ForenSeq™ Universal Analysis Software Guide

VEROGEN PROPRIETARY
Document # VD2018007 Rev. A
June 2018

Revision History

Document #	Date	Description of Change
Document # VD2018007 A	June 2018	Updated document number and other document numbers listed in this guide to Verogen document numbers. Updated technical support information with Verogen contacts. Updated the location of SDS information on the Verogen website. Updated ForenSeq Universal Analysis Software descriptions to software v1.3. Added the following information: • Sample History feature for user action, system event tracking, sample-level reports, and phenotype estimation, as well as enabling Sample History. • Project-level flanking region reports for STRs and iSNPs, and enabling automatic generation of flanking region reports. • CODIS reports, including sample-level reports, project-level reports, and report defaults. • Data management, including information about exporting and importing data. • Information for loci detected using ForenSeq DNA Primer Mix A or DNA Primer Mix B.

Document #	Date	Description of Change
Document # 15053876 v01	August 2016	Updated screen shots throughout the guide to reflect changes in the user interface for ForenSeq Universal Analysis Software v1.2. In the Supporting Information chapter, added directions for setting up an external data repository. Added a section on troubleshooting analysis errors. In the Run Metrics Tab section, revised the description of light orange in the sample representation tab graph to indicate that loci with low coverage might not have enough signal for all alleles to be distinguished from noise. Updated descriptions of icons for Allele Count and Interpretation Threshold. Updated descriptions of what causes interpretation threshold, allele count, and mixture detection quality indicators in STR genotype calling. Updated descriptions of what causes the interpretation threshold quality indicator in SNP genotype calling. Noted that the low coverage indicator can be seen in results from version 1.1 or earlier of the ForenSeq Universal Analysis Software. Updated the following thresholds for aSNPs in the Supporting Information section: • For rs1572018, changed the analytical threshold from > 5 to > 1.5, and the interpretation threshold from > 15 to > 4.5. • For rs735480, changed the analytical threshold from > 3.3 to > 1.5, and the interpretation threshold from > 10 to > 4.5. • For rs3916235, changed the interpretation threshold from > 10 to > 4.5. In the Analysis Procedures section for STR and SNP genotype calling, added quality considerations for the analytical threshold, and when signal is not detected at the locus. Added the icons Analytical Threshold and Not Detected. In the Reports section, added information on the new phenotype estimation report that is generated from the Project page. In the Reports section, added descriptions of the new coverage worksheets in the Project Level Report section. Updated descriptions of icons for Allele Count and Interpretation Threshold. In the section STR Genotype Calling, added a table with additional details regarding how gender is determined.

Document #	Date	Description of Change
Part # 15053876 Rev. C	June 2015	Updated software descriptions to ForenSeq Universal Analysis Software v1.1. • Updated screen shots throughout the guide to reflect changes in the user interface. • Added a section for modifying and updating the gender of a sample. • Added the section Define Loci Used in Population Studies for selecting loci for population calculations statistics, and noted that the software queries both population groups and loci defined by the administrator at the time of calculation. • Added information on selecting the 2p and 2p-p^2 population statistics calculation rule in the Defining the Statistics Calculation Method section. • Added instructions on applying the 2p rule to homozygous loci. • Noted in the STR Genotype Calling section that negative amplification control samples and reagent blank samples are not assessed for gender and by default are "inconclusive". • Added details on the system dashboard for the run progress bar, quick access to run metrics, analysis icon, and About information. • Added information on generating sample genotype reports from the Project page for multiple samples. • Updated the description of the Analysis page display to include the analysis version. • .Added information on how to indicate review status in the Viewing Run Quality Information section. • Added information on opening the Sample Details page from sample representation, positive amplification control, and negative amplification control tabs. • Noted that isometric alleles are now highlighted with a gray box. • Updated the name of the Locus Panel page to the Define Content page. • Updated the name of the Locus Thresholds page to the Analysis Values page.

Document #	Date	Description of Change
Part # 15053876 Rev. B	February 2015	Updated screenshots throughout the guide to reflect changes in the user interface. Added a System Settings chapter that details information on notifications, user management, changing locus thresholds, defining locus content, and population group settings. Added a section on Hardy-Weinberg expectations and linkage equilibrium. Removed the Analysis Settings section and moved threshold settings information to the System Settings chapter of the guide. Added a Creating a New Analysis section. Updated intensity and length graphs information to reflect a change in how the software splits and combines the display of STRs and SNPs on the graph. Updated information on the typed alleles indicator, which changed from true and false to a toggle switch, and can also be toggled on and off by clicking on the bar of the bar chart representing an allele for an STR, or pie chart for a SNP allele. Updated random match probability information to distinguish between the generation of population statistics and the random match probability calculation method. Updated report generation information to include Sample Summary and Sample Detail reports. Divided the Additional Analyses chapter to create a chapter for population statistics, sample comparison, and phenotype and bio-geographical ancestry estimation. Added a troubleshooting section for population group file upload. Changed the Settings section to Locus Level Settings. Updated the explanation of loci typed in the General Locus Information section to include selected locus content. Updated the population statistics calculations information to include population group selection.
Part # 15053876 Rev. A	December 2014	Initial Release

Table of Contents

Getting Started

Chapter 1

Introduction

ForenSeq™ Universal Analysis Software is a complete DNA-to-data forensic software solution at the center of the MiSeq FGx™ Forensic Genomics System. The software performs analysis of sequenced genetic sample information for human identification and works in combination with the ForenSeq DNA Signature Prep Kit and the MiSeq FGx instrument.

The ForenSeq Universal Analysis Software comes pre-installed as a dedicated server with a user interface specific for forensic genomics that enables run setup, sample management, analysis, and report generation. ForenSeq Universal Analysis Software provides population statistics and automated sample comparison, as well as an optional feature for estimating biogeographical ancestry, hair color, and eye color.

NOTE

Some laboratories may choose to or need to perform internal validation studies in order to develop protocols and interpretation guidelines for casework and DNA databasing using the ForenSeq Universal Analysis Software. It is possible that some features and recommendations detailed in the guide are not part of your analysis requirements, analysis routine, or necessary for high-quality analyses.

Software Functions

The following is a workflow outline of functions administered by the software.

1. Create an account with a user name and password for access to the ForenSeq Universal Analysis Software and the MiSeq FGx instrument.
2. Input sample index combinations, sample types, and ForenSeq DNA primer mix selection for sequencing on the MiSeq FGx of amplicons generated with the ForenSeq DNA Signature Prep Kit. Enter information manually or import a text (*.txt) file.
3. The MiSeq FGx uses two software applications in tandem to produce images of clusters on the flow cell, perform image analysis, and call bases.
    a. During the run, MiSeq FGx Control Software captures images of clusters on the flow cell for image analysis, as well as operates the flow cell stage, gives commands to dispense reagents, and changes temperatures of the flow cell.
    b. Real-Time Analysis (RTA) software performs image analysis, base calling, and assigns a quality score to each base for each cycle as the run progresses. The completion of analysis by RTA and transfer of files initiates analysis on the ForenSeq Universal Analysis Software.
4. Using data from RTA, the ForenSeq Universal Analysis Software aligns reads to make allele and genotype calls. If interpretation or troubleshooting for a particular allele might be considered, the ForenSeq Universal Analysis Software provides quality indicators.
5. Review run analyses and generate reports on the ForenSeq Universal Analysis Software.

ForenSeq Universal Analysis Software Concepts

The following concepts and terms apply to the ForenSeq Universal Analysis Software.

Table 1 Concepts and Terms

Concept	Description
Amplicon	The result of PCR amplification of a targeted region of interest from input gDNA template.
Analysis Version	Analysis versions are a traceable, flexible way to reanalyze samples with new analysis settings. See *Creating a New Analysis* on page 33.
Analytical Threshold	This value (%) represents the lower limit of detection; a matter of internal laboratory policy-making.
Clusters	A clonal grouping of template DNA bound to the surface of a flow cell. Each cluster is seeded by a single, template DNA strand and is clonally amplified through bridge amplification until the cluster has roughly 1000 copies. Each cluster on the flow cell produces a single sequencing read. For example, 10,000 clusters on the flow cell produce 10,000 single reads.
ForenSeq DNA Primer Mix	A set of target-specific PCR primers. Primers are tagged oligos targeting forensically relevant DNA sequences. In the ForenSeq DNA Signature Prep Kit, ForenSeq DNA Primer Mix A (DPMA) contains primers for polymorphic STRs, identity SNPs (iSNPs), and Amelogenin. ForenSeq DNA Primer Mix B contains primers for those same polymorphic STRs, iSNPs, and Amelogenin, as well as phenotypic SNPs (pSNPs), and biogeographical ancestry SNPs (aSNPs).
Flow Cell Type	The layout of a flow cell used for a MiSeq FGx sequencing run. • The Standard MiSeq FGx flow cell type allows for the full complement of samples to be sequenced in a single sequencing run. • The Micro flow cell type supports sequencing pooled sample multiplexes of reduced numbers of samples. Both flow cell types use the same sequencing reactions and have equivalent date quality.
Genotype	Alleles at a locus; typically heterozygous, homozygous, or hemizygous.
Human Sequencing Control	Pool of human DNA STR loci used as a positive sequencing control in the ForenSeq DNA Signature Prep Kit. Human Sequencing Control (HSC) helps MiSeq FGx run completion and highlights possible sequencing issues.
Index	DNA tags that are attached to DNA sequences targeted by the ForenSeq DNA Signature Prep Kit, enabling multiple samples to be pooled together for sequencing and demultiplexed post-run. See *Demultiplexing* on page 126.
Interlocus balance	The balance of read counts between loci in a sample. This balance is measured as the % CV of the read counts across all of the loci in the multiplex. See *General Locus Information* on page 35.
Intralocus balance	The balance of read counts between typed alleles at a heterozygous locus. This balance is measured as the intensity of the minimum intensity typed allele divided by the intensity of the maximum intensity typed allele. If both are identical in intensity, the intralocus balance is 100%. See *Locus Detail Box Indicators* on page 36.
Interpretation Threshold	This value (%) may be utilized as a conservative allele calling threshold; a matter of internal laboratory policy-making.

Concept	Description
Isometric Allele	Two or more alleles at the same locus with an identical allele name (i.e., repeat length), but are distinguishable by the allele sequence.
Project	A project is a collection of analyzed results in the ForenSeq Universal Analysis Software and can be used to logically group and organize data. If desired, each sample in a run can be independently assigned to a different project, so that a run can have samples from more than one project. See *Viewing the Project Page* on page 21.
Read	A sequence read refers to the data string of A, T, C, and G bases corresponding to a ForenSeq amplicon from a sample DNA. Millions of reads can be generated in a sequencing run from multiple loci and multiple samples simultaneously.
Run Version	Run versions are a traceable, flexible way to recombine information to execute a new ForenSeq software analysis. See *Viewing Run Details* on page 14 and *Creating a New Run Version* on page 17.
Sample History	The Sample History feature enables visibility of certain user actions and system-initiated events that are logged by the software. A list of events is displayed in analysis, sample, and locus-level activity dialogs within the software. These events are also recorded in the sample-level genotype reports on the Sample History tab.
Sample Information	Information needed to set up and analyze a MiSeq FGx sequencing run, such as a list of samples and their index combinations, sample types, and ForenSeq DNA primer mix.
Sample Type	The type or function of a sample. Sample types include sample, positive amplification control, negative amplification control, and reagent blank.
SNP	Single Nucleotide Polymorphism. Variation of a single nucleotide base within a DNA sequence, relative to a known DNA reference sequence.
STR	Short Tandem Repeat. A DNA sequence containing a variable number (typically ≤ 50) of tandemly repeated short (2–6 bp) sequence motifs, such as (GATA)n.
Stutter	Polymerase slippage that can occur during PCR amplification of repetitive DNA sequences, the library preparation process, cluster generation, or when performing sequencing, and can create DNA amplification products that are less than or greater than the size of a parent allele.
System Events	System-initiated events logged as part of the Sample History feature at the analysis and sample-level activity (e.g., run completed, analysis initiated, population statistics calculation completed)
User Actions	User-initiated events logged as part of the Sample History feature at the analysis, sample, or locus-level activity (e.g., run created, project report issued, locus edited or commented). User Actions is also the name of the indicator icon that appears in the software to facilitate recognition of user-initiated events.

Viewing the ForenSeq Universal Analysis Software

The ForenSeq Universal Analysis Software interface is viewed through a web browser. To view the interface, open a web browser on a computer with access to the network used by the ForenSeq Universal Analysis Software and the MiSeq FGx instrument. Set screen resolution to a minimum of 992 pixels for optimal display.

NOTE
The application is optimized for use on the Google Chrome Browser.

To access the ForenSeq Universal Analysis Software, enter the server address in the web browser address bar. If the server is not listed in the Domain Name System (DNS), enter the server IP address of the network.

NOTE
No internet connection or internet access is required to view the software interface.

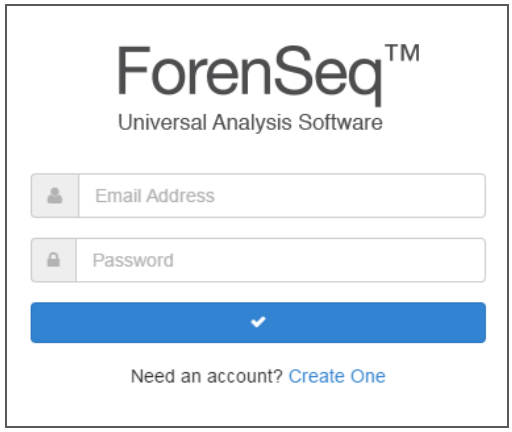
Logging in to the Software

1 Access the network server through a web browser.

2 Enter a user account name and password.
If you need an account, see *Creating an Account*.

3 Click the blue button with the checkmark.

Figure 1 Login Screen



4 To log out, click the drop-down list on your account email address at the top of the page and select **Log Out**.

NOTE
To change your password, see *Changing an Account Password* on page 100.

Creating an Account

1 Access the network server through a web browser.

2 When the ForenSeq Universal Analysis Software page opens, click **Create One** at the bottom of the dialog box.

3 Enter a valid user account name and password.

4 Click the blue button with the checkmark.
After you create an account, a user with administrator-level system access must approve your account before you can access the system. See *Enabling and Disabling Accounts* on page 99.

Viewing the System Dashboard

A successful login opens the system dashboard. At the top of the dashboard, the following information can be accessed:

- Analysis icon to access processing queue information, the last three analyses completed, and the status of the last analysis submitted
- Name of the currently logged-in user with a drop-down list that includes access for Log Out, About information, User Profile settings, and Maintenance page access for administrator-level users.
- Command to create a new run.

The system dashboard shows the following information and functionality organized within the Runs, Projects, and Samples tabs.

Runs tab

- List of runs arranged by run creation date with the most recent run appearing at the top of the list.
- Search tool for runs.
- Run filtering by run state (All, Created, Sequencing, and Completed) and Flow Cell Type (Standard and Micro).
- Quality icon next to run name that displays the run quality metrics. For more information, see *Run Metrics Tab* on page 29.

Figure 2 Runs Tab

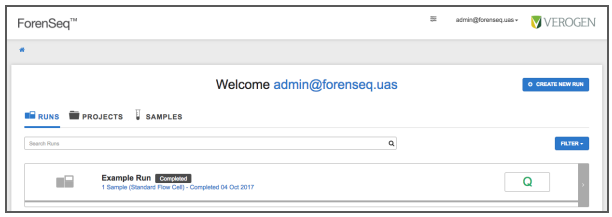
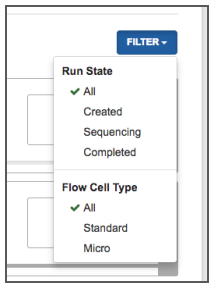


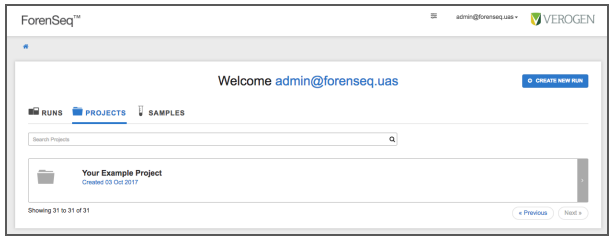
Figure 3 Runs Filter Options



Projects tab

- List of projects arranged by project creation date with the most recent project appearing at the top of the list.
- Search tool for projects.

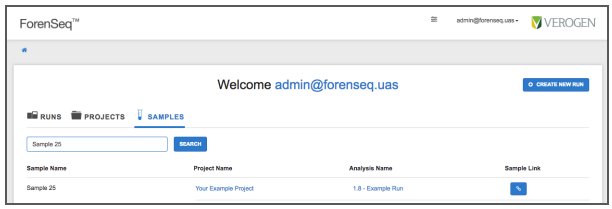
Figure 4 Projects Tab



Samples tab

- Search tool for sample names that provides links to the associated project, analysis, and sample details page for that sample.

Figure 5 Samples Tab



Example Workflow

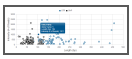
Create an account in the ForenSeq Universal Analysis Software. See *Logging in to the Software* on page 7.

Create a new run and save it. See *Creating a New Run* on page 11.

Log in to the MiSeq FGx instrument with the account information created in the ForenSeq Universal Analysis Software. See the *MiSeq FGx Instrument Reference Guide (document # VD2018006).*

On the MiSeq FGx instrument, select and sequence the run you created and saved in the ForenSeq Universal Analysis Software. See the *MiSeq FGx Instrument Reference Guide (document # VD2018006).*

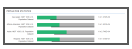
Review run quality information in the ForenSeq Universal Analysis Software. See *Viewing Run Quality Information* on page 22.



Review sample details. See *Viewing Sample Details Page* on page 34.

Optional Steps

Create a new run version or create a new analysis. See *Creating a New Run Version* on page 17 and *Creating a New Analysis* on page 33.



Generate population statistics. See *Population Statistics* on page 45.

Compare samples. See *Sample Comparison* on page 51.

Generate phenotype or biogeographical ancestry information. See *Phenotype and Biogeographical Ancestry Estimation* on page 61.

Generate project-level and sample genotype reports. See *Reports* on page 69.

Creating a New Run

Before performing a sequencing run on the MiSeq FGx instrument, enter information in ForenSeq Universal Analysis Software about the run and the samples in the run.

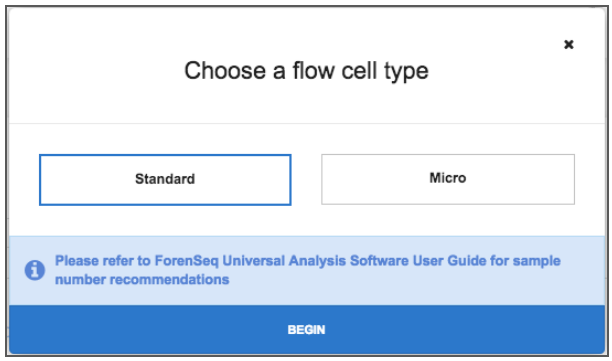
When creating a new run, select either Standard or Micro flow cell type. The decision of which flow cell type to use is based primarily on the total number of samples per run and desired coverage level for each sample. ForenSeq DNA Primer Mix A (DPMA) or ForenSeq DNA Primer Mix B (DMPB) libraries can be run on either flow cell type.

After logging in, the software opens to the system dashboard.

1 Click **Create New Run**.

2 Select the flow cell type to be used, and click **Begin**. See Table 2 for recommendations.

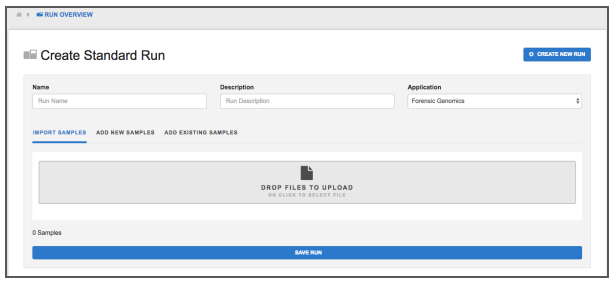
Figure 6 Flow Cell Type



3 Enter the following information for sequencing and analysis.

a Name—Enter a run name.
b [Optional] Description—Enter a run description.
c Application—Select Forensic Genomics.
d Add sample information using one of the methods described in , or a combination of the methods.

Figure 7 New Run Information



Before you save a run, you can edit any of the run details or sample details, or remove a sample. To make changes, click in any field on the Create A Run page. To remove a sample, click the X on the right side of the sample information.

4 Click **Save Run**.

NOTE
The run version that is created for the initial analysis and is sequenced is version 1.0. For example, if the name of the run is Example Run, the run version is 1.0 - Example Run, and appears on the Project page. If you create additional versions of the run, the runs are numbered in sequence. See *Creating a New Run Version* on page 17.

Flow Cell Types

Two types of MiSeq FGx Reagent Kits are available. The standard MiSeq FGx Reagent Kit provides the standard flow cell type and the MiSeq FGx Reagent Micro Kit provides the micro flow cell type.

The following sample multiplexing levels are recommended for each flow cell type.

Table 2 Flow Cell Types

Flow Cell Type	Description
Standard	ForenSeq DNA Primer Mix A: up to 96 samples per MiSeq FGx run. ForenSeq DNA Primer Mix B: up to 32 samples per MiSeq FGx run.
Micro	ForenSeq DNA Primer Mix A: up to 36 samples per MiSeq FGx run. ForenSeq DNA Primer Mix B: up to 12 samples per MiSeq FGx run.

Create a Run

Table 3 Create a Run

Method	Description	Steps
Import Samples	If you created a tab-delimited file (*.txt) with sample sheet information, import the sample sheet. See *Entering Sample Information* on page 13.	• Click the **Import Samples** tab. • Click **Drop Files to Upload or Click to Select File**. • Navigate to the (*.txt) file and click **Open**. • Click the blue **Import Samples** button.
Add New Samples	Add new sample information.	• Click the **Add New Samples** tab. • Enter a **Sample Name**. • Enter a **Project Name**. If the project already exists, use the search field to find the project name. • [Optional] Enter a **Sample Description**. • Select an **i7 Index**. • Select an **i5 Index**. • Select a **Sample Type** from the drop-down list. • Select a **Mix Type** from the drop-down list. • When sample information is complete, click **Add New Sample**.

Method	Description	Steps
Add Existing Samples	Add samples already created.	• Click the **Add Existing Samples** tab. • Enter an existing **Sample Name**. Use the search field to find the sample name. • Enter a **Project Name**. If the project already exists, use the auto complete drop-down list to find the project name. • [Optional] Enter a **Sample Description**. • Select an **i7 Index**. • Select an **i5 Index**. • Select a **Sample Type** from the drop-down list. • Select a **Mix Type** from the drop-down list. • When sample information is complete, click **Add New Sample**.

Entering Sample Information

Sample information can be directly input on the Create a Run page. You can also create a tab delimited *.txt file, and import it into the software on the same page.

To create a *.txt file for a run, make columns of information using these headers in the columns.

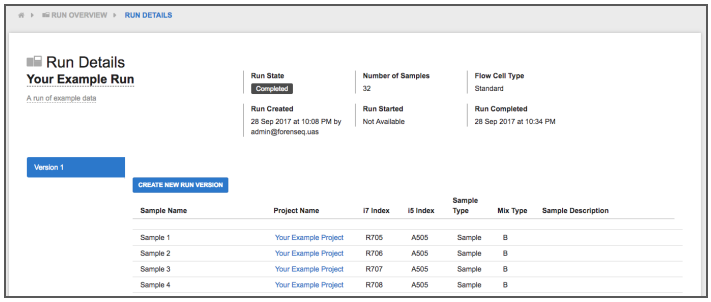
SampleName	Project	i7Index	i5Index	SampleType	SampleDescription	MixType
Sample1	Tst Prj	R701	A501	Sample	Sample 1	A
Sample2	Tst Prj	R701	A502	Positive Amplification Control	PAC 1	B
Sample3	Tst Prj	R701	A503	Negative Amplification Control	NAC 1	A
Sample4	Tst Prj	R701	A504	Reagent Blank	RA 1	B

Viewing Run Details

To view the Run Details page, click the name of the run on the system dashboard. There are several ways to find a run on the system dashboard:

- The system dashboard lists runs when the Runs tab is selected.
- To narrow your search of the run list, click the Filter button to the right of the Search Runs box to filter by Run State (All, Created, Sequencing, or Completed) and Flow Cell Type (All, Standard, or Micro).
- On the system dashboard, use page navigation at the bottom of the page, under the runs column.
- Enter the run name in the Search Runs box.

Figure 8 Run Details



The Run Details page contains the following run information:

- Name of the run.
- Description of the run (if an optional description was included).
- State of the run. See *Viewing Run Status* on page 15
- Number of samples in the run. If more than one run version, the number of samples in the most recent version of the run.
- Flow Cell Type (Standard or Micro). See Table 2 on page 12.
- Time and date the run is created in the ForenSeq Universal Analysis Software.
- User name that created the run in the ForenSeq Universal Analysis Software.
- Time and date the sequencing started on the MiSeq FGx.
- Time and date the sequencing completed on the MiSeq FGx.
- Run version.
- A list of samples in that version of the run.

Figure 9 Run Information on Run Details

Viewing Run Status

After you create and save a run, a box next to the name of the run indicates the current run status.

Table 4 Run Status Descriptions

Run Status	Description
Created	Run setup information is saved in the ForenSeq Universal Analysis Software.
Sequencing	Sequencing of the run on the MiSeq FGx instrument is in progress.
Sequencing Paused	Sequencing of the run is on pause on the MiSeq FGx instrument. To pause a run, see the *MiSeq FGx Instrument Reference Guide (document # VD2018006)*.
Completed	Sequencing of the run on the MiSeq FGx instrument is complete.
Completed-Error	The run did not finish sequencing on the MiSeq FGx instrument because it was manually stopped or because of an error.

Changing the Name or Description of a Run

1. Navigate to the Run Details page.
2. Click the name or description of the run to start the edit mode.
3. Enter the new name in the name field, or the new description in the description field.
4. Click the blue checkmark box to save the change, or click the X box if you do not want the save the change.

Creating a New Run Version

The Create New Run Version feature is a traceable, flexible way to create additional analyses of your samples. In a new run version, you can use existing samples or add new samples. You can update run index combinations, sample types, project assignment, and ForenSeq DNA primer mix types. When a new run version is set up, you can execute an analysis with the new information. Create A New Run Version is available after a run completes sequencing.

1 Open the Run Details page. See *Viewing Run Details* on page 14.

2 Click the **Create New Run Version** button.

3 Select **Edit Samples** to revise sample information. Add samples, remove samples, or modify existing sample information to describe the samples in a run.

4 Click **Save Changes**.

5 To execute an analysis on the new version, select **Execute Analysis**.

NOTE
After you execute analysis on a run version, that run version cannot be deleted.

6 To delete the run version before you execute analysis, select **Delete Version**.

NOTE
Version 1.0 of a run is the run version that is created for the initial analysis and is sequenced. If you create additional versions of the run, the runs are numbered in sequence. For example, if you create a new version of the run 1.0 - Example Run 1.0, it becomes 2.0 - Example Run after sequencing. If you create a third version to sequence of the same run, it becomes 3.0 - Example Run, and so forth.

