of the job.

    To return to the **Jobs** tab, click the **Jobs** breadcrumb link.

Figure 6–2   Location of the Jobs link

To get back to the initial list of jobs, click **Reset**.

View job inputs

■   To view details on the parameters provided when scheduling a job, in the **Details** page, click the **Inputs** button on the left.

■   To load the job details in the **Genomic Data Export** page, click the **Load input parameter values into Genomic Data Export page** link.

> **Note:**   You can only load these input parameters, when the job context matches the application's context.

View and download job outputs

This section shows the result files of a job. The following files are generated:

■   **Error log file** contains the log file of the application if any exception or error occurred while running the job.

■   **User log file** contains the log file, which shows the error occurred while generating the export files.

■   **Administrator log file** is for the DB administrator.

To download any of these files, click the file link under the **Files** column.

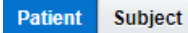**You can download error and user log files, but cannot download the admin log file**.
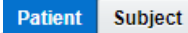
> **Note:**   When you try to export a file that does not have any patients or subjects linked to the specimen, an empty file is generated in **Schedule** mode.

Switch between Patient and Subject context

OHTR can be run in either Patient or Study Subject context. In Patient context, all queries are directed at patient tables in the Cohort Data Model (CDM) schema, while in the Subject context, all queries are routed to the subject tables in the Cohort Data Model (CDM).

- To change context, click either one of the buttons (![Patient Subject]), in the upper-right corner of the screen. By default, the context is set to **Patient**.

- The new context is available until the next time you change it.

- Switching context clears the data loaded in the Cohort Viewer tabs, but not the criteria of an active query in the Cohort Query screen. Only applicable criteria for the selected context is displayed.

Example 6–3

```
Encounter criteria is not supported in the Subject context and will be disabled.
When you switch back to the Patient context, it is enabled again.
```

- For a subject or a patient to be considered deleted and not visible for selection, a given subject or patient's delete flag in CDM should be set to *Y*.