Sample and Run Results

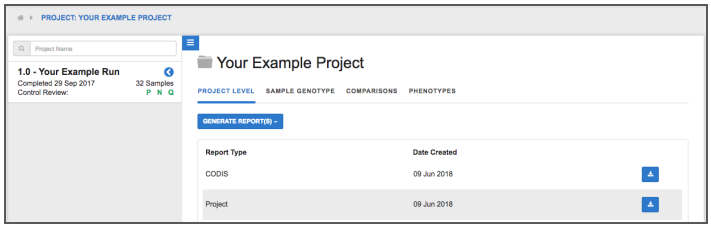
Chapter 2

Introduction

Information and results for a project are opened from the Project page. Quality metrics, sample representation, as well as positive and negative control results are accessible for each analysis. Analysis settings can be viewed, modified, or saved using a template, and renamed. Modified settings can then be used to reanalyze samples in a project. The software also features charts and tables of STR and SNP locus results.

Viewing the Project Page

Figure 10 Projects Page



When you add samples to a run, you associate every sample to a project. Because each sample can be independently assigned to a project, a run can have samples that are included in more than one project. Information and results for a project appear in a series of pages in the software interface from the Project page. To view project-level results in a report, see *Project-Level Genotype Report* on page 71.

Navigating to the Project Page

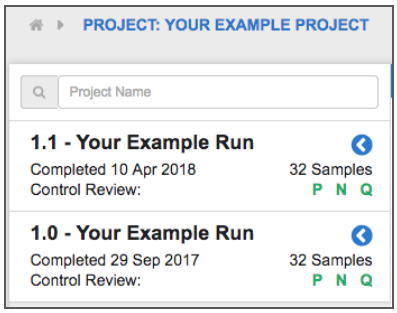
There are several ways to find a Project page.

- Access a Project page from the Run Details page of your completed sequencing run.
- Access the Project page from the system dashboard, which lists projects when the Projects tab is selected.
- On the system dashboard, use page navigation at the bottom of the page, under the projects column to find the relevant project. Click the name of a project to access the Project page.
- Enter the project name in the Search Projects field. Click the name of a project to access the Project page.
- On the Project page, use the Project Name search field to navigate to other projects you want to view. Click the name of a project to access the Project page.

Viewing Run Quality Information

The analyses in a project along with the associated version number are listed on the left panel of the Project page. Click the name of an analysis to view run quality metrics and sample metrics.

Figure 11 Analysis Left Panel



The following run information shows in the left panel below the analysis name.

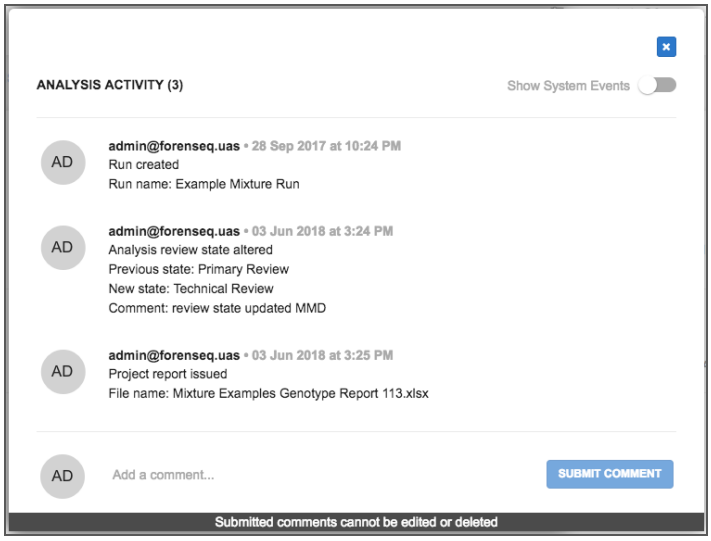
Analysis Info	Description
Completed Date	Date the analysis completed or Processing, if the analysis is not complete.
Samples	Number of samples in the analysis for the project.
Control Review	Snapshot of the overall quality indication for the following: • Positive control samples (P) • Negative control samples (N) • Quality metrics for the run (Q) Green indicates that all the metrics were within an acceptable range. Orange indicates that at least 1 metric was not within the predefined range for acceptability and requires further investigation. Gray indicates that there are no metrics assessed with the analysis.

After clicking an analysis on the left panel of the Project page, an Analysis page opens. At the top of the Analysis page, the analysis name and version are presented as a header. Below the analysis name, the analysis history, review state, analysis settings, version of the analysis module, status (state) of the analysis, and flow cell type are displayed.

Sample History is a feature that can be enabled by a ForenSeq Universal Analysis Software administrator. When visibility of the feature is enabled, the Analysis Activity dialog lists system-initiated events (e.g., run completed) and user actions (e.g., project report issued). To view the Analysis Activity, select the user actions icon on the Analysis page.

Figure 12 Analysis Page

Figure 13 Analysis Activity



The user actions icon provides an indicator for the number of analysis-level user actions that have been logged, such as comments, project reports, and creation of new analysis versions. Selecting the icon displays the user responsible, time and date, and details for each event. Additional system events can be displayed by adjusting the toggle in the upper-right corner of the dialog. A comment can be added to the analysis by typing it in the box at the bottom of the screen and selecting **Submit Comment**.

NOTE
After a comment is submitted, it is permanent and cannot be modified or deleted.

Reviewers can indicate a review status for each analysis. The default review state is Primary Review, with the option to select Technical Review, and then Review Complete.

Analysis Settings indicate whether an existing analysis template or user-applied override values are in use. If the analysis settings are from an analysis template, the name of the template is displayed. See *Changing Locus Thresholds* on page 101.

Figure 14 Analysis Header

The following table lists possible analysis states.

Table 5 State of Analysis

State	Description
New	Analysis by the software did not begin yet.
Queued	Analysis is in line to begin processing.
Processing	Analysis by the software is in progress.
Completed	Analysis by the software completed successfully.
Errored	Analysis did not complete successfully. Analysis stopped with no results because of an error.

After selecting an analysis name on the left panel, a box with the following color-coded run quality indicators appears on the page, from left to right. Click the icon for further details.

Figure 15 Analysis Header

Table 6 Run and Control Sample Quality Icons

Icon	Description
P	Positive control quality indicator for positive amplification controls and HSC metrics. See *Viewing Positive Control Metrics*.
N	Negative control quality indicator for reagent blank samples and negative amplification control samples. See *Viewing Negative Control Metrics* on page 27.
Q	Displays overall quality metrics for the run. See *Viewing Quality Metrics* on page 28.

Viewing Positive Control Metrics

The analyses in a project, along with the analysis version number, are listed on the left panel of the Project page. To view results for positive control samples, click the analysis name on the Project page to open the Analysis page, and then on the Positive Control icon (P). Click the tabs to display positive amplification control and HSC.

Figure 16 Positive Control Icon

Positive Control metrics are available after the analysis is complete. Color indicators on the positive control icon (P) show the outcome of control metrics.

Table 7 Positive Control Icon Color Indicator

Color	Positive Amplification Control Samples	HSC Samples
Green	All samples designated positive amplification control in the run, assigned to the analysis, have a green status. All STRs and SNPs are typed and all alleles are concordant with the known Control DNA 2800M.	Overall the sample has sufficient intensity coverage, and the genotype for each locus is concordant.

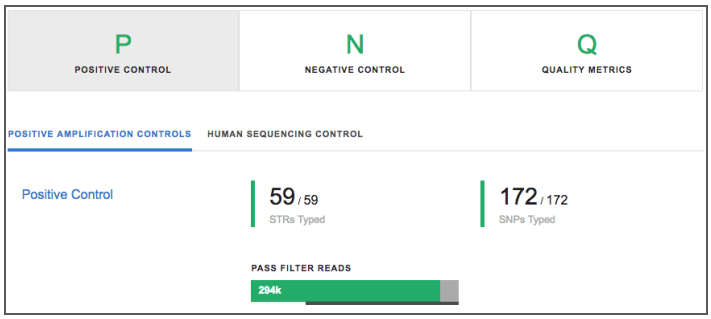
Color	Positive Amplification Control Samples	HSC Samples
Orange	Orange status indicates that at least one STR or SNP is not typed, is typed but discordant with Control DNA 2800M, or the number of reads in the sample is less than 85,000.	The sample might not have sufficient read coverage, or the genotype for a locus is not concordant. Lists all effected loci.

Positive Amplification Control Tab

The Positive Amplification Control tab lists all samples identified as positive amplification control. Each sample is compared with the known Control DNA 2800M multi-locus genotype for each of the STRs and SNPs in the ForenSeq DNA primer mix that is assigned to the sample.

NOTE
Control DNA 2800M is a control for single-source male gDNA and is the positive amplification control DNA provided with the ForenSeq DNA Signature Prep Kit.

Figure 17 Positive Amplification Control Tab



Next to each sample name is a short vertical color bar with the number of STRs and SNPs typed out of the total possible number of STR loci and SNP loci.

Figure 18 Positive Control STRs and SNPs Typed

Table 8 Positive Amplification Control Vertical Bar Color Indicators for STRs

Color	Indication
Green	All STRs in the sample are typed and concordant with the known Control DNA 2800M.
Orange	Indicates that at least one STR in the sample is not typed, or is typed yet discordant with the known Control DNA 2800M multilocus genotype.

Table 9 Positive Amplification Control Vertical Bar Color Indicators for SNPs

Color	Indication
Green	All SNPs in the sample are typed and concordant with the known Control DNA 2800M.
Orange	Indicates that at least one SNP in the sample is not typed, or is typed yet discordant with the known Control DNA 2800M multilocus genotype.

For each positive amplification control sample, the following information is displayed:

- The name of each positive amplification control sample, which can be clicked to open the Sample Details page
- The number of typed STRs and SNPs out of the total possible number of STRs and SNPs in the ForenSeq primer mix
- A list of discordant loci compared to Control DNA 2800M
- The number of reads

If warranted, a discordance table shows all discordant STRs and SNPs compared to Control DNA 2800M and includes the following information:

- Locus name
- Genotype observed in the sample
- Genotype expected from the Control DNA 2800M positive amplification control

A locus is determined to be discordant for the following reasons:

- 1 or more loci in the positive amplification control is not called
- The genotype of 1 or more positive amplification control loci differs from the known Control DNA 2800M genotype

Each positive amplification control sample has a horizontal bar that indicates the number of reads. The dark gray shadow bar below the bar indicates a guideline for 85,000 reads and above.

Figure 19 Positive Control Number of Reads

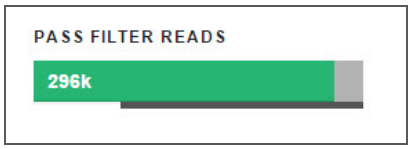


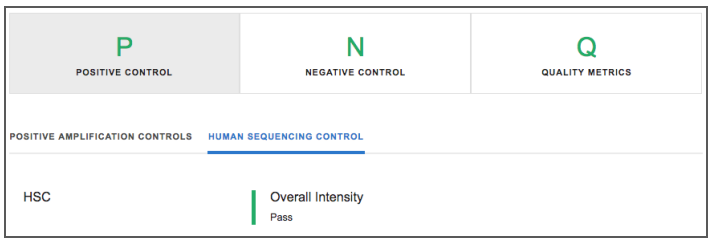
Table 10 Positive Amplification Control Number of Reads Bar Color Indicators

Color	Indication
Green	The total number of reads is above 85,000 reads.
Orange	The total number of reads is below 85,000 reads. Orange does not necessarily mean that the positive control sample has failed. It is possible to use these data, however Verogen recommends verifying the sample data. Look at the data in more detail on the Sample Details page and review the quality results on the Quality Metrics page. See *Viewing Sample Details Page* on page 34, and *Viewing Quality Metrics* on page 28.

ForenSeq Human Sequencing Control Tab

The Human Sequencing Control tab indicates whether HSC meets criteria for intensity, and genotype concordance.

Figure 20 Human Sequencing Control Tab



The loci in the sequencing control are assessed for an expected minimum intensity level and the correct genotype call. If the overall intensity status fails, then a list of relevant loci and the locus lengths is displayed.

Next to the sample name HSC is a short vertical color bar.

Figure 21 HSC Overall Intensity

Table 11 Human Sequencing Control Vertical Bar Color Indicators

Color	Indication
Green	Indicates that the sample meets or exceeds minimum intensity level criterion and genotype concordance.
Orange	Indicates that the sample does not meet 1 or more criteria. If the bar is orange, then it is possible that the overall intensity for HSC is lower than expected. If any loci are listed on the page, then those loci do not meet intensity or genotype concordance criteria, and their locus name and length in base pairs (bp) are listed. For a complete list of HSC loci, see *Human Sequencing Control Loci* on page 144.

Viewing Negative Control Metrics

The analyses in a project, along with the analysis version number, are listed on the left side of the navigation bar of the Project page. To view results for negative control samples, click the analysis name on the Project page to open the Analysis page, and then on the Negative Control icon (N). Click the tabs to display reagent blanks and negative amplification control metrics.

Figure 22 Negative Control Icon

Negative Control metrics are available after analysis is complete. A color indicator on the negative control icon (N) shows the outcome of control metrics.

Table 12 Negative Control Icon Color Indicators

Color	Indication
Green	All samples designated reagent blank or negative amplification control in the run, assigned to the analysis, have a green status. Green status indicates that in each control sample, no SNPs or STRs are typed.
Orange	At least one sample designated reagent blank or negative amplification control in the analysis has an orange status. Orange status indicates that at least one STR or SNP is typed.
Grey	No samples are designated reagent blank or negative amplification control in the analysis assigned to the project.

Reagent Blanks Tab

The Reagent Blanks tab lists all samples identified as reagent blank. Each sample is compared with STRs and SNPs in the ForenSeq DNA primer mix assigned to the sample. Next to each sample name, the tab shows how many STRs and SNPs are typed out of the total possible number of target loci. You can click the sample name to open the Sample Details page.

Figure 23 Reagent Blanks Tab

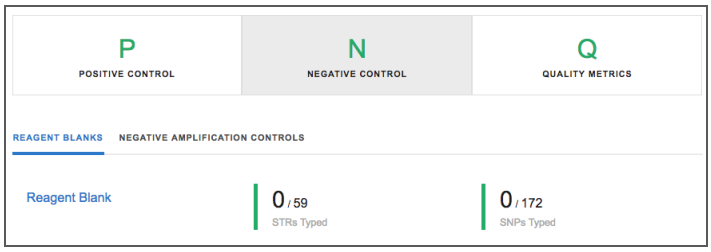


Table 13 Reagent Blanks Tab Vertical Bar Color Indicators

Color	Indication
Green	Indicates that no STR or SNP locus is typed.
Orange	Indicates that at least one STR or SNP locus is typed.

Negative Amplification Controls Tab

All samples identified as negative amplification control are listed in the Negative Amplification Controls tab. You can click the sample name to open the Sample Details page. Each negative amplification control sample is compared with the STRs and SNPs in the ForenSeq DNA primer mix assigned to that sample. Next to each sample name, the tab shows how many STRs and SNPs are typed out of the total possible number of loci targeted by the ForenSeq DNA Signature Prep Kit.

Figure 24 Negative Amplification Controls Tab

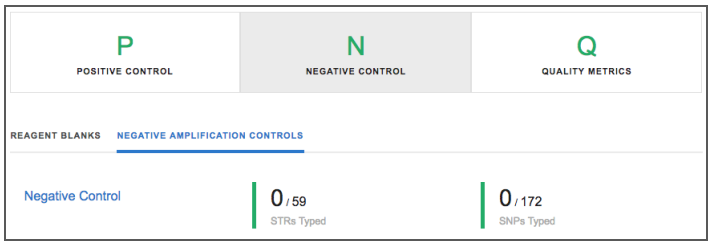


Table 14 Negative Amplifications Controls Tab Vertical Bar Color Indicators

Color	Indication
Green	Indicates that no STR or SNP locus is typed.
Orange	Indicates that at least one STR or SNP locus is typed.

Viewing Quality Metrics

The analyses in a project, along with the analysis version number, are listed on the left side of the Project page. To view overall run quality metrics and sample representation, click the

name of an analysis on the Project page to open the Analysis page, and then on the Quality Metrics icon (Q). Click the tabs to display Run Metrics and Sample Representation. Run metrics information can also be accessed on the system dashboard by hovering on the Q next to the run name.

Figure 25 Quality Metrics Icon

Run Quality Metrics show the intensity (number of reads) of each sample in the analysis, and a number of cycles completed indicator. This page enables remote monitoring of the run, and all the run metrics available during sequencing on the MiSeq FGx instrument.

As the run progresses, a gray run progress bar displays on the system dashboard under the run name, and the MiSeq FGx instrument updates the ForenSeq Universal Analysis Software with details of run progress. The metrics presented mirror the information displayed on the MiSeq FGx instrument during sequencing. After sequencing of the run completes, the information for the run is preserved for run quality assessment, as this Quality Metrics page remains static so that results from the run are available.

A color indicator on the quality metrics icon (Q) shows the outcome of quality metrics.

Table 15 Quality Metrics Icon Color Indicators

Color	Indication
Green	All quality metrics are in the recommended range for a run.
Orange	One or more quality metric is not in the recommended range for a run.

Run Metrics Tab

To view overall run quality, click the Run Metrics tab. Horizontal bars show the acceptable range of values for cluster density, clusters passing filter, phasing, and prephasing. A color dot indicates Read and Index quality.

Figure 26 Run Metrics Tab

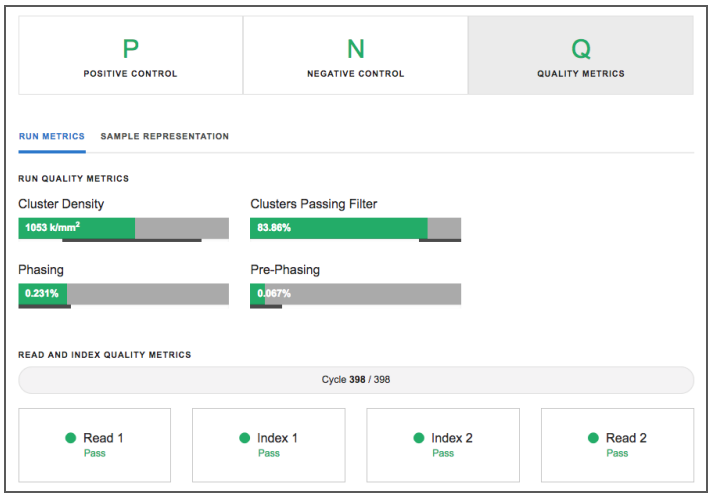
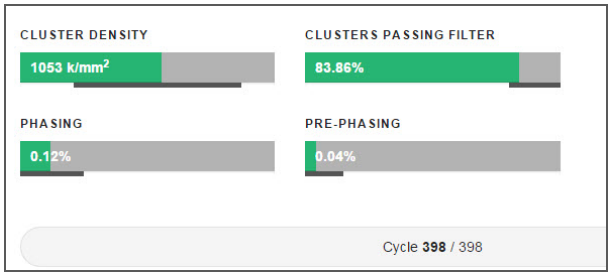


Table 16 Run Metrics Horizontal Bar Color Indicators

Color	Indication
Green	Green indicates that values are within the acceptable range.
Orange	Orange indicates the possibility that further evaluation of the quality of the run is recommended.
Dark Gray	The dark gray shadow bar below the result bar indicates the passing range.

Figure 27 Run Metric Indicators



- **Cluster Density (K/mm²)**—Shows the number of clusters per square millimeter for the run. For ForenSeq runs, a target cluster density range of 400–1650 K/mm² is recommended. Cluster density values outside of the target range can still produce results that are sufficient to use for analysis. Values that deviate substantially from the target range can negatively impact other quality metrics, and decrease the quantity of valuable data from the run.
- **Clusters Passing Filter (%)**—Shows the percentage of clusters passing filter based on the Illumina chastity filter, which measures quality. The filter can detect low quality base calls. Data appears only after cycle 25.

NOTE
The chastity of a base call is the ratio of the intensity of the greatest signal divided by the sum of the 2 greatest signals. If more 1 one base call has a chastity value of less than 0.6 in the first 25 cycles, reads do not pass the quality filter.

For ForenSeq samples, a target clusters passing filter value of ≥ 80% is recommended. Clusters passing filter values that are outside of the target range can still produce results that are sufficient to use for analysis. Values that deviate substantially from the target range can negatively impact other quality metrics, and decrease the quantity of data from the run.

- **Phasing (%)**— Shows the percentage of molecules in a cluster that fall behind the current cycle within Read 1 and Read 2. Low percentages indicate good run statistics. For ForenSeq samples, a phasing value of ≤ 0.25% is recommended. Phasing values outside of the target range can still produce results that are sufficient to use for analysis. See *Phasing and Prephasing* on page 125.
- **Prephasing (%)**— Shows the percentage of molecules in a cluster that run ahead of the current cycle within Read 1 and Read 2. Low percentages indicate good run statistics. For ForenSeq samples, a prephasing value of ≤ 0.15% is recommended. Prephasing values outside of the target range can still produce results that are sufficient to use for analysis. See *Phasing and Prephasing* on page 125.
- **Cycle**— Shows the number of sequencing cycles completed. One cycle includes the chemical addition and imaging of one base for each cluster on a flow cell. A total of 398 cycles are performed on a ForenSeq run.

Table 17 Read and Index Metrics Color Indicators

Color	Indication
Green	The average quality for assessed reads is within the recommended range.
Orange	The average quality for assessed reads is not within the recommended range.
Gray	The read or index did not occur yet in the sequencing run.

Figure 28 Read and Index Status Indicator

- **Read 1**— Read 1 follows the Read 1 sequencing protocol using the MiSeq FGx Reagent Kit or MiSeq FGx Reagent Micro Kit. The Read 1 sequencing primer is annealed to the template strand during the cluster generation step. The RTA software evaluates the first 50 cycles of the Read 1 segment of the run for quality. Read 1 quality metrics are displayed only after cycle 50 is complete.
- **Index 1**— The Read 1 product is removed, and the Index 1 (i7) sequencing primer is annealed to the same template strand as in Read 1. Following Index Read preparation, the Index 1 (i7) Read performs 8 cycles of sequencing. The RTA software evaluates all 8 cycles of the Index 1 segment of the run for quality. Quality results are displayed only after cycle 359 is complete.
- **Index 2**— The Index 1 (i7) Read product is removed, and the template anneals to the grafted P5 primer on the surface of the MiSeq FGx flow cell. The run proceeds through an additional seven chemistry-only cycles in which no imaging occurs, followed by eight cycles of sequencing. The RTA software evaluates all eight cycles of the Index 2 (i5) segment of the run for quality. Quality results are displayed only after cycle 367 is complete.
- **Read 2**— The Index Read 2 product is extended to copy the original template strand. Then, the original template strand is removed and the Read 2 sequencing primer is annealed. Read 2 continues for 30 cycles to sequence through the reverse PCR primer SBS reagents. The RTA software evaluates all cycles of the Read 2 segment of the run for quality.

Table 18 Cycle Numbers and Corresponding Sequencing Phases

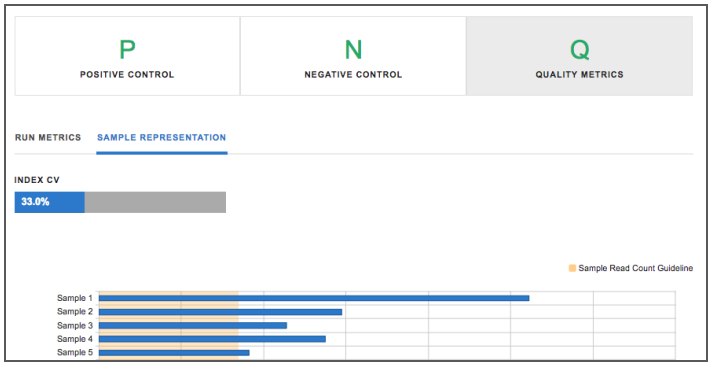
Cycle Number	Sequencing Phase
1 to 351	Read 1
352 to 359	Index 1
360 to 367	Index 2
368 to 398	Read 2

Sample Representation Tab

Quality Metrics for the run show intensity (number of reads) of each sample in the analysis, and a number of cycles completed indicator. This page enables remote monitoring

of the run, and all the run metrics available during sequencing on the MiSeq FGx instrument. Click the tabs to display Run Metrics and Sample Representation.

Figure 29 Sample Representation Tab



The quality Sample Representation tab shows an overall view of the number of reads, and read distribution for each sample in the analysis.

- **Index CV**— Displayed as a percentage, the number of reads that are assigned to each sample. CV is the coefficient of variation for the number of read counts across all indexes. Index CV represents the distribution of read counts of the samples in the run.

Figure 30 Index CV

Below the Index CV bar is a bar graph that shows the number of reads for each sample. To view the exact number of reads for a sample, mouse over the bar. You can click the bar or sample name to open the Sample Details page.

Figure 31 Sample Representation Graph

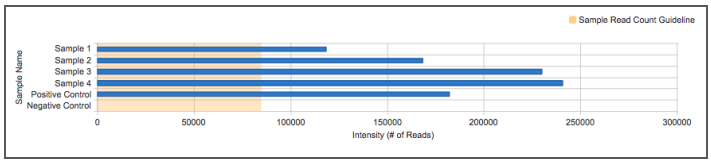


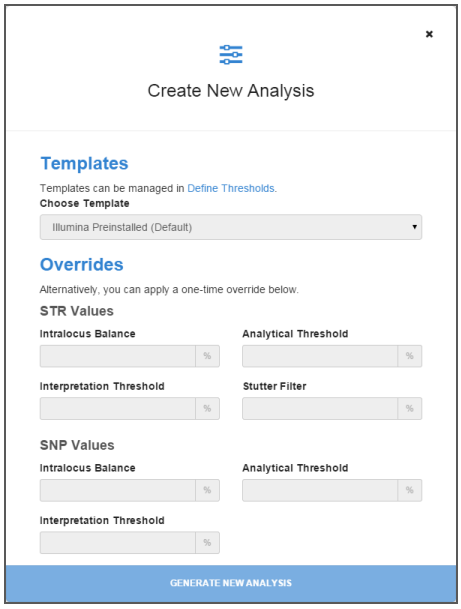
Table 19 Color Indicators of the Sample Representation Tab Graph

Color	Indication
Light Orange	The light orange shaded region from 0 to 85,000 total reads indicates the Sample Read Count Guideline region. Interpret samples below 85,000 reads with caution, as loci with low coverage might not have enough signal for alleles to be distinguished from noise.
Blue	A sample with a blue bar in the graph indicates a sample that meets signal intensity recommendations.
Dark Orange	A sample with a dark orange bar in the graph indicates a sample that does not meet signal intensity recommendations. Samples with less than 85,000 total reads may still include enough data for interpretation.

Creating a New Analysis

The Create New Analysis feature is a traceable, flexible way to create alternate versions of major analyses. In a new analysis, you can select a different system settings template or apply a one-time override to system settings. The creation of a template requires administrator access. See *Creating a Locus Threshold Template* on page 101. Create New Analysis is available after a run completes sequencing.

Figure 32 Create New Analysis



If you apply a one-time override, you have the option of changing intralocus balance, analytical threshold, and interpretation threshold settings for STRs and SNPs. You can also change the stutter filter setting for STRs.

The Create New Analysis command is accessed on the Analysis page. See *Viewing Run Quality Information* on page 22.

1. Select the name of a major analysis on the left panel.
   Major analyses version numbers end in 0. For example, an analysis with the name *3.0 - Your Example Run* is a major analysis. An analysis with the name *3.1 - Your Example Run* is not a major analysis, and no new analysis can be created from it.
2. Click **Create New Analysis**.
3. Select from the Choose Template drop-down list or enter one-time override values in the Overrides section.
4. Click **Generate New Analysis**.

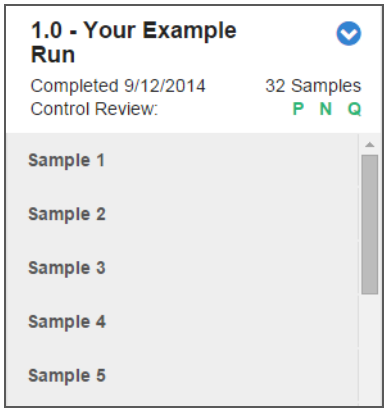
NOTE
After you execute analysis, the analysis version cannot be deleted.

NOTE
The run version created for an initial sequencing run is version 1.0, and the initial analysis of that run is version 1.0. If you create additional analyses for a run, the analyses are numbered in sequence as in 1.1, 1.2, etc. For example, a new analysis of 1.0 is named version 1.1. A second analysis of 1.0 is named version 1.2, and so forth. The Create New Analysis command can be applied to major analyses only, such as 1.0, 2.0, 3.0, etc.

Viewing Sample Details Page

To view sample locus results, click the Project page. See *Viewing the Project Page* on page 21. Locate an analysis on the left pane of the page, and click the blue arrow to open the list of samples in the run. A blue down arrow indicates that the sample list for the analysis is open. To close the list of samples, click the blue arrow. A blue arrow pointing towards the left indicates that the sample list is closed.

Figure 33 Samples List



To view the Sample Details page with STR and iSNP results, click a sample in the list. The sample name, index numbers, gender, sample type, ForenSeq DNA primer mix type, and description are displayed at the top of the page.

Figure 34 Samples Name

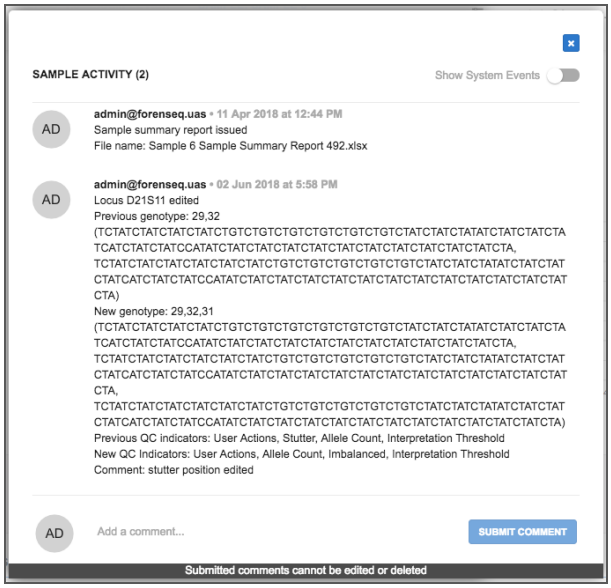
Enabling visibility of the Sample History feature requires ForenSeq Universal Analysis Software administrator privileges. When enabled, the Sample Activity dialog containing a list of system-initiated events (e.g., population statistics calculation completed) and user actions (e.g., locus edited) can be viewed by selecting the user actions icon on the Sample Details page.

The user actions icon provides an indicator for the number of sample-level user actions that have been logged, such as comments, sample reports, and locus edits. Selecting the icon displays the user responsible, time and date, and details for each event. Additional system events are displayed by adjusting the toggle in the upper-right corner of the dialog.

To add a comment to a sample, use the text box at the bottom of the screen, and select **Submit Comment**.

NOTE
After a comment is submitted, it is permanent and cannot be modified or deleted.

Figure 35 Sample Activity



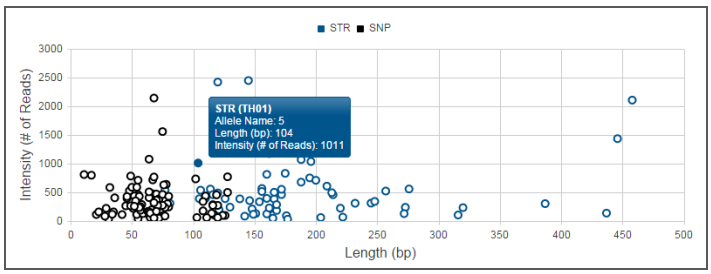
Reviewers can indicate a review status for each sample. The review states are Primary Review (default), Technical Review, and Review Complete.

Intensity and Length Graphs

To see sample intensity and length information for STRs graphed separately from SNPs, click the blue STR or black SNP squares above the graph to turn the display off or on. Using the mouse, hover over the squares to highlight the display of STRs or SNPs.

Hover over a dot in the graph to view the exact length in base pairs and precise intensity level for a data point. The display for STRs includes the allele name.

Figure 36 Intensity and Length Graphs



- **Intensity (# of Reads)**—For each typed STR or iSNP allele in a sample, the graph shows signal intensity reported as the number of reads.
- **Locus Length (bp)**—For each typed allele in a sample, the graph shows the length of the amplicon as the number of base pairs including the PCR primers for STR alleles and excluding PCR primers for iSNPs.

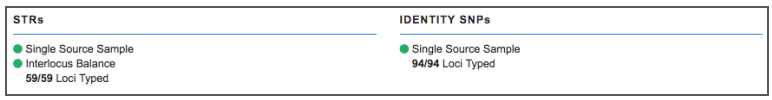
  For information about loci length, see *Autosomal STRs* on page 155 and *Identity Informative SNPs* on page 156.

General Locus Information

The following STR and iSNP information is displayed above the sample locus detail tables:

- **Single-Source Sample**—The single-source indicator is based on the results from the ForenSeq Universal Analysis Software algorithms for mixture detection. This approximation is performed separately for STRs and SNPs. For STRs, the number of loci with allele count quality control indicators must be > 5. For SNPs, the number of imbalance quality control indicators for all SNP loci must be > 10. See *STR Genotype Calling* on page 127 and *SNP Genotype Calling* on page 129. The calculation of the single-source indicator is updated if manual updates are made to the genotypes. See *Updating and Modifying Typed Results* on page 43.

Figure 37 Single-Source and Interlocus Balance Indicator Icons



NOTE
Evaluate and verify single-source sample results based on your internal guidelines.

Table 20 Single-Source Color Indicators

Color	Indication
Green	Indicates that the sample might be a single-source sample.
Orange	Indicates that the sample might contain multiple DNA contributors.

- **Interlocus Balance**—For STRs, the balance of the read counts between STR loci in a sample. This balance is measured as the % coefficient of variation (CV) of the read counts across all of the STR loci in the ForenSeq multiplex. The % CV is calculated as the standard deviation of the total read counts of each locus divided by the mean of locus read counts.

Table 21 Interlocus Balance Color Indicators

Color	Indication
Green	The balance is in the recommended range.
Orange	The interlocus balance metric falls outside of the optimum range.

- **Loci Typed**— Number of loci typed (by the analysis software or updated manually) out of all STR loci. Any loci that are deselected at the time of analysis creation are not included in the number of typed loci. See *Selecting Loci Content for Analysis* on page 105. For SNPs, the number of loci typed represents the iSNPs only. Phenotypic SNP (pSNP) and biogeographical ancestry (aSNP) information is available in the *Phenotype Estimation SNP Sample Details Table* on page 63. For more information on ForenSeq DNA primer mix types, see *Entering Sample Information* on page 13.

Locus Detail Box Indicators

Each box in the STR and iSNP sample locus details table has a color border that indicates whether the locus is within or outside of the ForenSeq Universal Analysis Software quality metrics. If applicable, quality control indicator icons are displayed in the box.

Table 22 Locus Detail Box Color Indicators

Color	Indication
Gray	Quality control indicators are not active for any locus.
Orange	One or more quality control indicators are active for the locus.

If applicable, a quality control indicator icon is displayed in an STR or iSNP locus detail box. For more information about quality indicator icons, see *STR Genotype Calling* on page 127 and *SNP Genotype Calling* on page 129.

Table 23 Sample Details Quality Control Indicator Icons

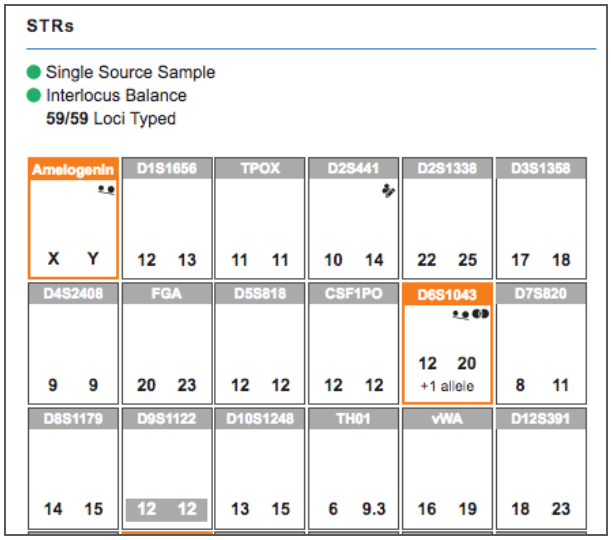
Symbol	Symbol Name	Indication
	Stutter	For STRs, the stutter filter percentage is exceeded by one or more sequences at a stutter position of a possible parent allele. See *STR Genotype Calling* on page 127.
	Allele Count	For STRs, depending on the assigned gender of the sample, there are more alleles above the analytical threshold than expected, or that can be attributed to stutter. Serves as a potential tool for mixture detection and resolution. See *Allele Counting* on page 127.
	Imbalanced	Read count ratio falls below the user-defined intralocus balance in the analysis settings. For more information on imbalance, see *STR Genotype Calling* on page 127 and *SNP Genotype Calling* on page 129.
	Low Coverage	Signal above the interpretation threshold was not detected as defined in the analysis settings. See *STR Genotype Calling* on page 127 and *SNP Genotype Calling* on page 129. The low coverage indicator can be seen in results from version 1.1 or earlier of the ForenSeq Universal Analysis Software.
	Interpretation Threshold	Indicates that there is at least one allele that is above the analytical threshold, but below the interpretation threshold, that is not attributed to stutter.
	User Actions	A user-initiated action occurred at the analysis, sample, or locus level. For example, at least one allele was manually edited as typed or not typed, or a user comment was submitted. See *Updating and Modifying Typed Results* on page 43.
	Analytical Threshold	Indicates that the locus has signal below the analysis threshold, and no alleles above the interpretation threshold.
	Not Detected	No signal was detected for the locus.

STR Sample Details Table

The STR Sample Detail table consists of boxes for each target locus targeted in the ForenSeq DNA Primer Mix used from the ForenSeq DNA Signature Prep Kit (DPMA or DPMB). All boxes are labeled with the locus name at the top. Click a box to see locus details. To view sample locus details in a report, see *Sample Genotype Report* on page 78.

A box for Amelogenin is in the upper-left corner. All other boxes display STR data and are ordered by their ascending location in the genome.

Figure 38 STR Sample Details Table



For autosomal STRs in single-source samples, two different numbers in a box indicate a heterozygous genotype. Two identical numbers in a box indicate a homozygous genotype, based on STR length alone. Review the locus details for further information and potential sequence variation.

If more than two alleles are typed, the two most abundant alleles are displayed and a plus sign is added with the number of additional alleles under the displayed alleles. STRs with isometric alleles are highlighted with a gray box.

Locus Details

Click a box to open a locus details table to view the signal intensity and the sequence for the alleles at the locus. The display is color-coded and the rows in the table match the colors of the bars in the chart.

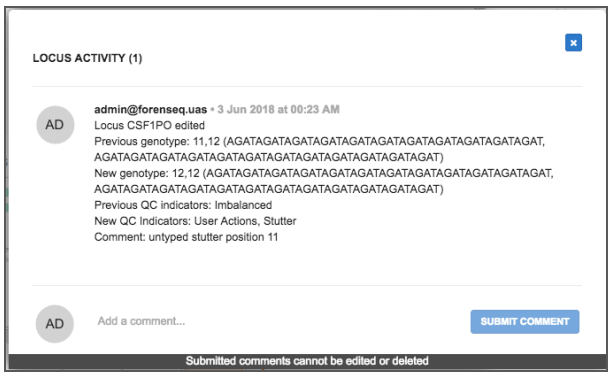
Figure 39 Locus Details Bar Chart

Enabling visibility of the Sample History feature requires ForenSeq Universal Analysis Software administrator privileges. When enabled, the Locus Activity dialog containing a list of user actions (e.g., locus edited) can be viewed by selecting the user actions icon above the locus details table. The user actions icon provides an indicator for the number of locus-level user actions that have been logged, such as locus edits and comments. Selecting the icon displays the user responsible, time and date, and details for each event.

To add a comment to a locus, use the text box at the bottom of the screen, and select **Submit Comment**.

NOTE
After a comment is submitted, it is permanent and cannot be modified or deleted.

Figure 40 Locus Activity



The rows in the locus details table can be sorted by allele number, or by intensity, which is the number of reads. Click the Allele column heading to sort the data according to the allele call. Click the Intensity column heading to sort rows according to the number of reads (intensity).

In the locus details table, a toggle switch in the Typed column that displays a blue background and is positioned on the right side indicates that the allele is contributing to the genotype for the locus. A toggle switch in the Typed column that displays a gray background and is positioned on the left side indicates that the allele is not contributing to the genotype of the locus. The table also lists potential stutter and DNA repeat sequences.

The locus details bar chart shows potential alleles along the X axis. The analytical threshold is in dark gray and interpretation threshold for intensity is in light gray. Hover over the information icon to the left of the legend to see analytical and interpretation threshold values. To view the precise signal intensity of a potential allele, hover over the bar chart and list in the locus details table. When STR alleles of the same length, but of a different sequence, are detected (i.e., isometric alleles), they are displayed in the bar chart as a single bar with a horizontal line delineating each sequence variant (allele). Letters highlighted in white indicates isometric SNPs in the sequence.

You can change the typed status of a sample in a locus detail box in the STR Sample Details table. For more information, see *Updating and Modifying Typed Results* on page 43.

Figure 41 Locus Details Bar Chart

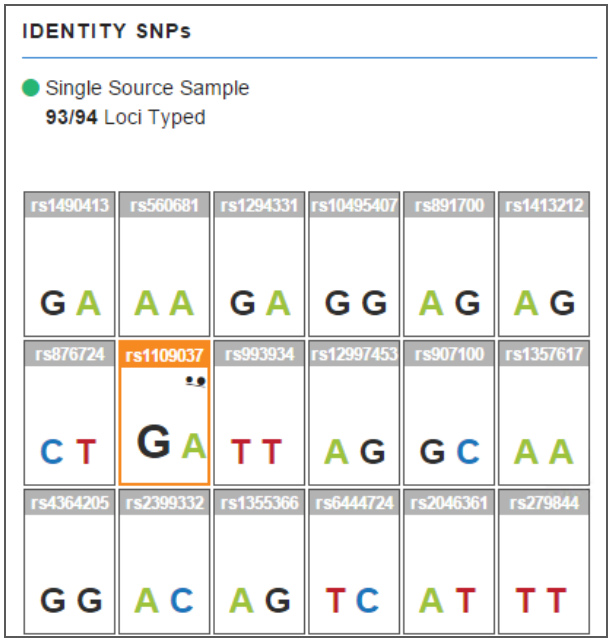
Table 24 STR Locus Details Table and Bar Chart Color Indicators

Color	Indication
Blue	Typed allele.
Brown	DNA sequence that is not typed, as its read count is greater than the analytical threshold, but less than the stutter filter % and is potentially stutter, or is the same length as a typed allele.
Pink	DNA sequence that is not typed and consists of read counts greater than the analytical threshold and stutter filter %. It meets analytical threshold and stutter filter limits. However, intensity is less than the interpretation threshold or there are no other typed alleles in the position of the parent allele relative to potential stutter. For more information on stutter filtering, see *STR Genotype Calling* on page 127.
Red	DNA sequence that is not typed and consists of read counts that are less than the analytical threshold, interpretation threshold, and stutter filter %, and is not the same repeat length as a typed allele. The signal does not fall below the stutter filter, relative to any typed allele. For more information on stutter filtering, see *STR Genotype Calling* on page 127.
Grey	Signal not typed. Read counts are less than the analytical threshold.

SNP Sample Details Table

The SNP Sample Details table consists of boxes for each target locus in the multiplex. All boxes are labeled with the locus name at the top. Click a box to see locus details. To view sample locus details in a report, see *Sample Genotype Report* on page 78.

Figure 42 SNP Sample Details



In each box, nucleotide abbreviations refer to the genotype of the locus. Signal imbalance is indicated by the font size of one letter (A, T, G, C) displayed larger than the other letter. The larger of the two letters has a greater intensity.

Click a box to open a locus details table and pie chart. The display is color-coded with the nucleotide letters in the locus details table corresponding to the colors of the pie chart.

Figure 43 Locus Details Pie Chart—Example 1

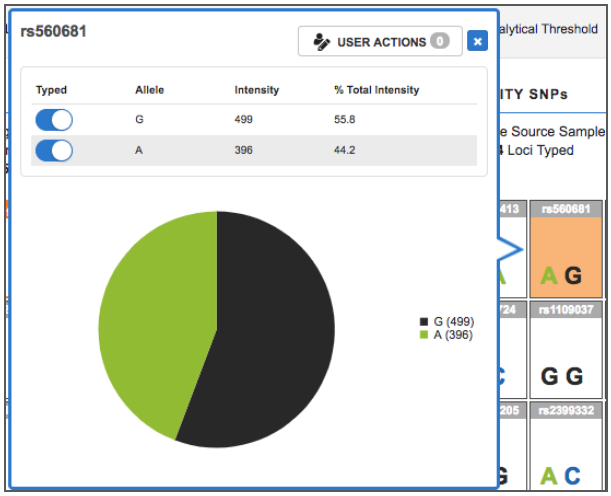
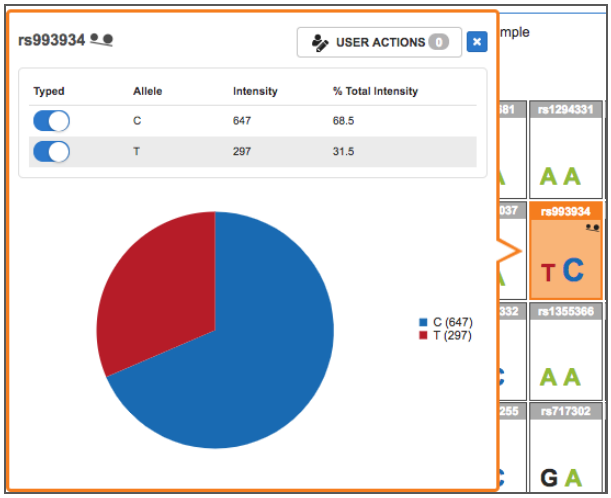


Figure 44 Locus Details Pie Chart—Example 2



Enabling visibility of the Sample History feature requires ForenSeq Universal Analysis Software administrator privileges. When enabled, the Locus Activity dialog containing a list of user actions (e.g., locus edited) can be viewed by selecting the user actions icon above the locus details table. The user actions icon provides an indicator for the number of locus-level user actions that have been logged, such as locus edits and comments. Selecting the icon displays the user responsible, time and date, and details for each event.

To add a comment to a locus, use the text box at the bottom of the screen, and select **Submit Comment**.

NOTE
After a comment is submitted, it is permanent and cannot be modified or deleted.

The rows in the locus detail box can be sorted by Typed, Intensity, or % Total Intensity by clicking their respective column headings.

The table and pie chart display the intensity and relative percentage of the signal at the SNP locus position. A toggle switch in the Typed column that displays a blue background and is in position on the right side indicates that analysis settings instructed the ForenSeq Universal Analysis Software to call a particular allele. A toggle switch in the Typed column that displays a gray background and is in position on the left side indicates that analysis settings that drive allele calling are not met.

The legend to the right of the pie chart, and the Intensity column in the table above display the intensity of possible SNP alleles. Using the mouse, hover over the pie chart segments or look in the % Total Intensity column of the table to see the percentage of total intensity at the locus.

Table 25 Nucleotide Colors in SNP Locus Details Pie Chart

Color	Nucleotide
Green	A
Red	T
Blue	C
Black	G

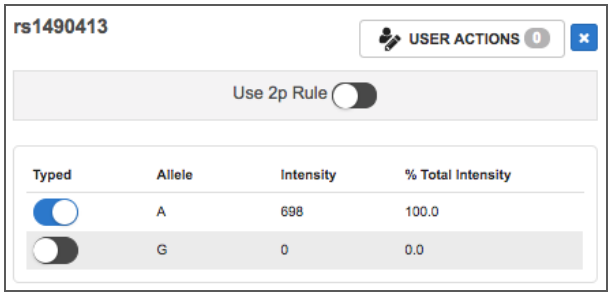
Updating and Modifying Typed Results

You can change the typed status of a sample in a locus detail box in the STR or SNP Sample Details table.

1 In the Sample details table, click the box to open the table and chart.
A toggle switch in the Typed column that displays a blue background and is in position on the right side indicates that the allele is typed. A toggle switch in the Typed column that displays a gray background and is in position on the left side indicates that the allele is not typed.

2 To change the type status of an allele, do one of the following:
- To change from typed to untyped, click the toggle switch so that the switch displays a gray background and is in position towards the left.
- To change from untyped to typed, click the toggle switch so that the switch displays a blue background and is in position towards the right.

Figure 45 SNP Locus Details Table



The type status of an allele can also be changed by clicking on the bar (in the case of an STR) or the part of a pie chart (in the case of a SNP) that represents the allele.

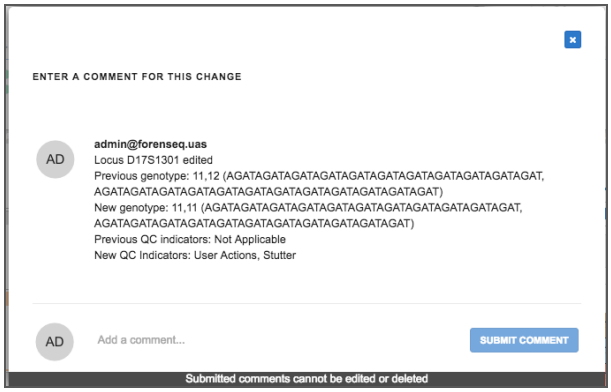
To apply the 2p rule to a homozygous locus, click the toggle switch so that the switch displays a blue background and is in position on the right. For more information, see *Defining the Statistics Calculation Method* on page 108 and *Applying the 2p Rule to a Homozygous Locus* on page 109.

When a change is made, a user actions icon resembling a person with a pencil appears in the box in the sample details table and the locus details box. If you return the allele to its original status (typed or untyped), the user actions icon no longer appears. The typed status of a sample cannot be modified while a sample or project report that includes the sample is generating.

When a genotype is updated with visibility of the Sample History feature enabled, a dialog opens that permits the addition of an optional comment for the change. If provided, the comment is associated with the change and reported in the sample-level reports. For more information, see *Sample Genotype Report* on page 78.

NOTE
After a comment is submitted, it is permanent and cannot be modified or deleted.

Figure 46 Optional Locus Comment



Updating and Modifying Gender Results

You can update the gender designation (XX, XY) of a sample on the Sample Details page. See *Viewing Sample Details Page* on page 34. The sample gender is displayed at the top of the page. When gender designation of a sample is updated, the update only applies to that particular analysis.

1 To update the gender designation of the sample, click **Gender**.
The Modify Gender Call box opens.

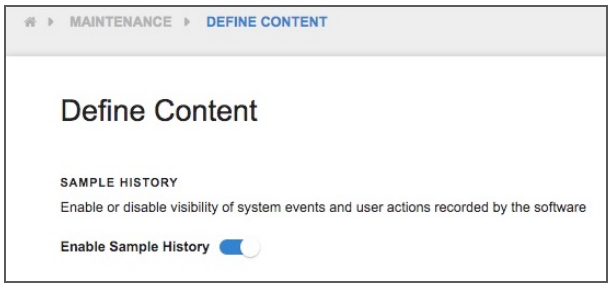
2 Select a gender.

3 Click **Submit**.

Sample History: User Action and System Event Tracking

Enabling visibility of the Sample History feature requires ForenSeq Universal Analysis Software administrator privileges and is accessible in the Define Content section of the Maintenance page. For more information, see *Defining Content Within the Application* on page 105.

Figure 47 Enable Visibility of the Sample History Feature



When enabled, certain user actions and system-initiated events that are logged by the software are visible as a list of events in analysis-level, sample-level, and locus-level activity dialogs. Additionally, these same events are recorded in the sample-level genotype reports (Sample History tab in Excel report).

If the Sample History feature is disabled, user action and system-initiated events continue to be tracked by the software; however, they are not visible as activity dialogs in the user interface and are not included in the sample reports. If visibility of the feature is enabled at a later date, all previously tracked activity associated with that sample will be visible in the activity dialogs and recorded in subsequently created sample reports.

Population Statistics

Introduction

ForenSeq Universal Analysis Software features the ability to calculate population statistics relative to an allele frequency database. After STR and iSNP multilocus genotyping is complete, population statistics can be calculated automatically. Population groups are installed with the software and their use can be excluded or included in calculations. Customized population groups can also be added for statistical calculation. See *Population Group Settings* on page 108.

Population Statistics Calculations

The ForenSeq Universal Analysis Software performs probability calculations for multilocus autosomal STR or iSNP profiles. The genotype for each locus must either meet the 2n copy number expected for autosomal loci, or must be uncalled. Autosomal STR calculations follow Scientific Working Group for DNA Analysis Methods (SWGDAM) guidelines. The model assumes that populations are in Hardy-Weinberg equilibrium, that loci are in linkage equilibrium with one another, and a basic correction is included for population substructure.

NOTE
If the copy number is inconsistent, such as three alleles typed for an autosomal STR, update and modify the locus calls to successfully calculate population statistics. See *Updating and Modifying Typed Results* on page 43.

NOTE
Population statistics calculations follow SWGDAM Interpretation Guidelines for Autosomal STR Typing by Forensic DNA Testing Laboratories.[1]

For homozygote genotypes, $p^2 + p(1 - p)\ \theta$ is used in following determinations by NRC II.[2]

NOTE
At this time, population statistics are calculated based on alleles defined by length, not by sequence. Population statistics for sequence-based alleles may be calculated from internal or published sequence-based allele frequency databases.

Population statistics calculations use the following information:

- Single-source autosomal STR or iSNP genotype calls
- Allele frequency estimates from populations reported by the National Institute of Standards and Technology (NIST) and other contributing literature[3,4,5]
- Minimum allele frequency, as defined by selected population groups
- Population substructure correction factor θ, when applicable to selected population groups

The ForenSeq Universal Analysis Software queries the population groups and loci defined by the administrator at the time of calculation.

Resources

1. Scientific Working Group on DNA Analysis Methods. Interpretation Guidelines for Autosomal STR Typing by Forensic DNA Testing Laboratories. Approved 2017 January. https://docs.wixstatic.com/ugd/4344b0_50e2749756a242528e6285a5bb478f4c.pdf.

2. National Research Council. The Evaluation of Forensic DNA Evidence. Washington, DC: The National Academies Press. 1996. https://doi.org/10.17226/5141.

3. Hill CR, Duewer DL, Kline MC, Coble MD, Butler JM. U.S. population data for 29 autosomal STR loci. *Forensic Sci Int Genet.* 2013;7(3):e82–e83.

4. Novroski N, King J, Churchill J, Seah L, Budowle B. Characterization of genetic sequence variation of 58 STR loci in four major population groups. *Forensic Sci Int Genet.* 2016 Sept;25:214-226.

5. Churchill J, Novroski N, King J, Seah L, Budowle B. Population and performance analyses of four major populations with Illumina's FGx Forensic Genomics System. *Forensic Sci Int Genet.* 2017 June;30:81-92.

Generating Population Statistics

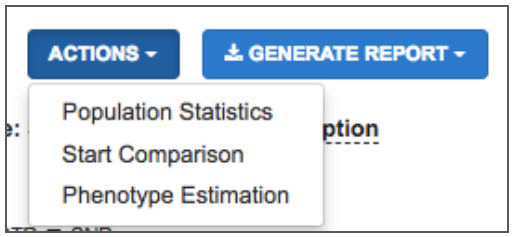
Follow these instructions to calculate population statistics.

1 From the Project page, locate an analysis on the left pane, and click the blue arrow to open a list of samples. See *Viewing the Project Page* on page 21.

2 Select a sample from the project.

3 Click the blue Actions drop-down list button and select **Population Statistics**.

Figure 48 Population Statistics



4 Select Population Groups to include in the random match probability calculation.

5 Click **Generate**.

6 A population statistics processing bar shows calculation progress.

7 Population statistics results are displayed in horizontal bars.

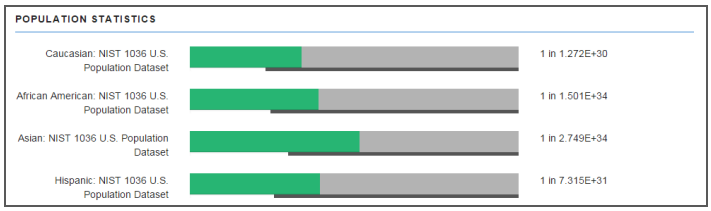
NOTE
Population statistics generation can take several minutes. While waiting, you can navigate to other pages and perform other functions in the software.

8 To view population statistics results later, return to the Project page and the Comparisons tab, and select the relevant analysis.

Population Statistics Results

Population group names that are used in the calculation are displayed on the left. Rarity of probability is represented with a horizontal green bar. The horizontal result bar is scaled to the range of possible frequencies for a population group. The range represents all values possible with the population group and the loci used and is represented on log10 scale. The gray bar represents the source attribution threshold set for the system. To modify source attribution threshold values, see *Changing the Source Attribution Threshold* on page 109.

Figure 49 Population Statistics Results



Hardy-Weinberg Expectations and Linkage Equilibrium

Results of preliminary data analyses of the autosomal STRs and iSNPs in the ForenSeq DNA Signature Prep Kit are consistent with previous experience and publications with genetic identity DNA markers. The autosomal genetic identity markers in the ForenSeq kit generally meet expectations of independence at the population level. The recommendations of the NRC Report II (1996) should be followed for estimating the rarity of a multilocus DNA profile.[1] For kinship analyses, physically close markers may not be inherited independently. There are two options recommended for the use of genetic data for kinship analyses: (1) Incorporate the recombination rate and maximum likelihood estimates of haplotype frequencies for the pair of loci. Note that if the data do not involve double heterozygotes for any individual, then the recombination rate is not needed in the computations. LE assumption allows haplotype frequency estimates from the product of allele frequencies; or (2) use only one of the two loci in a kinship analysis, which would be the more informative of the two in a specific case.[2]

Resources

1. Committee on DNA Forensic Science: An Update, National Research Council. The evaluation of DNA forensic evidence. *Washington D.C.: National Academies Press*; 1996.

2. Dr. Bruce Budowle, Ph.D, Executive Director of the University of North Texas Institute of Applied Genetics, and Dr. Ranajit Chakraborty, Ph.D, Director of the Center for Computational Genomics of the University of North Texas Institute of Applied Genetics, personal communication. 2015 February.

Sample Comparison

Chapter 4

Introduction

The ForenSeq Universal Analysis Software sample comparison feature evaluates the genotyping results between two samples using autosomal, X and Y STRs and iSNPs. You can compare samples from the same project, or different projects. Sample comparison compares typed STRs and typed iSNPs from the two samples. If ForenSeq DNA Primer Mix B is used, aSNPs and pSNPs can be used in phenotype estimation, but are not included in the comparison function. A typed STR and iSNP locus is designated as discordant, within the context of a sample comparison, if the genotype is not shared by both samples.

Generating a Sample Comparison

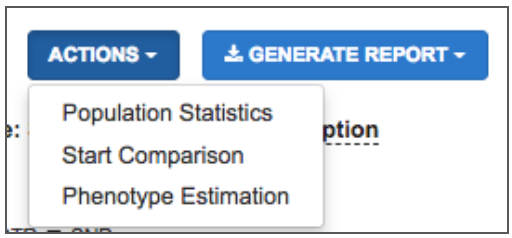
Follow these instructions to compare two single-source samples.

1 From the Project page, locate an analysis on the left pane, and click the blue arrow to show a list of samples. See *Viewing the Project Page* on page 21.

2 Select a sample from the project. In the sample comparison, this sample is called Sample A.

3 Click the blue Actions drop-down list button and select **Start Comparison**.

Figure 50 Start Comparison



4 Click the blue **Choose Sample** button.
A Sample Compare Selection box opens.

5 In the Sample Compare Selection box, select a project from the Project Name drop-down list.

6 A list of samples opens that are associated with the project. Use the search field or scroll through the list of samples from the project to find the second sample you want to compare. In the sample comparison, this sample is called Sample B.

7 Sample comparison starts automatically when you click the blue **Select Sample** button.

8 A sample comparison progress bar indicates how much time is left until the comparison is complete.
The sample comparison page opens.

9 When the comparison is complete, results are displayed.

NOTE
Sample comparison generation can take several minutes. While waiting for sample comparison to process, you can navigate to other pages, and perform other functions in the software.

To view sample comparison results later, return to the Project page and click the Comparisons tab. A link to the results page is presented next to a description of the compared samples and the date.

After the sample comparison is complete, you can select another sample to compare by clicking the blue Choose Sample button showing beneath the name of the second sample on the Sample Comparison page.

Sample Comparison Results

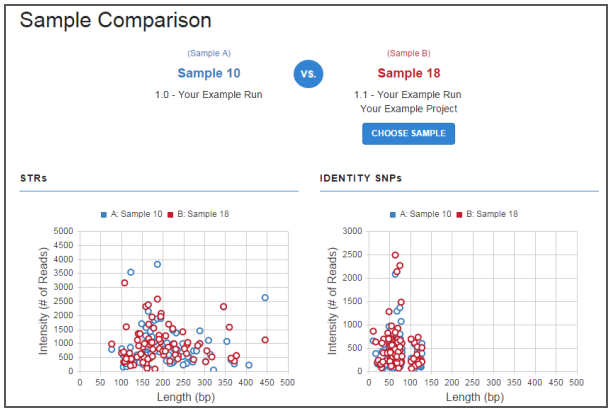
The results for sample comparison are color coded. Blue is for sample A, and red is for sample B. Sample comparison results are illustrated in three ways:

- Length and Intensity Graph
- Venn Diagram of Typed Loci Available for Comparison
- Table of Discordance

Length and Intensity Graph

Sample intensity (number of reads) and length (bp) information for typed STRs and iSNPs are graphed separately, with STRs on the left, and iSNPs on the right. This graph behaves similarly to the length and intensity graphs on the Sample Details page. For more information, see *Intensity and Length Graphs* on page 35. On the Sample Comparison page, the comparison samples are overlaid on top of each other.

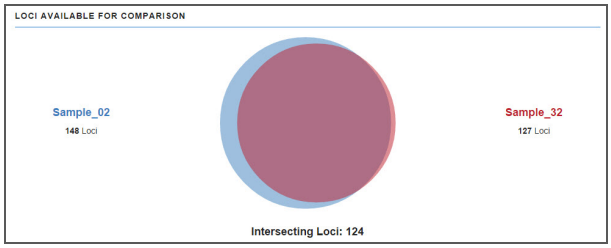
Figure 51 Length and Intensity Graph



Venn Diagram of Typed Loci Available for Comparison

A Venn diagram shows the intersection of typed loci from each sample. The total number of typed loci in each sample is represented by a circle. On each side of the diagram is the name of the sample and the total number of typed STR loci and typed iSNPs.

Figure 52 Venn Diagram of Typed Loci



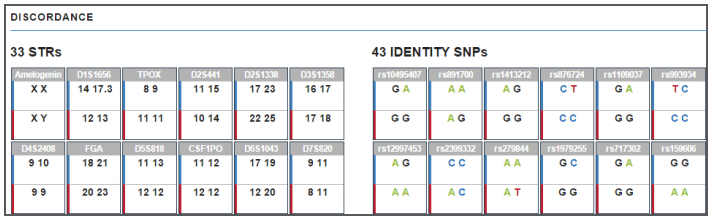
The number of intersecting loci is displayed below the Venn diagram circles. These are the loci that are called in both samples, and are visually represented by the overlap of the sample circles.

Table of Discordance

If there are typed loci that do not have the same genotype in both samples, they are displayed in a table showing discordance. Selecting a discordant locus presents the results

for both samples at the locus for further comparison.

Figure 53 Table of Discordance

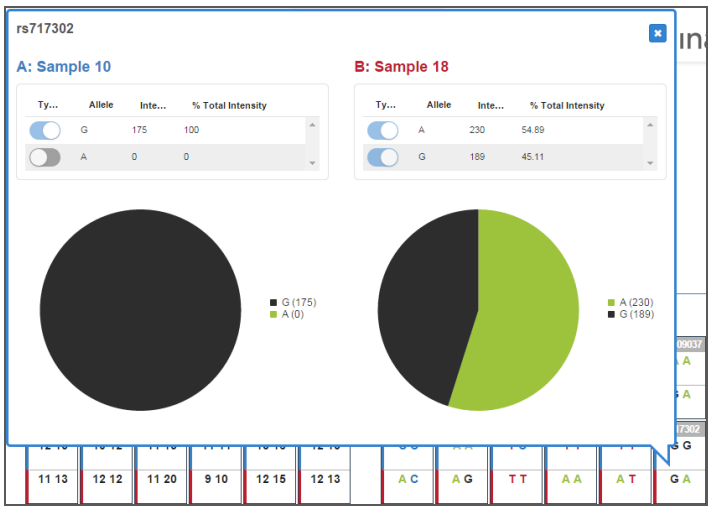


Tables in the STR detail boxes can be sorted by allele and intensity by clicking their respective column headings.

Figure 54 STR Comparison Details

Tables in the SNP detail boxes can be sorted by allele, intensity, and % total intensity by clicking their respective column headings.

Figure 55 SNP Comparison Details



Generating Sample Compare Population Statistics

After a sample comparison is complete, you can generate population statistics from the Sample Compare page for the intersection loci typed in Sample A. All of the loci that overlap between the samples are used in the calculation, regardless of discordance. To generate population statistics for a subset of loci, update and modify locus typing on the Sample Details page. See *Updating and Modifying Typed Results* on page 43.

NOTE
To enable generating a comparison, the software requires that each genotype reflects the locus copy number.

1. When a Sample Comparison is complete, scroll to the bottom of the Sample Comparison results page to the Population Statistics section and select the blue **Generate Population Statistics** button.
2. Select Population Groups to include in population statistics calculation.
3. Click **Generate**.
4. A population statistics progress bar indicates how much time is left until calculations are complete.
5. Population statistics results are displayed in horizontal bars.

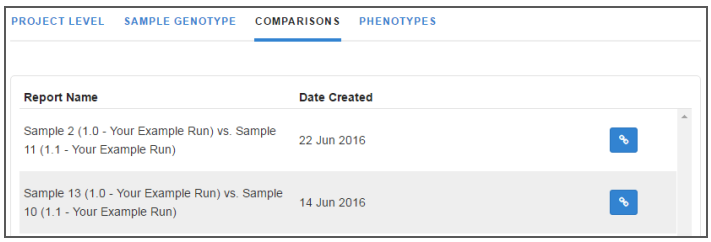
NOTE
Population statistics generation can take several minutes. While waiting for population statistics to process, you can navigate to other pages, and perform other functions in the software.

To view population statistics results later, return to the Project page and click the Comparisons tab.

Comparisons Tab

The Comparisons tab lists all sample comparisons performed on samples in a project. A blue link button connects to sample comparison results, which are identical to the results that open when a sample comparison is complete. See *Sample Comparison Results* on page 54 and *Generating Sample Compare Population Statistics* on page 56.

Figure 56 Comparisons Tab



1 On the Project page, select the **Comparisons** tab. See *Viewing the Project Page* on page 21.

2 Click the blue **Link** button in the list to display sample comparison results for a sample in the project.

NOTE
If you performed a sample comparison in two different projects, access the sample comparison report from the Project page of sample A.

Phenotype and Biogeographical Ancestry Estimation

Chapter 5

Introduction

The phenotype estimation feature analyzes pSNPs and aSNPs to display genotype results for SNPs that are indicative of hair color, eye color, and biogeographical ancestry relative to major population groups. The use of this feature is optional.

Phenotype and Biogeographical Ancestry Estimation

Probabilities for hair and eye color estimation are obtained from the HIrisPlex model, a multinomial logistic regression model.[1,2,3] Biogeographical ancestry estimation is obtained by principal component analysis (PCA). The model was trained on the European, East Asian, and African (excepting ASW) super populations of the 1000 Genomes data from Phase I of the project.[4] The unknown sample is projected along with the Ad-Mixed Americans super population, for context, onto the pre-trained first two components based on its aSNP genotype calls.

NOTE
For information on 1000 Genomes populations, see 1000genomes.org.

Resources

1. Walsh S, Chaitanya L, Clarisse L, et al. Developmental validation of the HIrisPlex system: DNA-based eye and hair colour prediction for forensic and anthropological usage. *Forensic Sci Int Genet.* 2014 Mar;9:150–161.

2. Liu F, van Duijn K, Vingerling JR, et al. Eye color and the prediction of complex phenotypes from genotypes. *Curr Biol.* 2009 Mar 10;19(5):R192–193.

3. Walsh S, Lui F, Ballantyne KN, van Oven M, Lao O, Kayser M. IrisPlex: A sensitive DNA tool for accurate prediction of blue and brown eye colour in the absence of ancestry information. *Forensic Sci Int Genet.* 2011 June;5(3):170–180.

4. Abecasis GR, Altshuler D, Auton A, et al. A map of human genome variation from population-scale sequencing. *Nature.* 2010 October;28;467(7319):1061–1073.

Generating pSNP and aSNP Information

Perform the following steps to generate phenotype estimation SNP information:

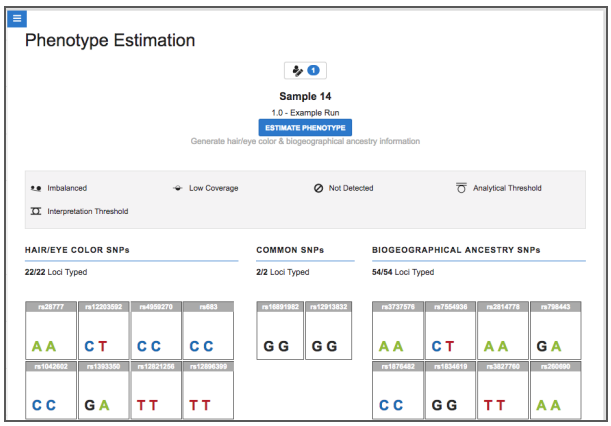
1. From the Project page, locate an analysis on the left pane, and click the blue arrow to open a list of samples. See *Viewing the Project Page* on page 21.
2. Select a sample from the project.
3. Click the blue **Actions** drop-down list button.
4. Select **Phenotype Estimation**.
5. The Phenotype Estimation table opens.

NOTE
The phenotype estimation feature is available for samples analyzed with ForenSeq DNA Primer Mix B.

Phenotype Estimation SNP Sample Details Table

The Phenotype Estimation SNP Sample Details table shows data for pSNPs and aSNPs. To generate a phenotype estimation from the SNP information, see *Generating Phenotype Estimations* on page 65.

Figure 57 Phenotype Estimation SNP Sample Details



Three columns of SNPs are displayed in the Phenotype Estimation sample details table:

- Hair and eye color (pSNPs)
- SNPs for estimating hair color, eye color, and biogeographical ancestry, listed as Common SNPs
- Biogeographical ancestry (aSNPs)

NOTE
Genotype results for hair, eye color, and common SNPs are required to produce a result for hair and eye color estimation. If 1 of the SNPs is not typed, no result is generated. In contrast, only 1 biogeographical ancestry SNP is required to be typed to produce results for biogeographical ancestry. The absence of SNPs in the biogeographical ancestry estimation diminishes the accuracy of the estimation. The biogeographical ancestry estimation model is retrained on each execution using only the aSNPs with a multilocus genotype.

The following information is displayed above sample locus detail tables:

- **Loci Typed**— The number of pSNP and aSNP target loci identified in the run of all possible target loci in the ForenSeq DNA Primer Mix B multiplex.

Each box in the Phenotype Estimation sample details table has a color border that indicates whether the locus is within or outside of quality metrics guidelines. If applicable, quality control indicator icons are displayed in the box. Click on a locus box to display genotyping table and circle plot. For information on the table and pie charts in the SNP locus detail table, see *SNP Sample Details Table* on page 40.

Table 26 Locus Typed Box Color Indicators

Color	Indication
Gray	Quality control indicators are not active for any locus.
Orange	One or more quality control indicators are active for the locus.

Table 27 Sample Details Quality Control Indicator Icons

Symbol	Symbol Name	Indication
	Imbalanced	Allele balance percentage falls below the defined ratio for intralocus balance in the analysis settings. For more information on imbalance, see *SNP Genotype Calling* on page 129.
	Low Coverage	Signal above the interpretation threshold was not detected as defined in the analysis settings. The low coverage indicator can be seen in results from version 1.1 or earlier of the ForenSeq Universal Analysis Software. See *SNP Genotype Calling* on page 129.
	Interpretation Threshold	Indicates that the locus has at least 1 typed allele, and at least one allele is above the analytical threshold but below the interpretation threshold.
	User Actions	At least one allele was manually edited as typed or not typed, or a user comment was submitted. See *Updating and Modifying Typed Results* on page 43.
	Analytical Threshold	Indicates that the locus has signal below the analysis threshold, and no alleles above the interpretation threshold.
	Not Detected	No signal was detected for the locus.

Generating Phenotype Estimations

Perform the following steps to generate a phenotype estimation:

1 From the Phenotype Estimation page, click the blue **Estimate Phenotype** button. For information on how to access the Phenotype Estimation page, see *Phenotype and Biogeographical Ancestry Estimation* on page 61.

2 A phenotype estimation progress bar indicates how much time is left until the estimation is complete.

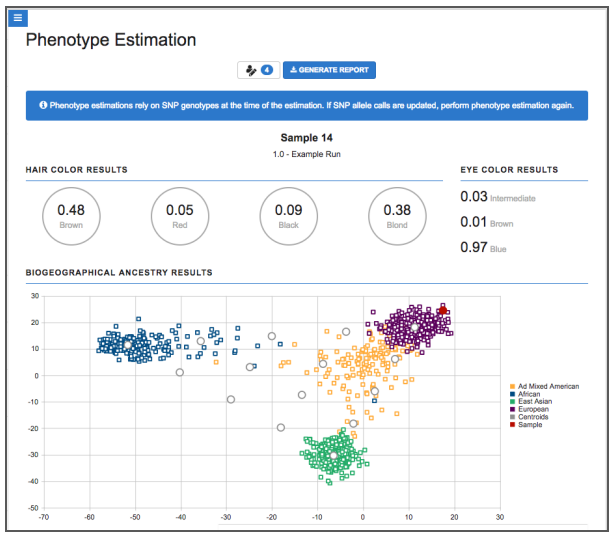
3 The Phenotype Estimation chart opens.

NOTE
If SNP allele calls are updated after you generate a phenotype estimation, perform phenotype estimation again.

- **Hair Color Results**—Hair phenotype percentages based in the HIrisPlex model described by Walsh et al.[1]
- **Eye Color Results**—Eye color phenotype percentages based in the HIrisPlex model described by Walsh et al.
- **Biogeographical Ancestry Results**—Graph of the result from the principal component analysis of the sample data set. Biogeographical ancestry estimation is obtained by principal component analysis (PCA). The model was trained on the European, East Asian, and African (excepting ASW) super populations of the 1000 Genomes data from Phase I of the project.[2] The unknown sample is projected along with the Ad-Mixed Americans super population, for context, onto the pre-trained first two components based on its aSNP genotype calls.

NOTE
For information on 1000 Genomes populations, see 1000genomes.org.

Figure 58 Phenotype Estimation



Centroids give perspective to the results of the estimation and provide logical groupings of populations. The centroids are represented by the gray centers of the clusters. There are three centroids for three major ancestries (AFR, ASN, EUR in the middle of the blue, purple, and green clusters), and then orientational centroids at the one-quarter intervals between these three large groups.

- **Sample**— Displayed on the chart in red. Your unknown sample.

- **Distance to Nearest Centroid**— Proximity of the sample to the nearest centroid. This measurement gives an indication of how related the sample is to the general grouping for the centroid. For comparison, the distance is provided for the 1000 Genome samples contributing to the centroid.
- **1000 Genomes populations with samples in centroid with sample**— Click the black arrow to access drop-down list below the graph.
- **Reference samples in centroid with sample**— Click the black arrow to access the drop-down list below the graph.

Resources

1. Walsh S, Chaitanya L, Clarisse L, et al. Developmental validation of the HIrisPlex system: DNA-based eye and hair colour prediction for forensic and anthropological usage. *Forensic Sci Int Genet.* 2014 Mar;9:150-161.

2. Abecasis GR, Altshuler D, Auton A, et al. A map of human genome variation from population-scale sequencing. *Nature.* 2010 October;28;467(7319):1061–1073.

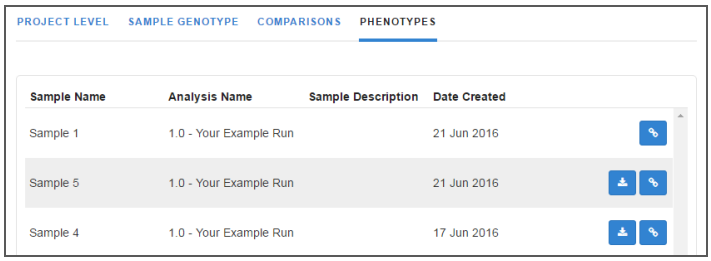
Phenotypes Tab

The Phenotypes tab is a list of all phenotype estimations performed on samples in a project. See *Phenotype and Biogeographical Ancestry Estimation* on page 61. A blue link button connects to phenotype estimation results, which are identical to the results that open when a phenotype estimation is complete. See *Generating Phenotype Estimations* on page 65.

1 On the Project page, select the **Phenotypes** tab. See *Viewing the Project Page* on page 21.

2 Click the blue **Link** button in the list to display phenotype estimation results for a sample in the project.

Figure 59 Phenotypes Tab

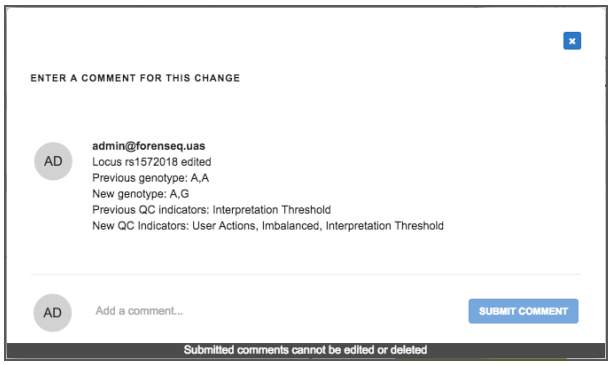


Sample History: Phenotype Estimation

The Sample History feature tracks phenotype estimation user actions and system-initiated events. System events are displayed by adjusting the toggle in the upper-right corner of the dialog window. With visibility of the Sample History feature enabled, the Phenotype Activity dialog of logged events is visible by clicking on the user actions icon at the top of the sample phenotype estimation page. Phenotype estimation Locus Activity is visible by clicking on the user actions icon at the top of the locus details box. Phenotype estimation Sample History is recorded in the phenotype estimation report (Phenotype History tab) and is not included in the sample-level genotype reports.

Figure 60 Phenotype Activity

Figure 61 Locus Activity



NOTE
Phenotype estimation user comments can be entered in the Locus Activity dialog.

NOTE
After a comment is submitted, it is permanent and cannot be modified or deleted.

Reports

Chapter 6

Introduction

The ForenSeq Universal Analysis Software features the ability to create and view the following reports:

- Project-level genotype report
- Project-level flanking region report
- Sample-level sample details genotype report
- Sample-level sample summary genotype report
- Sample-level phenotype estimation report
- Sample-level CODIS report
- Project-level CODIS report

After report generation, links to analyses and reports are available in separate tabs on the Project page for later reference.

To create a project-level genotype report or project-level CODIS report, click **Generate Report** on the Project page and select Project Details or CODIS from the drop-down list. Note that a project-level flanking region report is generated automatically when using an analysis settings template that specifies creation of the report. For information about the analysis settings template or flanking region report, see *Changing Locus Thresholds* on page 101 or *Project-Level Flanking Region Report* on page 75.

You can create sample-level reports from the following locations:

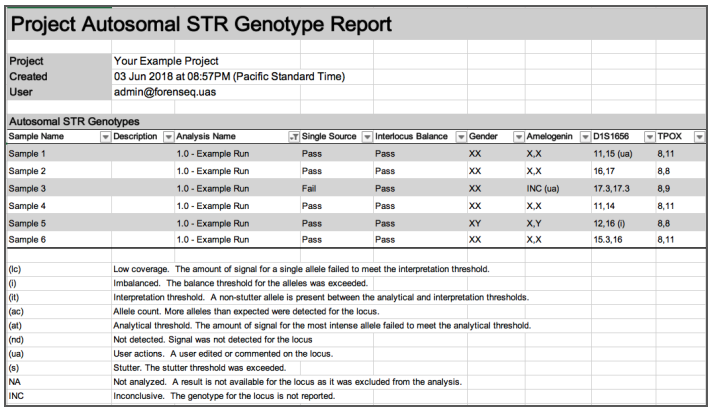
- The sample-level sample details genotype report can be created from the Project page or the Sample Details page.
- The sample-level sample summary genotype report can be created from the Project page or the Sample Details page.
- The sample-level phenotype estimation report can be created from the Phenotype Estimation results page.
- The sample-level CODIS report can be created from the Sample Details page.

Project-Level Genotype Report

The project-level genotype report provides results and supporting information for the STR loci and iSNPs across all samples and analyses in a project. The report is a Microsoft Excel workbook file that you can download and customize by using sorting filters and drop-down list. After project-level reports are created, they are available for download from the Project page.

The project-level genotype report workbook lists every sample name and its description, and includes single-source, interlocus balance, and gender results. The report shows the typed alleles for each locus of each sample of each analysis in a project. The worksheets with coverage information present the read counts and alleles for each typed allele from every sample in each analysis that is in a project.

Figure 62 Project-Level Genotype Report



Project Autosomal STR Genotype Report

Project	Your Example Project
Created	03 Jun 2018 at 08:57PM (Pacific Standard Time)
User	admin@forenseq.uas

Autosomal STR Genotypes

Sample Name	Description	Analysis Name	Single Source	Interlocus Balance	Gender	Amelogenin	D1S1656	TPOX
Sample 1		1.0 - Example Run	Pass	Pass	XX	X,X	11,15 (ua)	8,11
Sample 2		1.0 - Example Run	Pass	Pass	XX	X,X	16,17	8,8
Sample 3		1.0 - Example Run	Fail	Pass	XX	INC (ua)	17.3,17.3	8,9
Sample 4		1.0 - Example Run	Pass	Pass	XX	X,X	11,14	8,11
Sample 5		1.0 - Example Run	Pass	Pass	XY	X,Y	12,16 (i)	8,8
Sample 6		1.0 - Example Run	Pass	Pass	XX	X,X	15.3,16	8,11

(lc)	Low coverage. The amount of signal for a single allele failed to meet the interpretation threshold.
(i)	Imbalanced. The balance threshold for the alleles was exceeded.
(it)	Interpretation threshold. A non-stutter allele is present between the analytical and interpretation thresholds.
(ac)	Allele count. More alleles than expected were detected for the locus.
(at)	Analytical threshold. The amount of signal for the most intense allele failed to meet the analytical threshold.
(nd)	Not detected. Signal was not detected for the locus
(ua)	User actions. A user edited or commented on the locus.
(s)	Stutter. The stutter threshold was exceeded.
NA	Not analyzed. A result is not available for the locus as it was excluded from the analysis.
INC	Inconclusive. The genotype for the locus is not reported.

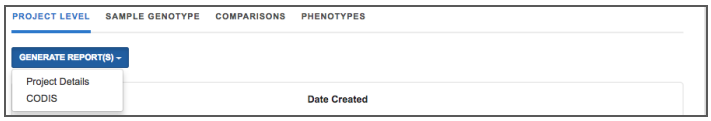
If multiple analyses of the samples in the project exist, the order of the samples in the workbook is arranged by analysis name, with the most recent analysis appearing at the top of the list.

The project-level genotype report workbook is subdivided into eight worksheets:

- Autosomal STRs
- Autosomal STR coverage
- Y-STRs
- Y-STR coverage
- X-STRS
- X-STR coverage
- iSNPs
- iSNP coverage

1 On the Project page, click **Generate Report** and select Project Details from the drop-down list. See *Viewing the Project Page* on page 21.

Figure 63 Generate Report Command



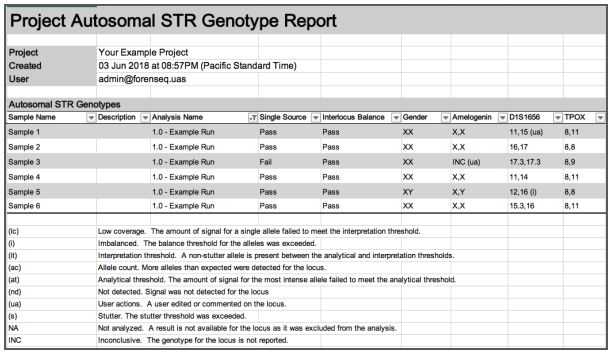
On the right side of the report list, a Pending button appears for each report generated. When the report is complete, the Pending button becomes a Download button.

2 Click **Download** to access the report.

Autosomal STR Worksheets for Project-Level Genotype Reports

The Autosomal STR worksheets in the project-level genotype report feature a genotype report and a coverage report. The reports show locus genotypes, quality control indicators, number of reads, and for STRs the nucleotide sequence of typed alleles. STRs with isometric alleles have sequence differences highlighted with a different font size.

Figure 64 Project-Level Genotype Report



Project Autosomal STR Genotype Report

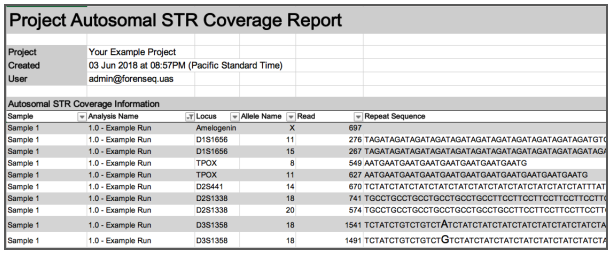
Project	Your Example Project
Created	03 Jun 2018 at 08:57PM (Pacific Standard Time)
User	admin@forenseq.uas

Autosomal STR Genotypes

Sample Name	Description	Analysis Name	Single Source	Interlocus Balance	Gender	Amelogenin	D1S1656	TPOX
Sample 1		1.0 - Example Run	Pass	Pass	XX	X,X	11,15 (ua)	8,11
Sample 2		1.0 - Example Run	Pass	Pass	XX	X,X	16,17	8,8
Sample 3		1.0 - Example Run	Fail	Pass	XX	INC (ua)	17.3,17.3	8,9
Sample 4		1.0 - Example Run	Pass	Pass	XX	X,X	11,14	8,11
Sample 5		1.0 - Example Run	Pass	Pass	XY	X,Y	12,16 (i)	8,8
Sample 6		1.0 - Example Run	Pass	Pass	XX	X,X	15.3,16	8,11

(lc)	Low coverage. The amount of signal for a single allele failed to meet the interpretation threshold.
(i)	Imbalanced. The balance threshold for the alleles was exceeded.
(it)	Interpretation threshold. A non-stutter allele is present between the analytical and interpretation thresholds.
(ac)	Allele count. More alleles than expected were detected for the locus.
(at)	Analytical threshold. The amount of signal for the most intense allele failed to meet the analytical threshold.
(nd)	Not detected. Signal was not detected for the locus
(ua)	User actions. A user edited or commented on the locus.
(s)	Stutter. The stutter threshold was exceeded.
NA	Not analyzed. A result is not available for the locus as it was excluded from the analysis.
INC	Inconclusive. The genotype for the locus is not reported.

Figure 65 Autosomal STR Coverage Report



Project Autosomal STR Coverage Report

Project	Your Example Project
Created	03 Jun 2018 at 08:57PM (Pacific Standard Time)
User	admin@forenseq.uas

Autosomal STR Coverage Information

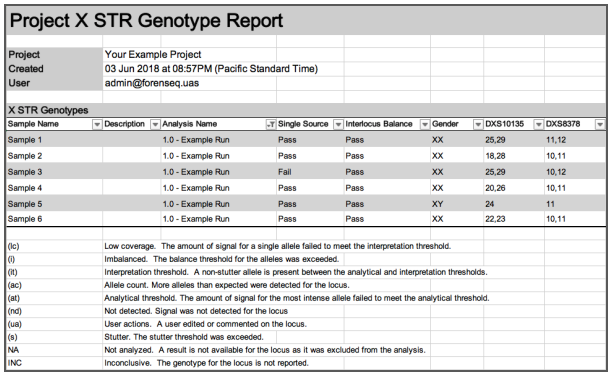
Sample	Analysis Name	Locus	Allele Name	Read	Repeat Sequence
Sample 1	1.0 - Example Run	Amelogenin	X	597	
Sample 1	1.0 - Example Run	D1S1656	11	276	TAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATGTC
Sample 1	1.0 - Example Run	D1S1656	15	267	TAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGA
Sample 1	1.0 - Example Run	TPOX	8	549	AATGAATGAATGAATGAATGAATGAATG
Sample 1	1.0 - Example Run	TPOX	11	627	AATGAATGAATGAATGAATGAATGAATGAATGAATGAATG
Sample 1	1.0 - Example Run	D2S441	14	670	TCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATTTAT
Sample 1	1.0 - Example Run	D2S1338	18	741	TGCCTGCCTGCCTGCCTGCCTTCCTTCCTTCCTTCCTTCCTTC
Sample 1	1.0 - Example Run	D2S1338	20	574	TGCCTGCCTGCCTGCCTGCCTGCCTTCCTTCCTTCCTTCCTT
Sample 1	1.0 - Example Run	D3S1358	18	1541	TCTATCTGTCTGTCTATCTATCTATCTATCTATCTATCTATCTATCTA
Sample 1	1.0 - Example Run	D3S1358	18	1491	TCTATCTGTCTGTCTGTCTATCTATCTATCTATCTATCTATCTATCTA

For information on the meaning of column headings and quality control indicators in the body of the worksheet, see *General Locus Information* on page 35 and *STR Sample Details Table* on page 37.

Y-STR and X-STR Worksheet for Project-Level Genotype Report

The Y-STR and X-STR worksheet in the project-level genotype report feature a genotype report and a coverage report. The reports show locus genotypes, quality control indicators, number of reads, and the nucleotide sequence of typed alleles for all samples and all analyses in a project. STRs with isometric alleles have sequence differences highlighted with a different font size. For information on the meaning of column headings and quality control indicators in the body of the worksheet, see *General Locus Information* on page 35 and *STR Sample Details Table* on page 37.

Figure 66 X STR Genotype Report



Project X STR Genotype Report

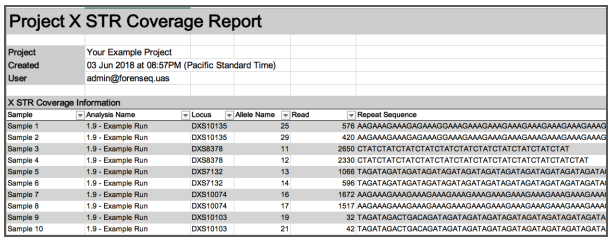
Project	Your Example Project
Created	03 Jun 2018 at 08:57PM (Pacific Standard Time)
User	admin@forenseq.uas

X STR Genotypes

Sample Name	Description	Analysis Name	Single Source	Interlocus Balance	Gender	DXS10135	DXS8378
Sample 1		1.0 - Example Run	Pass	Pass	XX	25,29	11,12
Sample 2		1.0 - Example Run	Pass	Pass	XX	18,28	10,11
Sample 3		1.0 - Example Run	Fail	Pass	XX	25,29	10,12
Sample 4		1.0 - Example Run	Pass	Pass	XX	20,26	10,11
Sample 5		1.0 - Example Run	Pass	Pass	XY	24	11
Sample 6		1.0 - Example Run	Pass	Pass	XX	22,23	10,11

(lc)	Low coverage. The amount of signal for a single allele failed to meet the interpretation threshold.
(i)	Imbalanced. The balance threshold for the alleles was exceeded.
(it)	Interpretation threshold. A non-stutter allele is present between the analytical and interpretation thresholds.
(ac)	Allele count. More alleles than expected were detected for the locus.
(at)	Analytical threshold. The amount of signal for the most intense allele failed to meet the analytical threshold.
(nd)	Not detected. Signal was not detected for the locus
(ua)	User actions. A user edited or commented on the locus.
(s)	Stutter. The stutter threshold was exceeded.
NA	Not analyzed. A result is not available for the locus as it was excluded from the analysis.
INC	Inconclusive. The genotype for the locus is not reported.

Figure 67 X STR Coverage Report



Project X STR Coverage Report

Project	Your Example Project
Created	03 Jun 2018 at 08:57PM (Pacific Standard Time)
User	admin@forenseq.uas

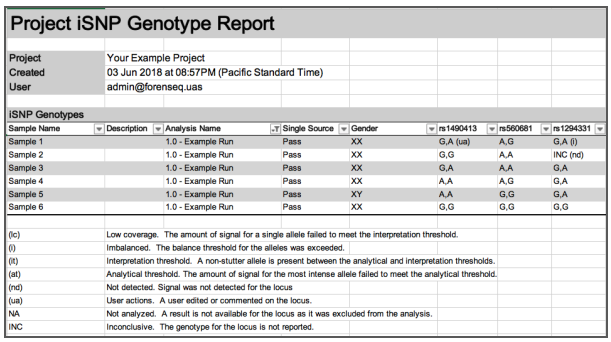
X STR Coverage Information

Sample	Analysis Name	Locus	Allele Name	Read	Repeat Sequence
Sample 1	1.9 - Example Run	DXS10135	25	576	AAGAAAGAAAGAAAGGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG
Sample 2	1.9 - Example Run	DXS10135	29	420	AAGAAAGAAAGAAAGGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG
Sample 3	1.9 - Example Run	DXS8378	11	2650	CTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTAT
Sample 4	1.9 - Example Run	DXS8378	12	2330	CTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTAT
Sample 5	1.9 - Example Run	DXS7132	13	1066	TAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATA
Sample 6	1.9 - Example Run	DXS7132	14	596	TAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATA
Sample 7	1.9 - Example Run	DXS10074	16	1672	AAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAA
Sample 8	1.9 - Example Run	DXS10074	17	1517	AAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAA
Sample 9	1.9 - Example Run	DXS10103	19	32	TAGATAGACTGACAGATAGATAGATAGATAGATAGATAGATAGATAGATA
Sample 10	1.9 - Example Run	DXS10103	21	42	TAGATAGACTGACAGATAGATAGATAGATAGATAGATAGATAGATAGATA

iSNP Worksheet for Project-Level Genotype Report

The iSNP worksheet in the project-level genotype report feature a genotype report and a coverage report. The reports show genotypes, quality control indicators, and the number of reads for typed alleles. For information on the meaning of column headings in the body of the worksheet, see *General Locus Information* on page 35 and *SNP Sample Details Table* on page 40.

Figure 68 iSNP Genotype Report



Project iSNP Genotype Report

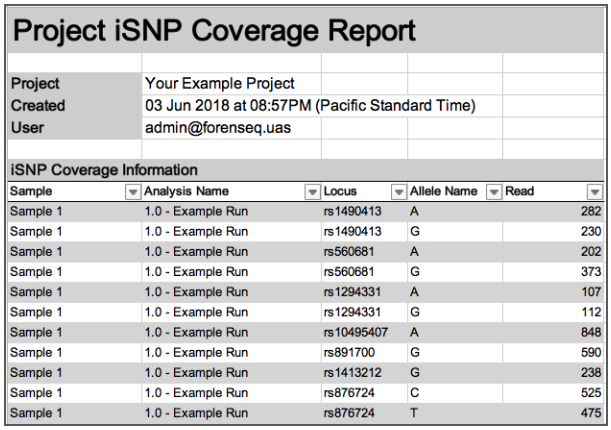
Project	Your Example Project
Created	03 Jun 2018 at 08:57PM (Pacific Standard Time)
User	admin@forenseq.uas

iSNP Genotypes

Sample Name	Description	Analysis Name	Single Source	Gender	rs1490413	rs560681	rs1294331
Sample 1		1.0 - Example Run	Pass	XX	G,A (ua)	A,G	G,A (i)
Sample 2		1.0 - Example Run	Pass	XX	G,G	A,A	INC (nd)
Sample 3		1.0 - Example Run	Pass	XX	G,A	A,A	G,A
Sample 4		1.0 - Example Run	Pass	XX	A,A	A,G	G,A
Sample 5		1.0 - Example Run	Pass	XY	A,A	G,G	G,A
Sample 6		1.0 - Example Run	Pass	XX	G,G	G,G	G,G

(lc)	Low coverage. The amount of signal for a single allele failed to meet the interpretation threshold.
(i)	Imbalanced. The balance threshold for the alleles was exceeded.
(it)	Interpretation threshold. A non-stutter allele is present between the analytical and interpretation thresholds.
(at)	Analytical threshold. The amount of signal for the most intense allele failed to meet the analytical threshold.
(nd)	Not detected. Signal was not detected for the locus
(ua)	User actions. A user edited or commented on the locus.
NA	Not analyzed. A result is not available for the locus as it was excluded from the analysis.
INC	Inconclusive. The genotype for the locus is not reported.

Figure 69 iSNP Coverage Report



Project iSNP Coverage Report

Project	Your Example Project
Created	03 Jun 2018 at 08:57PM (Pacific Standard Time)
User	admin@forenseq.uas

iSNP Coverage Information

Sample	Analysis Name	Locus	Allele Name	Read
Sample 1	1.0 - Example Run	rs1490413	A	282
Sample 1	1.0 - Example Run	rs1490413	G	230
Sample 1	1.0 - Example Run	rs560681	A	202
Sample 1	1.0 - Example Run	rs560681	G	373
Sample 1	1.0 - Example Run	rs1294331	A	107
Sample 1	1.0 - Example Run	rs1294331	G	112
Sample 1	1.0 - Example Run	rs10495407	A	848
Sample 1	1.0 - Example Run	rs891700	G	590
Sample 1	1.0 - Example Run	rs1413212	G	238
Sample 1	1.0 - Example Run	rs876724	C	525
Sample 1	1.0 - Example Run	rs876724	T	475

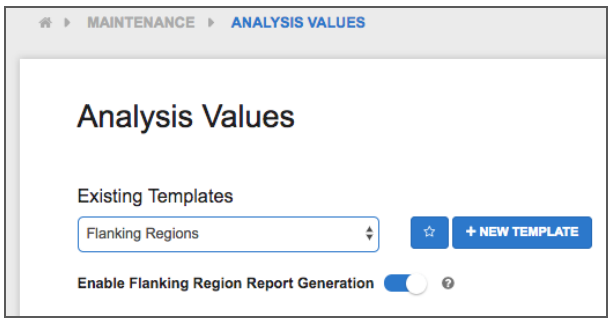
Project-Level Flanking Region Report

The project-level flanking region report provides flanking region sequence information for ForenSeq STR and iSNP amplicons outside of the ForenSeq Universal Analysis Software user interface. The project-level flanking region report contains every flanking region-enabled analysis within the project as a Microsoft Excel workbook. Analyses that did not have flanking region report generation enabled are not present in the report.

The sequence data provided in the project-level flanking region report include the nucleotides between the primers for each amplicon. This additional data may indicate potential variants unavailable with traditional genotyping technology. For each reported amplicon, the strand of the read is provided to facilitate alignment or comparison to a reference genome.

When specified in the analysis settings template, a project-level flanking region report is generated automatically at the conclusion of analysis. After project-level flanking region reports are created, they are available for download from the Project page. See *Viewing the Project Page* on page 21.

Figure 70 Flanking Region Analysis Template



When a flanking region report-enabled analysis is processing, a Pending button appears on the right side of the report list for the project-level flanking region report that is being generated during analysis. When the analysis and report are complete, the Pending button becomes a Download button. Click **Download** to access the report.

The Project Level Flanking Region report workbook is subdivided into five worksheets:

- Autosomal STR coverage
- X-STR coverage
- Y-STR coverage
- iSNP coverage
- Text format (*.txt)

NOTE
Project-level flanking region report generation requires an analysis that uses a flanking region report-enabled analysis template. To enable flanking region report generation for an analysis template, an administrator user can select the desired template in the Analysis Values section of the Maintenance page and toggle on (blue position) **Enable Flanking Region Report Generation**. See *Defining Content Within the Application* on page 105.

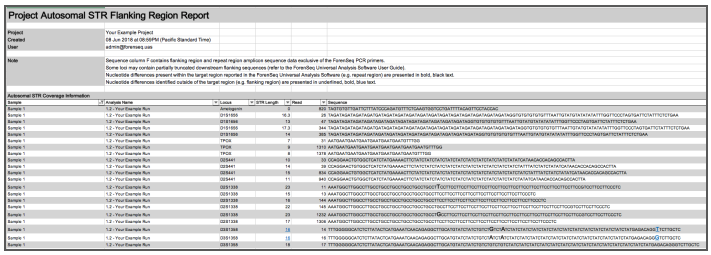
Autosomal STR Worksheet for Project-Level Flanking Region Report

The Autosomal STR worksheet in the project-level flanking region report contains amplicon sequence data exclusive of the ForenSeq PCR primers for the analyzed autosomal STR loci of each sample. The worksheet provides the sample name, analysis name, locus name, STR length, number of reads supporting the sequence, amplicon sequence, and strand direction for every unique sequence of the reported loci.

Nucleotide differences detected between two or more sequences of the same length at the same locus are displayed with the following designations:

- Nucleotide differences present inside of the region reported in the ForenSeq Universal Analysis Software are presented in **bold**, black text.
- Nucleotide differences present outside of the region reported in the ForenSeq Universal Analysis Software are presented in underlined, **bold**, blue text. Sequences uniquely identified by these positions are denoted in the STR Length column with underlined, blue text.

Figure 71 Autosomal STR Flanking Region Report



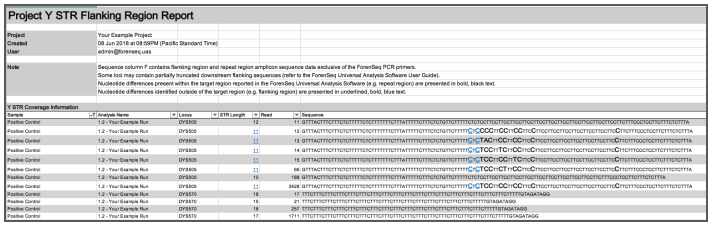
Y-STR and X-STR Worksheets for Project-Level Flanking Region Report

The Y-STR and X-STR worksheets in the project-level flanking region report contains amplicon sequence data exclusive of the ForenSeq PCR primers for the analyzed Y-STR and X-STR loci of each sample. The worksheets provide the sample name, analysis name, locus name, STR length, number of reads supporting the sequence, amplicon sequence, and strand direction for every unique sequence of the reported loci.

Nucleotide differences detected between two or more sequences of the same length at the same locus are displayed with the following designations:

- Nucleotide differences present inside of the region reported in the ForenSeq Universal Analysis Software are presented in **bold**, black text.
- Nucleotide differences present outside of the region reported in the ForenSeq Universal Analysis Software are presented in underlined, **bold**, blue text. Sequences uniquely identified by these positions are denoted in the STR Length column with underlined, blue text.

Figure 72 Y STR Flanking Region Report



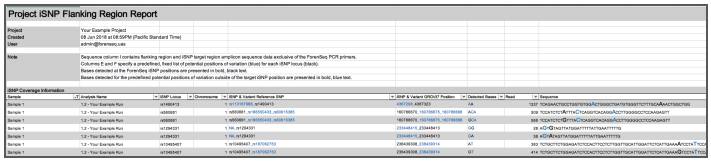
iSNP Worksheet for Project-Level Flanking Region Report

The iSNP worksheet in the project-level flanking region report contains amplicon sequence data exclusive of the ForenSeq PCR primers for the analyzed iSNP loci of each sample. The worksheet provides the sample name, analysis name, iSNP locus name, chromosome, predefined positions of potential variation[1,2] with reference SNP identifier and coordinate, detected bases, number of reads supporting the sequence, amplicon sequence, and strand direction for every unique sequence of the reported loci.

Previously identified positions of potential variation[1,2] are specifically interrogated and the detected bases at each position reported. Nucleotide differences detected between two or more sequences from the same amplicon are displayed with the following designations:

- Nucleotide differences present at the ForenSeq DNA Signature Prep Kit positions are presented in **bold**, black text. These results are presented in the iSNP & Variant Reference SNP, iSNP & Variant GRCh37 Position, and Detected Bases columns in black text.
- Nucleotide differences present at the positions of potential variation are presented in **bold**, blue text. These results are presented in the iSNP & Variant Reference SNP, iSNP & Variant GRCh37 Position, and Detected Bases columns in blue text.
- Nucleotide differences present at other positions in the amplicon may exist and are reported within the sequence. These positions are not presented with special formatting or identified in other report columns.

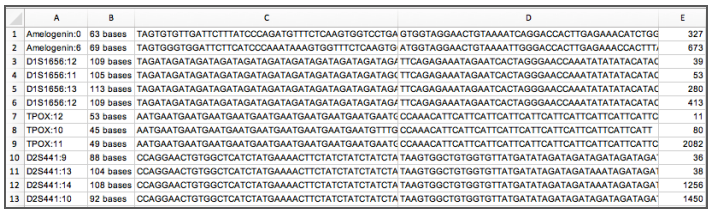
Figure 73 iSNP Flanking Region Report



Text Format Worksheet for Project-Level Flanking Region Report

The text format (*.txt) worksheet in the project-level flanking region report is designed for import into a system that supports the format. The worksheet provides the STR and iSNP loci names, STR length, total reported length in bases, forward read sequence, reverse read sequence, and the total read count for each sequence.

Figure 74 Text Format Tab of Flanking Region Report



	A	B	C	D	E
1	Amelogenin:0	63 bases	TAGTGTGTTGATTCTTTATCCCAGATGTTTCTCAAGTGGTCCTGA	GTGGTAGGAACTGTAAAATCAGGACCACTTGAGAAACATCTGG	327
2	Amelogenin:6	69 bases	TAGTGGGTGGATTCTTCATCCCAAATAAAGTGGTTTCTCAAGTG	ATGGTAGGAACTGTAAAATTGGGACCACTTGAGAAACCACTTT	673
3	D1S1656:12	109 bases	TAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAG	TTCAGAGAAATAGAATCACTAGGGAACCAAATATATATACATAC	39
4	D1S1656:11	105 bases	TAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAG	TTCAGAGAAATAGAATCACTAGGGAACCAAATATATATACATAC	53
5	D1S1656:13	113 bases	TAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAG	TTCAGAGAAATAGAATCACTAGGGAACCAAATATATATACATAC	280
6	D1S1656:12	109 bases	TAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAG	TTCAGAGAAATAGAATCACTAGGGAACCAAATATATATACATAC	413
7	TPOX:12	53 bases	AATGAATGAATGAATGAATGAATGAATGAATGAATGAATC	CCAAACATTCATTCATTCATTCATTCATTCATTCATTC	11
8	TPOX:10	45 bases	AATGAATGAATGAATGAATGAATGAATGAATGTTTG	CCAAACATTCATTCATTCATTCATTCATTCATTCATT	80
9	TPOX:11	49 bases	AATGAATGAATGAATGAATGAATGAATGAATGAATC	CCAAACATTCATTCATTCATTCATTCATTCATTCATTC	2082
10	D2S441:9	88 bases	CCAGGAACTGTGGCTCATCTATGAAAACTTCTATCTATCTATCTA	TAAGTGGCTGTGGTGTTATGATATAGATAGATAGATAGATAGA	36
11	D2S441:13	104 bases	CCAGGAACTGTGGCTCATCTATGAAAACTTCTATCTATCTATCTA	TAAGTGGCTGTGGTGTTATGATATAGATAGATAGATAAATAGATAGA	38
12	D2S441:14	108 bases	CCAGGAACTGTGGCTCATCTATGAAAACTTCTATCTATCTATCTA	TAAGTGGCTGTGGTGTTATGATATAGATAGATAAATAGATAGA	1256
13	D2S441:10	92 bases	CCAGGAACTGTGGCTCATCTATGAAAACTTCTATCTATCTATCTA	TAAGTGGCTGTGGTGTTATGATATAGATAGATAGATAGATAGA	1450

Resources

1. King J, Churchill J, Novroski N, Zeng X, Warshauer D, Seah L, Budowle B. Increasing the discrimination power of ancestry- and identity-informative SNP loci within the ForenSeq DNA Signature Prep Kit. *Forensic Sci Int Genet* (2018), https://doi.org/10.1016/j.fsigen.2018.06.005.

2. Single Nucleotide Polymorphism Database (dbSNP). https://www.ncbi.nlm.nih.gov/projects/SNP/.

Sample Genotype Report

The sample genotype report is composed of data and charts of results and supporting information for the STR and iSNP loci of a single sample. The report provides the same information as in the sample locus details table. See *STR Sample Details Table* on page 37 and *SNP Sample Details Table* on page 40.

The sample details and sample summary reports are a Microsoft Excel workbook file that can be download and customized by using sorting filters, and drop-down lists. After Sample Genotype reports are created, they are available for download from the Project page on the Sample Genotype tab.

The workbook is subdivided into 10 worksheets:

- Autosomal STRs
- Autosomal STR figure
- Y-STRs
- Y-STR figure
- X-STRs
- X-STR figure
- iSNPs
- iSNP figure
- Sample History
- Settings

1. On the Project page, select the **Sample Genotype** tab.
2. Click **Generate Report**.
3. Select **Sample Details** or **Sample Summary**.
   The Sample Details report includes typed and untyped alleles. The Sample Summary report includes typed alleles only.
4. In the Generate Reports box, select an analysis name from the drop-down list and select samples.
5. Click **Generate**.
   On the right side of the report list, a Pending button appears for each report generated. When the report is complete, the Pending button becomes a Download button.
6. Click **Download** to access the report.

NOTE
If Excel gives you a Protected View warning when you open a Sample Genotype report, click the Enable Editing button to view all the features of the workbook.

Sample Genotype reports for single samples can also be generated from the Sample Details page. Sample-level CODIS reports must be generated from the Sample Details page. For more information, see *Sample-Level CODIS Report* on page 87.

Autosomal STR Worksheets for Sample-Level Reports

The autosomal STR worksheets in the dample genotype report show locus genotypes, quality control indicators, number of reads, and for STRs the nucleotide sequence of typed alleles. STRs with isometric alleles have sequence differences highlighted with a different font size. For information on the meaning of column headings and quality control indicators in the body of the worksheet, see *General Locus Information* on page 35 and *STR Sample Details Table* on page 37.

Figure 75 Autosomal STR Sample Report Locus Information

Sample Autosomal STR Report

Sample	Sample 10	
Project	Your Example Project	
Analysis	1.0 - Example Run	
Run	Example Run	
Gender	XY	
Created	04 Jun 2018 at 11:35AM (Pacific Standard Time)	
User	admin@forenseq.uas	
Autosomal STR Locus Information		
Loci Typed	28 / 28	
Single Source	Pass	
Interlocus Balance	Pass	
Locus	**Genotype**	**QC Indicators**
Amelogenin	X,Y	
D1S1656	12,16	imbalanced
TPOX	8,8	
D2S441	10,11	
D2S1338	20,23	
D3S1358	14,17	
D4S2408	9,10	
FGA	20,20	
D5S818	12,13	

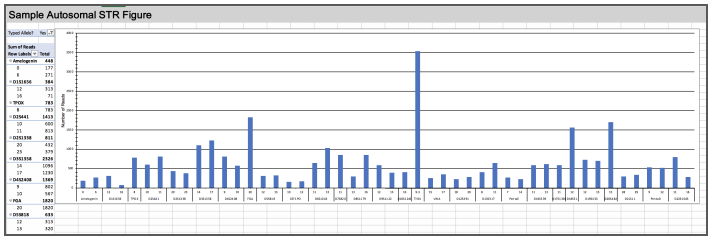
Figure 76 Autosomal STR Sample Report Coverage Information

Coverage Information

Locus	Allele Name	Typed Allele?	Reads	Repeat Sequence
Amelogenin	0	Yes	177	
Amelogenin	6	Yes	271	AAAGTG
D1S1656	12	Yes	313	TAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATA
D1S1656	16	Yes	71	TAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATA
TPOX	8	Yes	783	AATGAATGAATGAATGAATGAATGAATG
D2S441	10	Yes	600	TCTATCTATCTATCTATCTATCTATCTATCTATCTGTCTA
D2S441	11	Yes	813	TCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCT
D2S1338	20	Yes	432	TGCCTGCCTGCCTGCCTGCCTGCCTGCCTTCCTTCCTTCCTTCC
D2S1338	23	Yes	379	TGCCTGCCTGCCTGCCTGCCTGCCTTCCTTCCTTCCTTCC
D3S1358	14	Yes	1096	TCTATCTGTCTGTCTATCTATCTATCTATCTATCTATCTATC
D3S1358	17	Yes	1230	TCTATCTGTCTGTCTGTCTATCTATCTATCTATCTATCTATC
D4S2408	9	Yes	802	ATCTATCTATCTATCTATCTATCTATCTATCTATCT
D4S2408	10	Yes	567	ATCTATCTATCTATCTATCTATCTATCTATCTATCTATCT
FGA	20	Yes	1820	TTTCTTTCTTTCTTTTTTCTCTTTCTTTCTTTCTTTCTTTCTTT
D5S818	12	Yes	313	AGATAGATAGATAGATAGATAGATAGATAGATAGATAGATA
D5S818	13	Yes	320	AGATAGATAGATAGATAGATAGATAGATAGATAGATAGATA

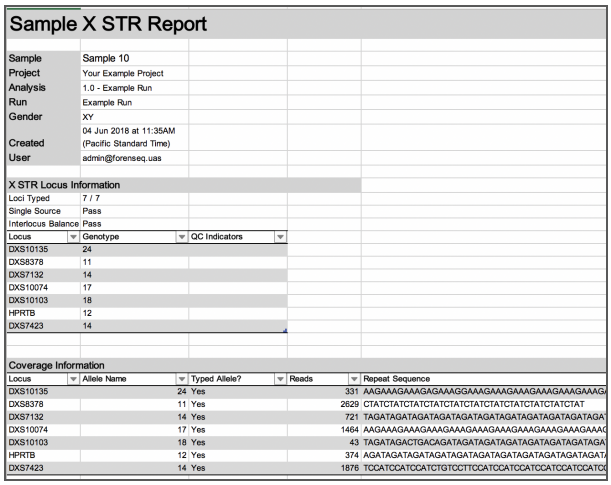
For STRs, the figure worksheet shows a pivot table and pivot chart. The data for the table and chart in the figure worksheet are populated from the autosomal STR worksheet. The pivot chart illustrates the typed alleles for all the autosomal STRs. The pivot table is a summation of the reads for the typed alleles at each locus. The Typed Allele column default setting shows the value of typed alleles (Typed Allele = Yes). However, the setting can be modified by clicking the filter icon to the right of the column heading.

Figure 77 Autosomal STR Figure Worksheet



Y-STR and X-STR Worksheets for Sample-Level Reports

Figure 78 X STR Sample Report



Sample X STR Report

Sample	Sample 10
Project	Your Example Project
Analysis	1.0 - Example Run
Run	Example Run
Gender	XY
Created	04 Jun 2018 at 11:35AM (Pacific Standard Time)
User	admin@forenseq.uas

X STR Locus Information

Loci Typed	7 / 7
Single Source	Pass
Interlocus Balance	Pass

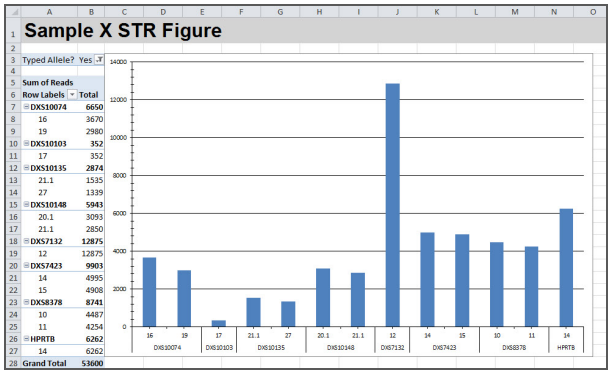
Locus	Genotype	QC Indicators
DXS10135	24	
DXS8378	11	
DXS7132	14	
DXS10074	17	
DXS10103	18	
HPRTB	12	
DXS7423	14	

Coverage Information

Locus	Allele Name	Typed Allele?	Reads	Repeat Sequence
DXS10135	24	Yes	331	AAGAAAGAAAGAGAAAGGAAAGAAAGAAAGAAAGAAAGAAAGA
DXS8378	11	Yes	2629	CTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTAT
DXS7132	14	Yes	721	TAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGA
DXS10074	17	Yes	1464	AAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG
DXS10103	18	Yes	43	TAGATAGACTGACAGATAGATAGATAGATAGATAGATAGATAGATAGA
HPRTB	12	Yes	374	AGATAGATAGATAGATAGATAGATAGATAGATAGATAGATAGATA
DXS7423	14	Yes	1876	TCCATCCATCCATCTGTCCTTCCATCCATCCATCCATCCATCCATC

The Y-STR and X-STR worksheets in the sample genotype report show locus genotypes, quality control indicators, number of reads, and the nucleotide sequence of typed alleles. STRs with isometric alleles have sequence differences highlighted with a different font size. For information on the meaning of column headings and quality control indicators in the body of the worksheet, see *General Locus Information* on page 35 and *STR Sample Details Table* on page 37.

Figure 79 X STR Figure Worksheet



Sample X STR Figure

Typed Allele?	Yes

Sum of Reads	
Row Labels	Total
DXS10074	6650
16	3670
19	2980
DXS10103	352
17	352
DXS10135	2874
21.1	1535
27	1339
DXS10148	5943
20.1	3093
21.1	2850
DXS7132	12875
12	12875
DXS7423	9903
14	4995
15	4908
DXS8378	8741
10	4487
11	4254
HPRTB	6262
14	6262
Grand Total	53600

The figure worksheets show a pivot table and pivot chart for Y-STRs and X-STRs. The data for the table and chart in the figure worksheets are populated from their respective worksheets. The pivot chart illustrates the information in the table on the left of the read sums per locus. The pivot table is a summation of the reads for an allele at each locus.

iSNP Worksheet for Sample-Level Reports

The iSNP worksheets in the sample genotype report show genotypes, quality control indicators, and the number of reads for typed alleles. For information on the meaning of column headings in the body of the worksheet, see *General Locus Information* on page 35 and *SNP Sample Details Table* on page 40.

Figure 80 iSNP Sample Report Locus Information

Sample iSNP Report

Sample	Sample 10	
Project	Your Example Project	
Analysis	1.0 - Example Run	
Run	Example Run	
Gender	XY	
Created	04 Jun 2018 at 11:35AM (Pacific Standard Time)	
User	admin@forenseq.uas	
iSNP Locus Information		
Loci Typed	92 / 94	
Single Source	Pass	

Locus	Genotype	QC Indicators
rs1490413	A,A	
rs560681	G,G	
rs1294331	G,A	
rs10495407	G,G	
rs891700	G,G	
rs1413212	G,G	interpretation threshold
rs876724	C,C	

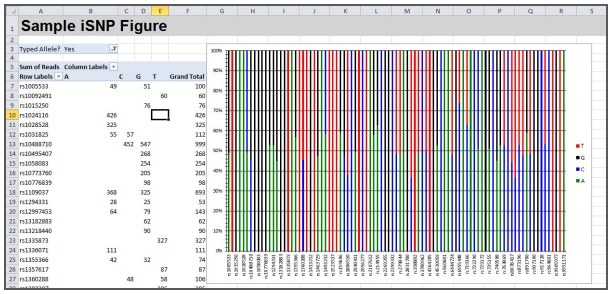
Figure 81 iSNP Sample Report Coverage Information

Coverage Information

Locus	Allele Name	Typed Allele?	Reads
rs1490413	A	Yes	273
rs560681	G	Yes	368
rs1294331	A	Yes	74
rs1294331	G	Yes	45
rs10495407	G	Yes	542
rs891700	G	Yes	289
rs1413212	G	Yes	142
rs876724	C	Yes	797

The figure worksheet shows a pivot table and pivot chart for iSNPs. The data for the table and chart in the figure worksheet are populated from the worksheet. The pivot chart illustrates the information in the table on the left of the read sums per locus. In the pivot table, each of the iSNP loci are listed. The first column is the locus name, identified as a reference SNP (rs) number. The second, third, fourth, and fifth columns list the possible alleles (ACGT) and the number of reads. The number of reads for each sample is listed by typed SNP allele in the Grand Total column. The chart is a 100% stacked bar chart representation of the percentage each allele contributes to a genotype.

Figure 82 iSNP Figure Worksheet



Sample History for Sample-Level Reports

The Sample History worksheet in the sample summary and sample details genotype reports shows the history of activity for the sample. The system-initiated events and user actions tracked by the software are displayed for the sample. The time and date, recorded event activity, and the user responsible are presented for each event.

Figure 83 Sample History

Sample History

Sample	Sample 10	
Project	Your Example Project	
Analysis	1.0 - Example Run	
Run	Example Run	
Gender	XY	
Created	04 Jun 2018 at 11:35AM (Pacific Standard Time)	
User	admin@forenseq.uas	

Sample History Log

Date and Time of Event	Recorded Activity	User
	Run Created	
03 Oct 2017 at 10:31 PM	Run name: Example Run	admin@forenseq.uas
	Analysis Created	
	Analysis name: 1.0 - Example Run	
	Analysis settings: Illumina Preinstalled	
03 Oct 2017 at 10:31 PM	Total Samples: 27	System Event
	Run Completed	
04 Oct 2017 at 6:24 AM	Run name: Example Run	System Event
	Analysis Submitted	
04 Oct 2017 at 6:24 AM	Analysis name: 1.0 - Example Run	System Event

NOTE

The Sample History worksheet is generated only when visibility of the Sample History feature is enabled. For more information on how to enable or disable visibility of this feature, see *Sample History: User Action and System Event Tracking* on page 44.

Locus-Level Settings for Sample-Level Reports

The Settings worksheet displays the analytical threshold, interpretation threshold, and intralocus balance values for STRs and iSNPS, as well as the stutter filter values for STRs used for the analysis of the sample.

Figure 84 Locus Settings

Locus Settings

Secondary Analysis Module Version	1.0.0

STR General Settings	
Intralocus Balance (%)	60

SNP General Settings	
Intralocus Balance (%)	50

STR Thresholds

Locus	Analytical Threshold (%)	Interpretation Threshold (%)	Stutter Filter (%)
Amelogenin	1.5	4.5	0
D1S1656	1.5	4.5	25
TPOX	1.5	4.5	10
D2S441	1.5	4.5	7.5
D2S1338	1.5	4.5	20
D3S1358	1.5	4.5	15
D4S2408	1.5	4.5	7.5
FGA	1.5	4.5	25
D5S818	1.5	4.5	12.5

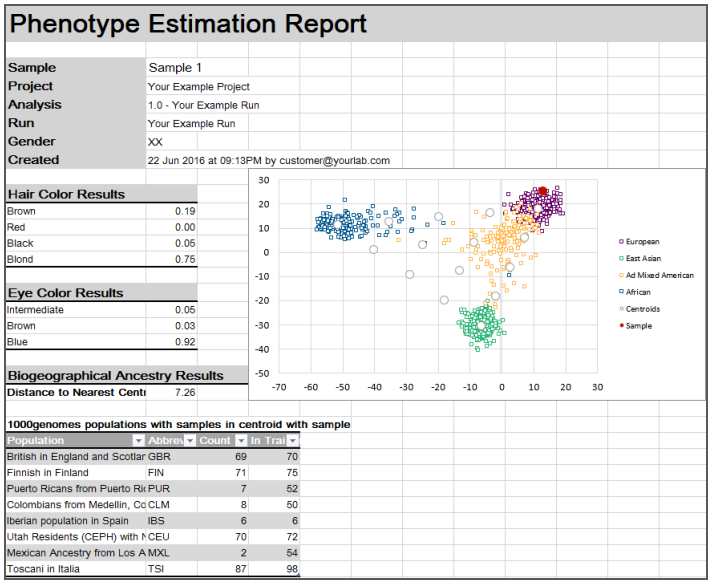
SNP Thresholds

Locus	Analytical Threshold (%)	Interpretation Threshold (%)
rs1480413	1.5	4.5
rs560681	1.5	4.5
rs1294331	1.5	4.5
rs10495407	1.5	4.5
rs891700	1.5	4.5
rs1413212	1.5	4.5
rs876724	1.5	4.5
rs1109037	1.5	4.5
rs993934	1.5	4.5

Phenotype Estimation Report

The Phenotype Estimation report is composed of data and charts of results and supporting information for a single sample. The report provides the same information as a phenotype estimation with the exception of reference samples in sample. See *Generating Phenotype Estimations* on page 65 and *Phenotype Estimation SNP Sample Details Table* on page 63.

Figure 85 Phenotype Estimation Report



Phenotype Estimation Report

Sample	Sample 1
Project	Your Example Project
Analysis	1.0 - Your Example Run
Run	Your Example Run
Gender	XX
Created	22 Jun 2016 at 09:13PM by customer@yourlab.com

Hair Color Results

Brown	0.19
Red	0.00
Black	0.05
Blond	0.75

Eye Color Results

Intermediate	0.05
Brown	0.03
Blue	0.92

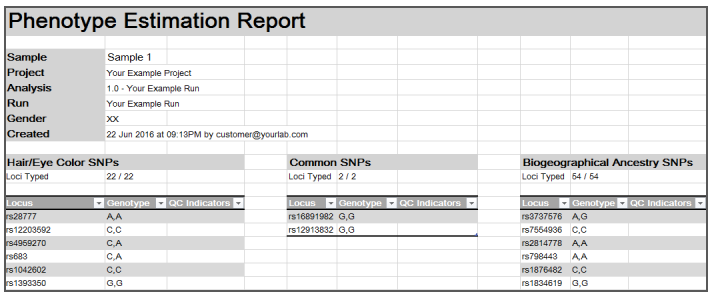
Biogeographical Ancestry Results

Distance to Nearest Centr	7.26

1000genomes populations with samples in centroid with sample

Population	Abbrev	Count	In Trai
British in England and Scotlar	GBR	69	70
Finnish in Finland	FIN	71	75
Puerto Ricans from Puerto Ric	PUR	7	52
Colombians from Medellin, Co	CLM	8	50
Iberian population in Spain	IBS	6	6
Utah Residents (CEPH) with N	CEU	70	72
Mexican Ancestry from Los A	MXL	2	54
Toscani in Italia	TSI	87	98

Figure 86 Phenotype Estimation Report Locus Information



Phenotype Estimation Report

Sample	Sample 1
Project	Your Example Project
Analysis	1.0 - Your Example Run
Run	Your Example Run
Gender	XX
Created	22 Jun 2016 at 09:13PM by customer@yourlab.com

Hair/Eye Color SNPs — Loci Typed 22 / 22

Locus	Genotype	QC Indicators
rs28777	A,A	
rs12203592	C,C	
rs4959270	C,A	
rs683	C,A	
rs1042602	C,C	
rs1393350	G,G	

Common SNPs — Loci Typed 2 / 2

Locus	Genotype	QC Indicators
rs16891982	G,G	
rs12913832	G,G	

Biogeographical Ancestry SNPs — Loci Typed 54 / 54

Locus	Genotype	QC Indicators
rs3737576	A,G	
rs7554936	C,C	
rs2814778	A,A	
rs798443	A,A	
rs1876482	C,C	
rs1834619	G,G	

Figure 87 Phenotype Estimation Report Coverage Information

Coverage Information

Locus	Allele Name	Typed Allele?	Reads
rs3737576	A	Yes	53
rs3737576	G	Yes	47
rs7554936	C	Yes	682
rs7554936	T	No	0
rs2814778	A	Yes	3870
rs2814778	G	No	0
rs798443	A	Yes	596
rs798443	G	No	0
rs1876482	C	Yes	1233
rs1876482	T	No	0

The report is a Microsoft Excel workbook file that you can download and customize by using sorting filters, and drop-down lists. After a phenotype estimation for a sample is performed, a report can be created from the estimation. Reports are available for download from the Project page on the Phenotype tab.

The workbook is subdivided into four worksheets:

- Estimation
- SNP Data
- SNP Balance Figure
- Settings

1 On the Project page, select the **Phenotypes** tab.

2 Click the blue link icon for a sample in the list.
The samples on the list are ones that have a phenotype estimation.

3 Click **Generate Report**.
When the report is complete, the Download icon displays.

4 Click **Download** to access the report.

Phenotype Estimation reports for single samples can also be generated from the Phenotype Estimation page after the estimation is complete. See *Generating Phenotype Estimations* on page 65.

NOTE
Typed alleles in the report reflect current data. If SNP allele calls are updated after you generate a phenotype estimation, perform phenotype estimation again and generate another report.

Estimation Worksheet for Phenotype Estimation Report

The Estimation worksheet shows the hair color, eye color, and biogeographical ancestry results from a phenotype estimation. The hair color and eye color results are the values provided by the phenotype estimation. The biogeographical ancestry estimation results include the distance to the nearest centroid, the 1000 genomes populations with samples in the centroid, and the PCA plot with the 1000 genomes populations and the sample plotted against the first two principal components.

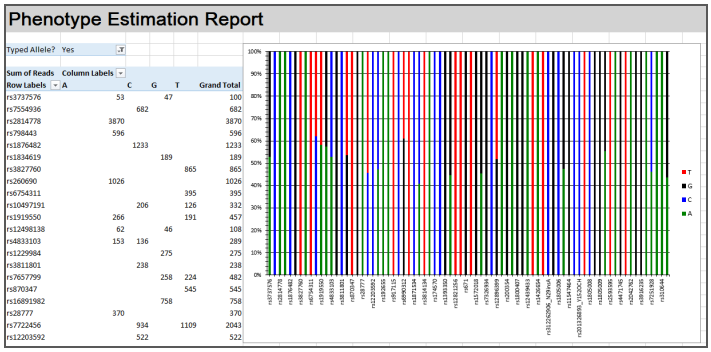
SNP Data Worksheet for Phenotype Estimation Report

The SNP Data worksheet shows the genotypes for the aiSNP and piSNP data of the estimated sample. The results are presented in a fashion similar to the ForenSeq Universal Analysis Software Hair and Eye Color SNPs, Common SNPs, and Biogeographical Ancestry SNPs presented in separate tables. For each locus, the genotype and any QC indicators are reported.

SNP Balance Worksheet for Phenotype Estimation Reports

The SNP balance figure worksheet shows a pivot table and pivot chart for SNPs in use for the phenotype estimation report. The data for the table and chart in the SNP balance figure worksheet are populated from the worksheet. The pivot chart illustrates the information in the table on the left of the read sums per locus. In the pivot table, each of the SNP loci are listed. The first column is the locus name, identified as an rs number. The second, third, fourth, and fifth columns list the possible alleles (ACGT) and the number of reads. The number of reads for each sample is listed by typed SNP allele in the Grand Total column. The chart is a 100% stacked bar chart representation of the percentage each allele contributes to a genotype.

Figure 88 SNP Balance Figure Worksheet



Phenotype History for Phenotype Estimation Reports

The Phenotype History worksheet in the sample summary and sample details genotype reports shows the history of phenotype activity for the sample. The system-initiated events and user actions tracked by the software are displayed for the sample. The time and date, recorded event activity, and the user responsible are presented for each event.

Figure 89 Phenotype History

Phenotype History

Sample	Sample 1
Project	Mixture Examples
Analysis	1.0 - Example Mixture Run
Run	Example Mixture Run
Gender	XY
Created	09 Jul 2018 at 08:41PM (Pacific Standard Time)
User	admin@forenseq.uas

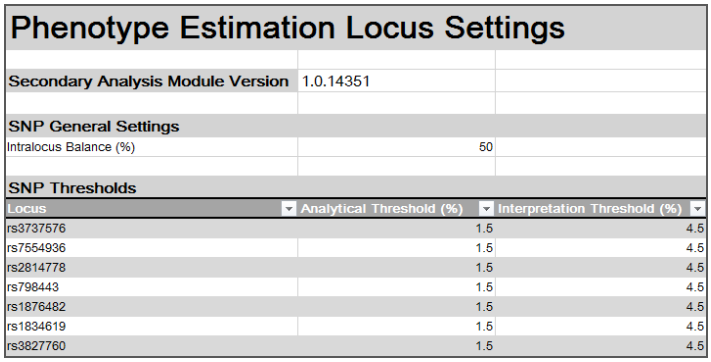
Phenotype History Log

Date and Time of Event	Recorded Activity	User
29 Sep 2017 at 3:24 AM	Run Created Run name: Example Mixture Run	exportuser@forenseq.uas
29 Sep 2017 at 3:24 AM	Analysis Created Analysis name: 1.0 - Example Mixture Run Analysis settings: Illumina Preinstalled Total Samples: 4	System Event
29 Sep 2017 at 8:31 PM	Run Completed Run name: Example Mixture Run	System Event
29 Sep 2017 at 8:31 PM	Analysis Submitted Analysis name: 1.0 - Example Mixture Run	System Event
29 Sep 2017 at 9:12 PM	Analysis Completed Successfully Analysis name: 1.0 - Example Mixture Run Analysis duration from submission: 40m.59s Analysis duration from initiation:	System Event
	Phenotype Estimation initiated Loci Evaluated: rs16891982 , rs28777 , rs12203592 , rs4959270 , rs683 , rs1042602 , rs1393350 , rs12821256 , rs12896399 , rs2402130 , rs1800407 , rs12913832 ,	

Locus Settings for Phenotype Estimation Reports

The Settings worksheet displays the analytical and interpretation thresholds for loci in the phenotype estimation of the sample. The worksheet also displays the intralocus balance setting for SNPs in the analysis.

Figure 90 Locus Settings



Phenotype Estimation Locus Settings

Secondary Analysis Module Version	1.0.14351

SNP General Settings

Intralocus Balance (%)	50

SNP Thresholds

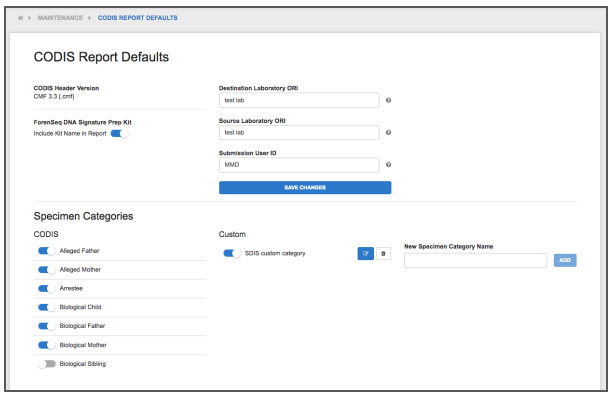
Locus	Analytical Threshold (%)	Interpretation Threshold (%)
rs3737576	1.5	4.5
rs7554936	1.5	4.5
rs2814778	1.5	4.5
rs798443	1.5	4.5
rs1876482	1.5	4.5
rs1834619	1.5	4.5
rs3827760	1.5	4.5

Sample-Level CODIS Report

The sample-level CODIS report is composed of analysis results in CODIS cmf format, version 3.3. The results for the 20 CODIS core loci, as well as Amelogenin and all of the additional typed autosomal and Y STR markers analyzed with the ForenSeq DNA Signature Prep Kit, are output in the report.[1] When initiated from the Sample Details page, a three-step CODIS report generator is used to capture information about the sample for CODIS upload. After a sample-level CODIS report is created, it is available for download from the Project page on the Sample Genotype tab. See *Viewing the Project Page* on page 21.

Default values for many of these fields can be defined by an administrator user in the CODIS Report Defaults section of the maintenance page. Specified values pre-populate in the report generator workflow to expedite report generation. See *CODIS Report Defaults* on page 113.

Figure 91 CODIS Defaults



The CODIS format allows entry of the PCR kit used as an optional field. The toggle at the bottom of the dialog allows the ForenSeq DNA Signature Prep Kit to be included in the report.

NOTE
If this is the first time uploading a CODIS report generated using the ForenSeq DNA Signature Prep Kit, your LDIS or SDIS might need to be configured to accept the value.

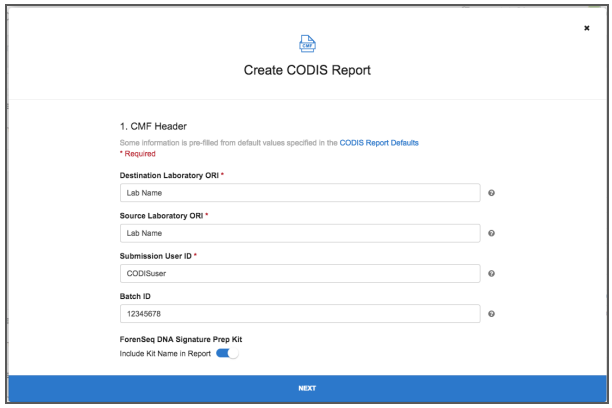
Generate a Sample-Level CODIS Report

Follow the prompts in the CODIS report wizard to generate a CODIS Report.

1 Populate the CMF header to include the following:
- Destination Laboratory ORI
- Source Laboratory ORI
- Submission User ID
- Batch ID
- Option to include the ForenSeq DNA Signature Prep Kit name in the report

Select **Next** to proceed.

Figure 92 Sample-Level CODIS Report Wizard—Step 1

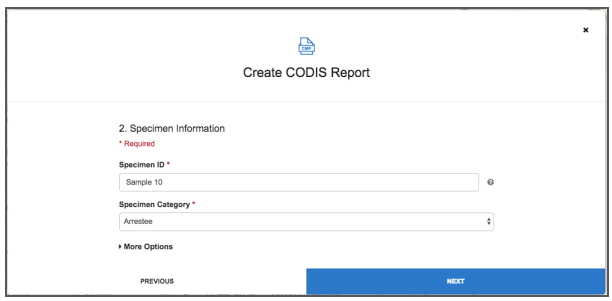


2 Enter Specimen Information.

- The sample name provided in the ForenSeq Universal Analysis Software is used as default for the Specimen ID and can be edited, if desired.
- Select the Specimen Category from the list of CODIS-approved options. The Specimen Category drop-down list is administrator user-configurable in the CODIS Report Defaults section with the option to include or exclude individual CODIS categories, as well as add custom category names.
- For samples with additional information to enter, expand the More Options selection to provide information regarding source ID, case ID, whether it should be considered a partial profile, NCIC number, ViCAP number, or specimen comments for the sample.

Select **Next** to proceed or **Previous** to return to step 1 of the report generator.

Figure 93 Sample-Level CODIS Report Wizard—Step 2

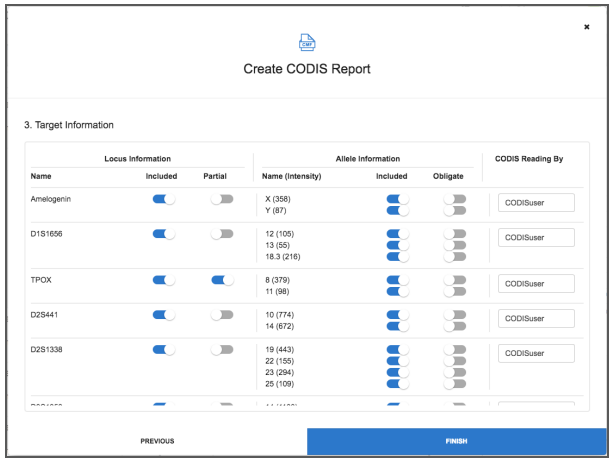


3 Enter Target Information about the loci and alleles.

- For locus information, use the toggle selections to include or exclude individual locus results from the report and to designate a locus as partial, if desired.
- For allele information, use the toggle selections to include or exclude individual allele results from the report, and to designate one allele as obligate, if desired. Note that obligate alleles are recognized in a "+" format in CODIS, indicating that the allele is required for a match to be reported.
- The CODIS Reading By user name is populated with the Submission User ID specified in step 1 of the report generator. This field can be edited, if desired.

Click **Finish** to generate the report or **Previous** to return to step 2 of the report generator.

Figure 94 Sample-Level CODIS Report Wizard—Step 3



Upon selecting Finish, a green confirmation is displayed to indicate that the CODIS report is being generated. The finished report is then available for download with other sample-level reports on the Project page.

Figure 95 Example CMF Sample-Level CODIS Report

```
<?xml version="1.0" encoding="utf-8"?>
<CODISImportFile xmlns:xsd="http://www.w3.org/2001/XMLSchema" xmlns:xsi="http:
schema">
  <HEADERVERSION>3.3</HEADERVERSION>
  <MESSAGETYPE>Import</MESSAGETYPE>
  <DESTINATIONORI>Lab Name</DESTINATIONORI>
  <SOURCELAB>Lab Name</SOURCELAB>
  <SUBMITBYUSERID>CODISuser</SUBMITBYUSERID>
  <SUBMITDATETIME>2018-07-10T00:15:52</SUBMITDATETIME>
  <BATCHID>12345678</BATCHID>
  <KIT>ForenSeq DNA Signature Prep Kit</KIT>
  <SPECIMEN SOURCEID="No">
    <SPECIMENID>Sample 1</SPECIMENID>
    <SPECIMENCATEGORY>Biological Father</SPECIMENCATEGORY>
    <LOCUS>
      <LOCUSNAME>Amelogenin</LOCUSNAME>
      <READINGBY>CODISuser</READINGBY>
      <READINGDATETIME>2018-07-10T00:15:52</READINGDATETIME>
      <ALLELE>
        <ALLELEVALUE>X</ALLELEVALUE>
      </ALLELE>
      <ALLELE>
        <ALLELEVALUE>Y</ALLELEVALUE>
      </ALLELE>
    </LOCUS>
    <LOCUS>
      <LOCUSNAME>D1S1656</LOCUSNAME>
      <READINGBY>CODISuser</READINGBY>
      <READINGDATETIME>2018-07-10T00:15:52</READINGDATETIME>
      <ALLELE>
        <ALLELEVALUE>12</ALLELEVALUE>
      </ALLELE>
      <ALLELE>
        <ALLELEVALUE>13</ALLELEVALUE>
      </ALLELE>
      <ALLELE>
        <ALLELEVALUE>18.3</ALLELEVALUE>
      </ALLELE>
    </LOCUS>
    <LOCUS PARTIALLOCUS="true">
```

Resources

1. D.R. Hares. Selection and implementation of expanded CODIS core loci in the United States. *Forensic Sci Int Genet.* 2015;9:150–161.

Project-Level CODIS Report

The project-level CODIS report is composed of analysis results for all selected samples in CODIS cmf format, version 3.3. The results for each sample for the 20 CODIS core loci, as well as Amelogenin and all of the additional typed autosomal and Y STR markers analyzed with the ForenSeq DNA Signature Prep Kit are output in a single report. When initiated from the Project Page, a three-step CODIS report generator is used to capture information about the samples selected for CODIS upload. After a project-level CODIS report is created, it is available for download from the Project page on the Project Level tab. See *Viewing the Project Page* on page 21.

Default values for many of these fields can be defined by an administrator user in the CODIS Report Defaults section of the maintenance page. Specified values pre-populate in the report generator workflow to expedite report generation. See *CODIS Report Defaults* on page 113.

The CODIS format allows entry of the PCR kit used as an optional field. The toggle at the bottom of the dialog allows the ForenSeq DNA Signature Prep Kit to be included in the report.

NOTE
If this is the first time uploading a CODIS report generated using the ForenSeq DNA Signature Prep Kit, your LDIS or SDIS might need to be configured to accept the value.

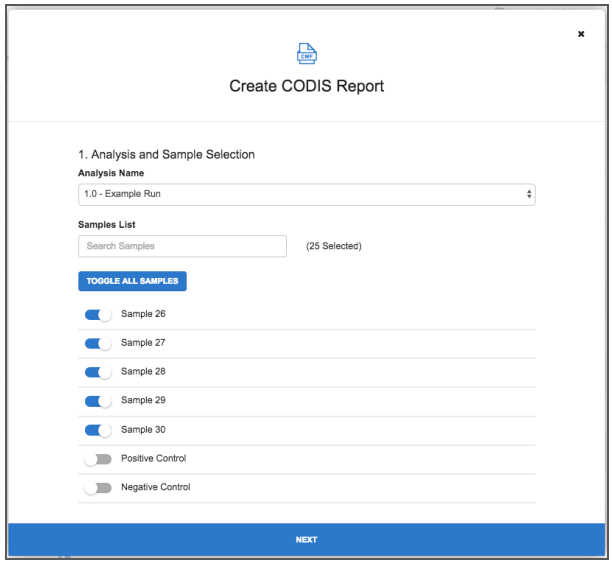
Generate a Project-Level CODIS Report

Follow the prompts in the CODIS report wizard to generate a CODIS report.

1 Select desired samples for inclusion in the project-level batch report.
- Select the analysis name where samples are located.
- Use the toggle options to select which samples from the analysis you would like to include in the report.

Select **Next** to proceed.

Figure 96 Project-Level CODIS Report Wizard—Step 1

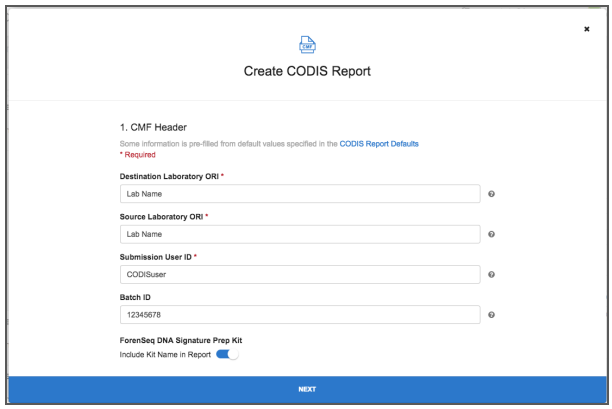


2 Populate the CMF header to include the following:
- Destination Laboratory ORI
- Source Laboratory ORI

- Submission User ID
- Batch ID
- Option to include the ForenSeq DNA Signature Prep Kit name in the report

Select **Next** to proceed or **Previous** to return to step 1 of the report generator.

Figure 97 Project-Level CODIS Report Wizard—Step 2



3 Enter Specimen Information.

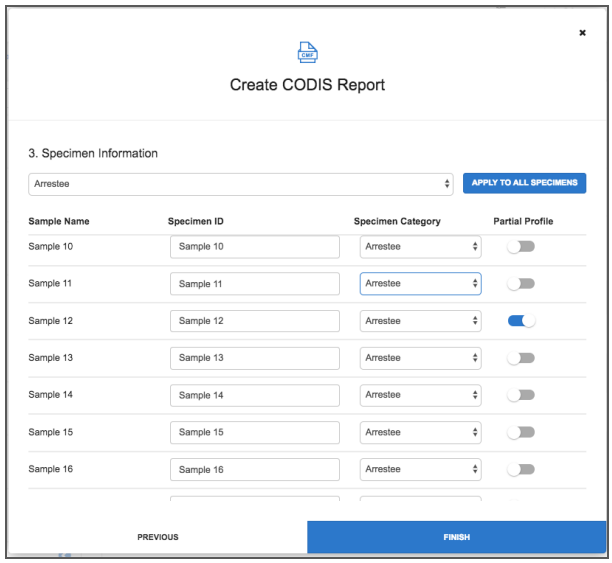
- The sample name provided in the ForenSeq Universal Analysis Software is used as default for the Specimen ID and can be edited, if desired.
- Select the Specimen Category from the list of CODIS-approved options. The Specimen Category drop-down list is administrator user-configurable in the CODIS Report Defaults section with the option to include or exclude individual CODIS categories, as well as add custom category names.

NOTE
The Apply to All Specimens option can be used to apply the same category to every sample, or the samples can be assigned categories individually.

- Use the toggle selections to designate a partial profile sample, if desired.

Click **Finish** to generate the report or **Previous** to return to step 2 of the report generator.

Figure 98 Project-Level CODIS Report Wizard—Step 3



Upon selecting Finish, a green confirmation is displayed to indicate that the CODIS report is being generated. The finished report is then available for download with other project-level reports on the Project page.

NOTE
If a project-level CODIS report is created for a single sample, the completed report will be available for download from the sample-level reports tab.

Figure 99 Example CMF Project-Level CODIS Report

```
<?xml version="1.0" encoding="utf-8"?>
<CODISImportFile xmlns:xsd="http://www.w3.org/2001/XMLSchema" xmlns:xsi="http:
schema">
  <HEADERVERSION>3.3</HEADERVERSION>
  <MESSAGETYPE>Import</MESSAGETYPE>
  <DESTINATIONORI>Lab Name</DESTINATIONORI>
  <SOURCELAB>Lab Name</SOURCELAB>
  <SUBMITBYUSERID>CODISuser</SUBMITBYUSERID>
  <SUBMITDATETIME>2018-07-10T00:15:52</SUBMITDATETIME>
  <BATCHID>12345678</BATCHID>
  <KIT>ForenSeq DNA Signature Prep Kit</KIT>
  <SPECIMEN SOURCEID="No">
    <SPECIMENID>Sample 1</SPECIMENID>
    <SPECIMENCATEGORY>Biological Father</SPECIMENCATEGORY>
    <LOCUS>
      <LOCUSNAME>Amelogenin</LOCUSNAME>
      <READINGBY>CODISuser</READINGBY>
      <READINGDATETIME>2018-07-10T00:15:52</READINGDATETIME>
      <ALLELE>
        <ALLELEVALUE>X</ALLELEVALUE>
      </ALLELE>
      <ALLELE>
        <ALLELEVALUE>Y</ALLELEVALUE>
      </ALLELE>
    </LOCUS>
    <LOCUS>
      <LOCUSNAME>D1S1656</LOCUSNAME>
      <READINGBY>CODISuser</READINGBY>
      <READINGDATETIME>2018-07-10T00:15:52</READINGDATETIME>
      <ALLELE>
        <ALLELEVALUE>12</ALLELEVALUE>
      </ALLELE>
      <ALLELE>
        <ALLELEVALUE>13</ALLELEVALUE>
      </ALLELE>
      <ALLELE>
        <ALLELEVALUE>18.3</ALLELEVALUE>
      </ALLELE>
    </LOCUS>
    <LOCUS PARTIALLOCUS="true">
      <LOCUSNAME>TPOX</LOCUSNAME>
```

CODIS Report Fields

Field	Definition	Required/Optional
Destination Laboratory ORI	Originating Agency Identifier (ORI) for the destination CODIS agency that will process the import file.	Required Field
Source Laboratory ORI	Originating Agency Identifier (ORI) for the agency processing the specimen.	Required Field
Submission User ID	The User ID of the person that generated the import file.	Required Field
Batch ID	An identifier for the batch to which the specimen belongs.	Optional Field
PCR Kit (ForenSeq DNA Signature Prep Kit)	The kit that was used to produce the specimen results.	Optional Field
Specimen ID	A unique identifier for the specimen within the import file.	Required Field
Specimen Category	List of CODIS-defined or custom specimen classification categories.	Required Field
Source ID	Indicates if the identity of the specimen contributor is known.	Optional Field
Case ID	An identifier for the case to which the specimen belongs.	Optional Field
Partial Profile	Indicates if the profile might have additional information.	Optional Field
NCIC Number	A unique number that is accepted for storage in the NCIC system.	Optional Field
ViCAP Number	A unique number that is accepted for storage in the ViCAP system.	Optional Field
Specimen Comment	Comment to provide information regarding the specimen.	Optional Field
Partial Locus	Indicates if the locus might have additional information.	Optional Field
CODIS reading by	Valid CODIS User ID of the person performing the reading.	Required Field
Obligate Allele	Indicates if the allele is required to be present for a match to be reported during searching. Only one allele may be designated.	Optional Field

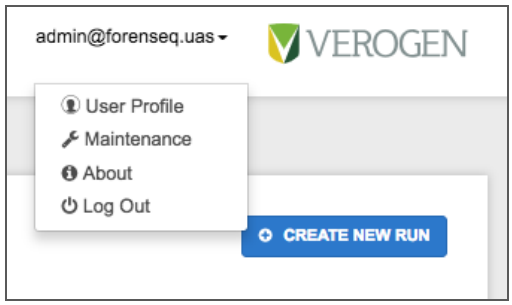
System Settings

Chapter 7

Introduction

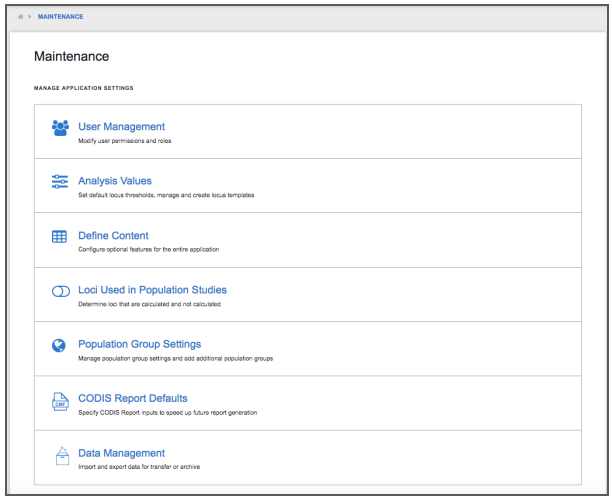
System settings are available from the Maintenance page. The Maintenance page and system settings features are only visible and available to users with administrator account privileges.

Figure 100 System Settings



To access the Maintenance page, click the arrow next to the user account drop-down list and select **Maintenance**.

Figure 101 Maintenance Page



The following system settings features are accessed from the Maintenance page.

Table 28 System Settings Features

Settings Name	Description
User Management	Activate or deactivate user accounts. Modify user permissions and roles, including the assignment of administrator privileges.
Analysis Values	Set default analysis thresholds. Create alternate analysis settings templates. Allow flanking region report generation in a template. Create locus templates.
Define Content	Configure optional features for the entire application. Enable visibility and reporting of Sample History tracked events. Control the loci that are displayed in the software and are analyzed by the software.
Loci Used in Population Studies	Control the loci that are included in population statistics calculations.
Population Group Settings	Select one or more population databases for the calculation of population statistics. You can select population databases included with the software, or upload your own.
CODIS Report Defaults	Specify default CODIS report field entries for report generation. Configure the specimen categories list.
Data Management	Export data from the ForenSeq Universal Analysis Software for transfer to another installation or data archival. Import previously exported data to the ForenSeq Universal Analysis Software.

Notifications

When a user is logged on as an administrator, pending notifications that require attention appears in a blue box in the top right corner of the page next to a bell. To view pending notifications, click the down arrow next to the number of notifications.

Figure 102 Administrator Notifications

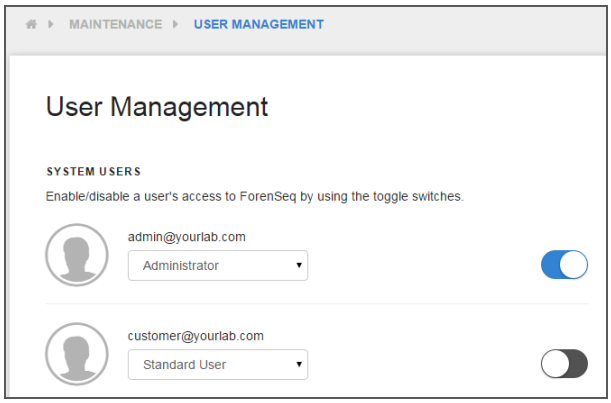
When you click a notification information alert, a page related to the notification opens. For example, if the notification is to approve a new user, the User Management page opens. After you click a notification information alert, the reminder does not appear again on the page.

User Management

Access to the system is controlled by account management on the User Management page. The User Management page is launched from the Maintenance page. Users with administrator privileges enable and disable system accounts and assign account access levels.

When a user creates an account on the log in page of the software, an administrator approves the account on the User Management page before the user can log in and access the system. An enabled account has access to the ForenSeq Universal Analysis Software and the MiSeq FGx.

Figure 103 User Management Page



All accounts are listed on the User Management page. The access level for each account is displayed in a drop-down list below the account ID. A toggle switch next to the account indicates if the account is enabled or disabled. An account with a gray toggle switch is disabled, and the user cannot log in.

Enabling and Disabling Accounts

A user creates an account on the login page of the software. An administrator enables the account before the user can log in and access the system. When the user creates an account, a message is displayed for the administrator user that the account is pending creation. Sign-in is possible when the administrator approves the account.

After an administrator enables the account, which is controlled on the User Management page, the user can log in. If the user attempts to log in again before the account is approved, a message is displayed that the account was not approved for login, and to contact the administrator.

To enable an account, click the toggle switch next to the account name so that the switch displays a blue background and is in position on the right side.

To disable an account, click the toggle switch next to the account name so that the switch displays a gray background and is in position on the left side.

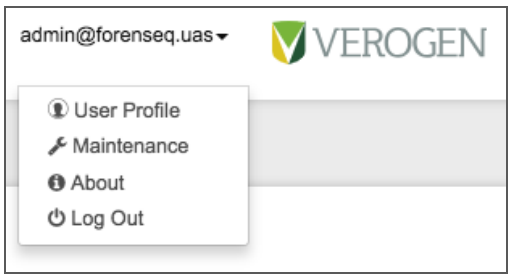
Assigning Account Access Levels

Each user is assigned to a group that determines access privileges for the system. A user with administrator privileges has all access privileges of a regular user, as well as the ability to configure system settings that are on pages launched from the Maintenance page. Administrator account privileges can be assigned or removed by other administrator users.

Changing an Account Password

The current password associated with an account can be changed when you are logged in with the current password. A password change takes effect the next time you are required to log in.

Figure 104 Account Drop-Down—User Profile



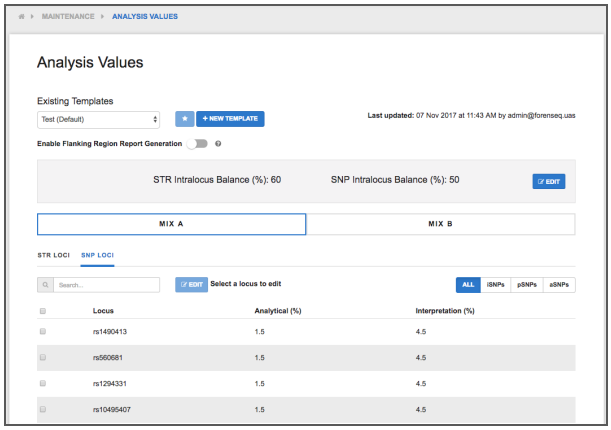
To change a user password, perform the following steps:

1. Select **User Profile** from the account drop-down list.
2. Enter the current password.
3. Enter the new password. The password must meet system requirements for a valid password.
4. In the Confirm New Password field, enter the new password again.
5. Select **Save**.

Changing Locus Thresholds

Locus thresholds and stutter filters (%) that impact automated genotype calling and generation of flanking region reports can be changed on the Analysis Values page. The Analysis Values page is launched from the Maintenance Page. See *Introduction* on page 96. Users with administrator privileges can change current thresholds or create new threshold templates. Changes to locus thresholds can then be applied to new analyses.

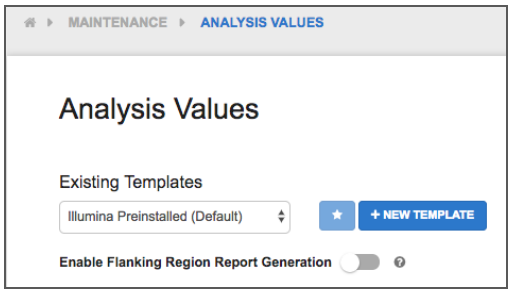
Figure 105 Analysis Values Page



Creating a Locus Threshold Template

A collection of locus threshold settings is saved under a single name as a template for easy access and reference within the system. When you create a new template, it is based on the settings for the template open at the time of creation. To access the analysis settings template used for a particular analysis, click on the linked analysis template name next to Analysis settings on the Project page.

Figure 106 Analysis Templates



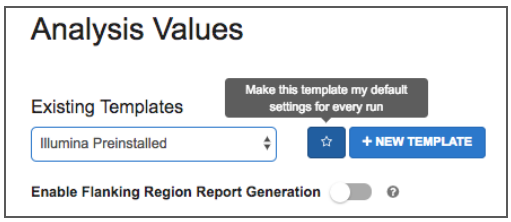
To create a new template, access the Analysis Values page from the Maintenance page.

1. From the Existing Templates drop-down list, select a template on which to base the new template.
2. Click **+New Template**.
3. Enter a name for the template.
4. Click **Save**.

Designating a Default Locus Threshold Template

Any analysis template can be designated as the system default template. The purpose of a default template is to define locus thresholds for major version analyses. A major version analysis is the first analysis conducted on a run version or at run completion.

Figure 107 Default Template



To designate a template as the default template, access the Analysis Values page from the Maintenance page.

1 Select the template from the Existing Templates drop-down list.

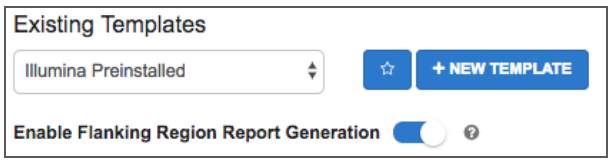
2 Click the blue star icon.

Enabling Flanking Region Report Generation

An analysis template can have flanking region report generation enabled, such that when any analysis with that template is executed, a project-level flanking region report generates automatically upon completion of analysis.

To enable flanking region report generation for a particular analysis template, access the Analysis Values page from the Maintenance page and toggle on (blue position) **Enable Flanking Region Report Generation**. All subsequently executed analyses with that template assigned will generate a flanking region report.

Figure 108 Flanking Region Report Toggle



When a flanking region-enabled analysis is executed:

- The regular, non-flanking region analysis data appears in the ForenSeq Universal Analysis Software Analysis page, exactly as it does with a flanking region disabled analysis.
- Flanking region sequence data is not depicted in the user interface and is not utilized to inform genotyping results in the ForenSeq Universal Analysis Software.
- The flanking region sequence data results are only available in the project-level Excel report generated at completion of analysis.

NOTE
When enabled, the flanking region report is automatically generated for each analysis conducted with the analysis template. The flanking region report is a project-level report; all samples in the project have the flanking regions output. For more information, see *Project-Level Flanking Region Report* on page 75.

Changing Analysis-Level Settings

STR intralocus balance settings and SNP intralocus balance settings in a template are used for all loci across an analysis. These analysis-level settings apply to all ForenSeq DNA primer mix types and loci.

To change the analysis-level settings for a template, access the Analysis Values page from the Maintenance page.

1 Select the template from the Existing Templates drop-down list.

2 Enter new STR Intralocus Balance or SNP Intralocus Balance values.
If the values were changed before, click **Edit** on the row that displays STR Intralocus Balance and SNP Intralocus Balance values. Then enter new STR Intralocus Balance or SNP Intralocus Balance values.

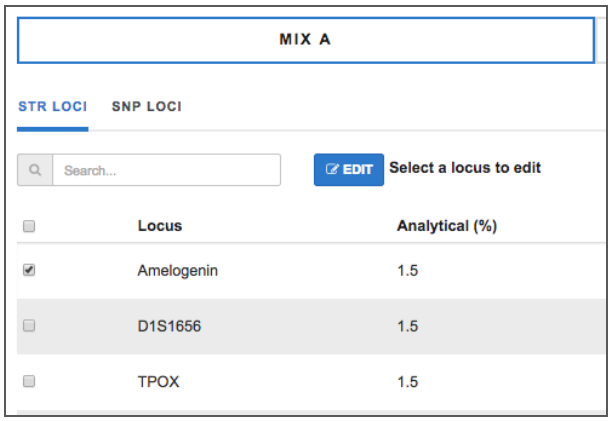
3 Click **Save**.

When you click Save, the new settings are applied to the template. The Last Updated time stamp in the upper-right corner of the page is refreshed with the new time and date, confirming when the change in analysis settings values is made to the template.

Changing Locus Level Settings

If you change the thresholds for an individual locus, it is important to consider whether you want to apply the change to both ForenSeq DNA Primer Mix A and ForenSeq DNA Primer Mix B.

Figure 109 Flanking Region Report Toggle



To change the locus level settings for a template, access the Analysis Values page from the Maintenance page.

1 Select the template from the Existing Templates drop-down list.

2 Select ForenSeq DNA Primer Mix A or ForenSeq DNA Primer Mix B tab.

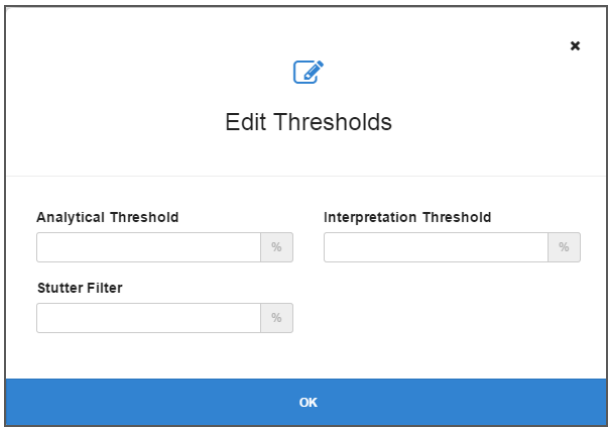
3 Select the STR or the SNP loci tab.
To find a particular locus, you can use the Search feature, or separate the loci by selecting a filter button above the list.
- For STRs, the list can be filtered for all, autosomal, X chromosome, or Y chromosome
- For SNPs, the list can be filtered for all, iSNPs, aSNPs, or pSNPs

4 To select loci for locus level changes, click the small box to the left of the locus name. To select all loci in the tab, click the small box at the top of the list on the left side.

5 Click **Edit**. An Edit Threshold dialog opens.

6 Enter new values for locus level settings.

Figure 110 Edit Threshold Dialog

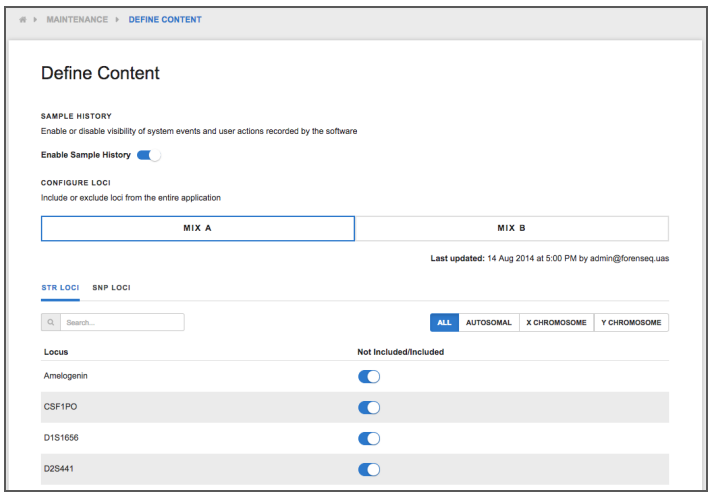


7 Click **OK**. When you click OK, the new settings are applied to the template. The Last Updated time stamp in the upper-right corner of the page is refreshed with the new time and date, confirming when the change in locus level settings values is made to the template.

Defining Content Within the Application

The Define Content page is accessed by an administrator user from the Maintenance page. Application-wide, configurable options may be defined, such as enabling visibility and reporting of Sample History activity, and specifying locus content for analysis.

Figure 111 Define Content Page



Enabling Sample History

To enable visibility of Sample History, toggle on (blue position) the **Enable Sample History** toggle at the top of the Define Content page.

When enabled, certain user actions and system-initiated events that are logged by the software are visible as a list of events in analysis-level, sample-level, and locus-level activity dialogs. Additionally, these same events are recorded in the sample-level genotype reports (Sample History tab in Excel report).

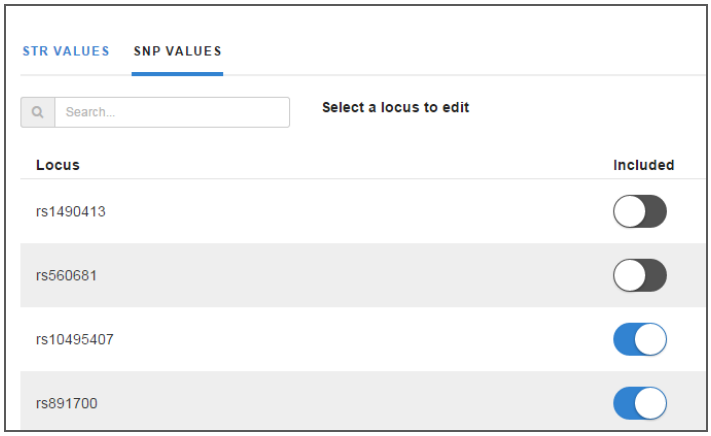
If the Sample History feature is disabled, user action and system-initiated events continue to be tracked by the software; however, they are not visible as activity dialogs in the user interface and are not included in the sample reports. If visibility of the feature is enabled at a later date, all previously tracked activity associated with that sample will be visible in the activity dialogs and recorded in subsequently created sample reports.

NOTE
If the Sample History feature is disabled *during data export*, the following information will be visible in the Sample History activity log *upon import*: run, analysis, report, and population statistic events or actions.

Selecting Loci Content for Analysis

The ForenSeq Universal Analysis Software is equipped to analyze all loci that are included in ForenSeq kits. While the software can analyze the entire range of loci, locus content can be defined by an administrator. Deselected loci are not analyzed and results for the loci are not generated.

Figure 112 Configure Loci Content for Analysis



The selection of loci included for analysis is controlled on the Define Content page. By default, the software includes all kit loci in analysis.

NOTE
The gender that is assigned to a sample is based on the presence of signal at a sufficient number of Y loci. When modifying the loci included for analysis, at least three X loci or three Y loci are required in order to determine gender. If not enough X or Y loci are active, gender cannot be determined for the samples, and all samples are assigned the status of *Inconclusive*. The reason that gender cannot be called in this instance is because the maximum copy number for a locus is the maximum copy number for an assigned gender call. For example, the X loci will have a copy number of 2.

NOTE
The deactivation of at least one pSNP prevents hair and eye color estimation, if you execute phenotype estimation. The complete set of ForenSeq pSNPs are required.

Selecting Loci to Include in Analysis

By default, the software includes all kit loci in analysis. To define loci in analyses, access the Define Content page from the Maintenance page.

1 Select ForenSeq DNA Primer Mix A or ForenSeq DNA Primer Mix B tab.

2 Select the STR or the SNP loci tab.
To find a particular locus, you can use the Search feature, or separate the loci by selecting a filter button above the list.
- For STRs—Filter the list for all, autosomal, X chromosome, or Y chromosome
- For SNPs—Filter the list for all, iSNPs, aSNPs, or pSNPs

3 To deselect a locus, click the toggle switch next to the locus name so that the switch displays a gray background and is in position on the left side.

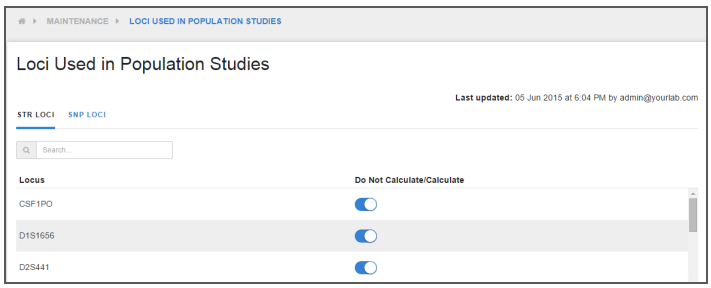
NOTE
Deselecting a locus is specific to the ForenSeq DNA primer mix type. If a locus is present in more than one mix type, consider adjusting both ForenSeq DNA primer mixes.

4 To select a locus to include for analysis, click the toggle switch next to the locus name so that the switch displays a blue background and is in position on the right side.

Define Loci for Population Studies

By default, the ForenSeq Universal Analysis Software uses autosomal STR and iSNP loci that are typed with the ForenSeq DNA Signature Prep Kit in calculating population statistics. Inclusion of loci for calculating population statistics can be defined by an administrator. Deselected loci are not used to calculate population statistics.

Figure 113 Loci Used in Population Studies

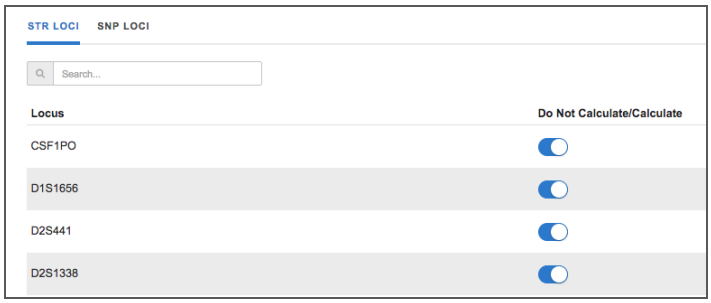


The selection of loci included for population statistics calculations is controlled on the Locus Used in Population Studies page.

Selecting Loci Content for Population Studies

To define loci used in population statistics, access the Loci Used in Population Studies page from the Maintenance page.

Figure 114 Loci Content for Population Studies



1 Select the STR or the SNP loci tab.
To find a particular locus, you can use the Search feature.

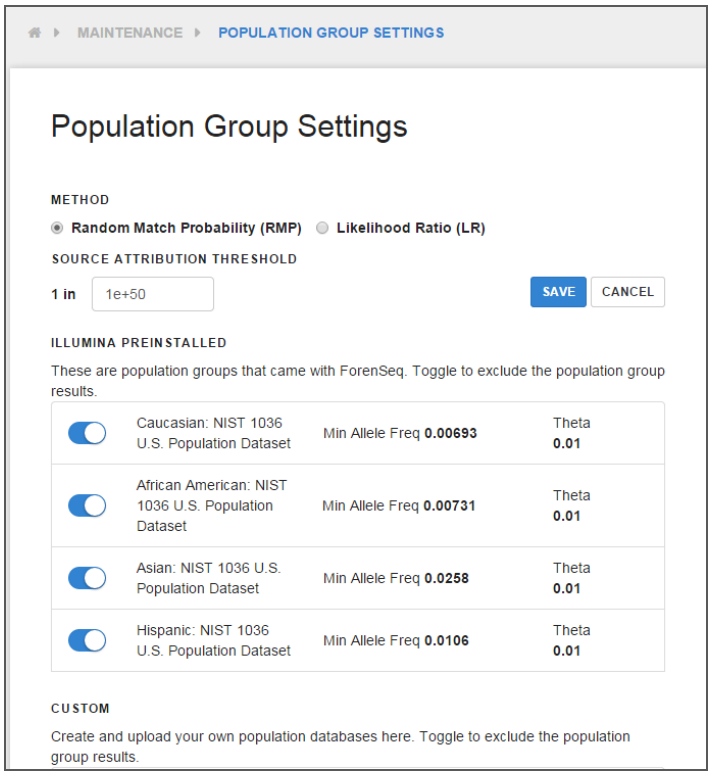
2 To deselect a locus, click the toggle switch next to the locus name so that the switch displays a gray background and is in position on the left side.

3 To select a locus to include, click the toggle switch next to the locus name so that the switch displays a blue background and is in position on the right side.

Population Group Settings

Population statistics calculations by the software require 2 components: at least 1 population group database to reference, and a method of calculation. The software comes installed with several population group databases that are ready for use. You can use population databases included with the software, or upload your own. Although random match probability is the default calculation method, you can select likelihood ratio instead. The software also supports the ability to define source attribution thresholds. Source attributions are relative to guidelines from Budowle et al.[1]

Figure 115 Populations Group Settings



Population group database selection, method of population statistics calculation, and source attribution thresholds are controlled on the Population Group Settings page. The Population Group Settings page is launched from the Maintenance Page. See *Introduction* on page 96.

Resources

1. Budowle B, Chakraborty R, Carmody G, Monson KL. Source Attribution of a Forensic DNA Profile. *Forensic Science Communications*. 2000;2(3).

Defining the Statistics Calculation Method

The ForenSeq Universal Analysis Software supports the calculation of population statistics by either random match probability or likelihood ratio methods, with use of the 2p or 2p - p^2 rule. Use of the likelihood ratio method creates the inverse of the result retrieved from that of the random match probability method.

Figure 116 Population Statistics Calculation Method

By default, the software uses random match probability and uses the 2p - p^2 setting for population statistics calculations. A change to the method used for calculation can be selected on the Population Group Settings page. If you select a different method of calculation, this method is used for all new calculations in the system going forward. Existing results remain unchanged.

NOTE
If you select Likelihood Ratio, the software automatically inverts the source attribution threshold setting. See *Changing the Source Attribution Threshold* on page 109.

The selection of calculation methods is controlled on the Population Group Settings page. To change the calculation method or calculation rule for population statistics, access the Population Groups Settings page from the Maintenance page.

1 Select **Random Match Probability (RMP)** or **Likelihood Ratio (LR)** in the Method area of the page.

2 Select **Use 2p** or **Use 2p - p^2** in the Population Statistics Calculation area of the page.

3 Click **Save**.

Applying the 2p Rule to a Homozygous Locus

When the genotype of a locus is homozygous (diploid), the formula for population statistics calculation is $P(A)^2 + P(A) \times (1 - P(A)) \times \theta$ where P(A) is the effective frequency of allele A in use for the reference population, and θ is the population structure correction factor.

The method of calculation can be changed for a homozygous locus by applying the 2p rule in the locus detail box. For directions, see *Updating and Modifying Typed Results* on page 43.

The results of applying the 2p rule to a homozygous locus in the locus detail box is dependent on the population statistics calculation setting on the Population Group Settings page. When the setting is Use 2p, and the 2p rule is applied in the locus detail box, the calculation formula becomes 2 x P(A). In contrast, when the population statistics calculation setting is Use 2p - p^2, and the 2p rule is applied in the locus detail box, the calculation formula becomes 2 x P(A) - P(A)^2.

Changing the Source Attribution Threshold

The source attribution threshold is defined to facilitate visualization of population group statistics in a display bar. Population statistics that are less than the frequency of the source attribution threshold are colored on the results bar in orange. The threshold is visible on the display as a gray bar below the result.

There are 2 source attribution threshold settings, each corresponding with a particular method of population group statistics calculation. Calculation of population statistics is performed by the software as either random match probability or likelihood ratio methods. Use of the likelihood ratio method creates the inverse of the result retrieved from that of the random match probability method. By default, as corresponds with the random match probability method, the source attribution threshold is set to > 1.

NOTE
If you select Random Match Probability, the software automatically inverts the source attribution threshold setting. See *Changing the Source Attribution Threshold* on page 109.

To change the source attribution threshold in use by the system, access the Population Group Settings page from the Maintenance page. Perform the following steps.

1 If random match probability is in use for population statistics calculations, enter a value > 1 in the Source Attribution Threshold field. If likelihood ratio is in use, enter a value < 1.

2 Click **Save**. Click Cancel to return to the source attribution threshold you changed.

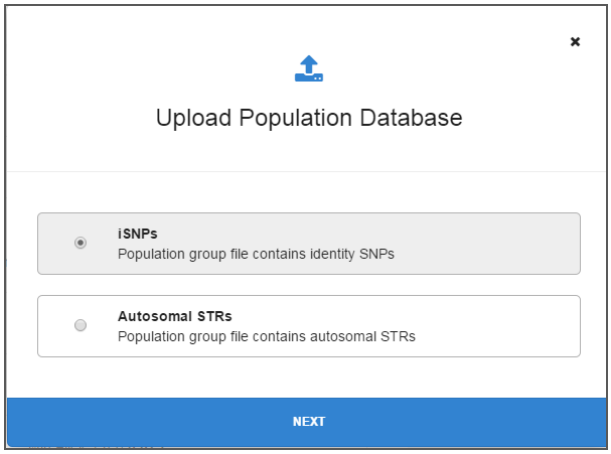
Adding Population Groups

In addition to the population groups installed with the software, you can add custom population groups of your own to use in calculating population statistics. To add a population group, access the Population Group Settings page from the Maintenance page.

1 Select the **Add Population Group** button in the Custom area of the page.

2 Click **iSNPs** or **Autosomal STRs** to define the population group you are adding.

Figure 117 Upload Population Database Page



3 Click **Next**.

4 Select a population group file for upload.

5 Click **Upload Population Database**.
When you upload the file, results of the file validation are displayed. If you receive an error message, in many cases you can still proceed with the upload, defining the population group without the samples with errors. See *Troubleshooting Population Group File Uploads* on page 140. To abandon uploading the population group, close the Upload Population Database window.

6 Click **Next**.

7 Enter a name for the population group in the Database Name field.
A default value is displayed in the Minimum Allele Frequency field. The default value is calculated from the population size using the equation Minimum Allele Frequency = 5/2N. N is the number of subjects in the group.

8 Click **Finish**.

Custom Population Group Requirements

To add a custom population group successfully, the software requires that the file conform to a particular format.

NOTE
Types of loci that are not supported for custom population group files include X STRs, Y STRs, aSNPs, pSNPs, and loci with 3 alleles.

Table 29 Requirements for Custom Population Group Files

Type of Requirement	Details
File Type	File contents are text tab-delimited (*.txt).
Type of STR Loci Supported	Autosomal STRs with genotype entries that contain numbers
Type of SNP Loci Supported	iSNPs with genotype entries that contain A, C, T, or G. Each genotype must contain exactly 2 alleles.
Homozygotes	Entry contains the allele 2 times. For example, enter A A, or 12 12.
Number of Subjects in the Population Group	> 3 and ≤ 2500
Locus Names	Match the name of the locus in the software.
Table Rows	The first row of the table contains the loci being defined. In the rows that follow, each subject identifier is unique to a row and is not repeated in other rows.
Table Columns	The first column of the table contains the subject identifiers for each row. Each subject must contain a genotype for each locus in the file.

Custom Population Group File Examples

There are 2 options for the entry of genotypes in custom population files. Make sure that the file is consistent, with each sample in the file entered in the table in the same way.

In the first table option, each allele is in an independent cell, so that each locus has 2 columns of data.

SampleCode	CFS1PO	CFS1PO	D10S1248	D10S1248	DS12S391	DS12S391
STRSample1	11	12	14	14	17	21
STRSample2	9	11	12	13	16	16
STRSample3	10	12	14	15	16	20
STRSample4	11	12	11	14	16	19
STRSample5	8	12	14	14	15	18

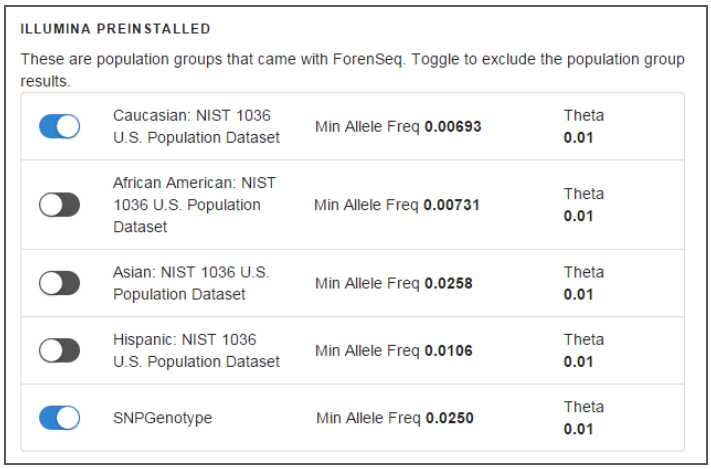
In the second table option, each locus is in a single column, with alleles separated by a comma.

SampleCode	rs1005533	rs10092491	rs1015250	rs1024116	rs1028528	rs1031825
SNPSample1	A,G	T,C	C,G	A,A	A,G	C,C
SNPSample2	G,A	T,T	G,C	A,A	G,A	A,C
SNPSample3	G,G	T,C	G,G	G,G	G,A	A,A
SNPSample4	G,A	C,T	C,G	A,G	A,G	C,C
SNPSample5	A,A	T,T	C,G	G,A	G,A	A,A

Selecting Population Groups

Population groups installed with the software and any population groups you defined for calculating population statistics are listed on the Population Group Settings page. The minimum allele frequency and Theta value is displayed next to each population group. By default, all population groups installed with the software are available for population statistics calculations.

Figure 118 Population Group Settings Page



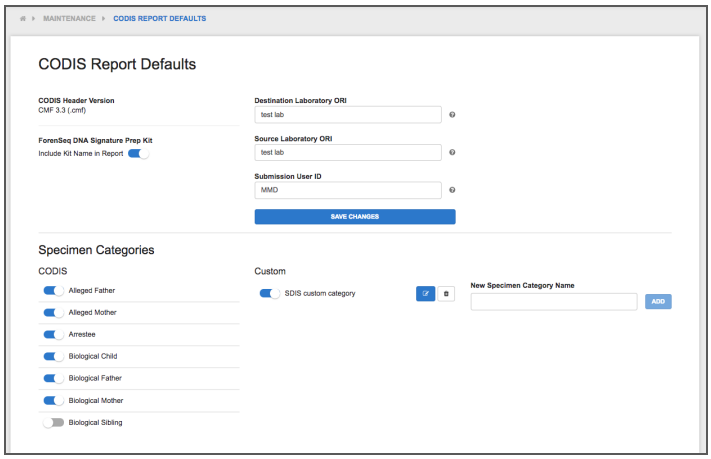
To deselect a population group from use for population statistic calculations, click the toggle switch next to the population group name so that the switch displays a gray background and is in position on the left side.

To select a population group for use for population statistic calculations, click the toggle switch next to the population name so that the switch displays a gray background and is in position on the left side.

CODIS Report Defaults

Values specified in CODIS Report Defaults pre-populate in the report generator workflow to expedite report generation. These default values can be changed at the time of report generation, if desired.

Figure 119 CODIS Report Defaults



CODIS report defaults include the following fields:

- Option to include the ForenSeq DNA Signature Prep Kit name
- Destination Laboratory ORI
- Source Laboratory ORI
- Submission User ID
- Configurable specimen categories

Specimen Category Selections

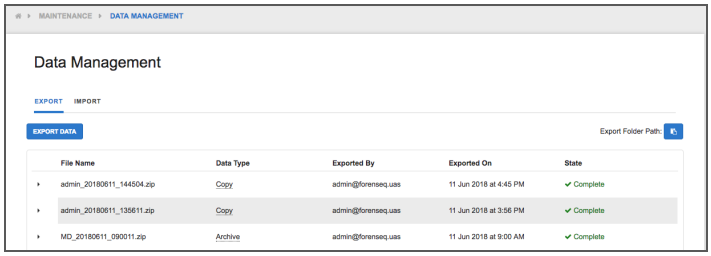
- To configure existing specimen category selections, use the toggle selections (blue position to the right indicates selected).
- To add a new custom specimen category, type the name of the category and select **Add**. The new category name appears under the Custom category header. To edit or remove Custom categories, select the edit or delete buttons next to the category name.

For more information about creating CODIS reports, see *Sample-Level CODIS Report* on page 87 and *Project-Level CODIS Report* on page 90.

Data Management

Importing or exporting data into or from the ForenSeq Universal Analysis Software is enabled through the Data Management functionality. Data can be exported for transfer to another ForenSeq software installation or for data archival. The exported data can then be imported into the ForenSeq software instance that generated the data package originally or to a different installation, including a third-party site.

Figure 120 Data Management Page



When exporting, data in the package can be copied or archived.

- Copying allows for data sharing across installations without removing any data from the ForenSeq Universal Analysis Software.
- Archiving removes the data from ForenSeq Universal Analysis Software, which can reclaim hard drive space and free up room for additional runs.

Exporting Data

The Data Management page has an EXPORT tab that shows a list of previously created data export files by File Name, Data Type, the user that created the data package (Exported By), date and time of the export (Exported On), Status (State), and a Summary of contents of the data export file.

To view the contents of a previously created data export, expand the details by clicking on the carat to the left of the File Name. A Summary pane provides details on the Runs and Analysis Items, and Configuration Options.

- The Runs section displays each run that was included in the data export.
- The Analysis Items section displays a hierarchical view of the samples included. To view the samples, expand each project and each analysis to view the samples for that analysis.
- The Configuration Options display the optional items that were selected for inclusion during the creation of the data export.

NOTE
When a run is included, all of the sequencing data for that run is added to the data export, including sequencing data for samples not included in the data export.

Data exports that have been created are stored on the ForenSeq Universal Analysis Software server. Selecting the Export Folder Path icon copies the path to the folder, where the data export file exists, to the clipboard for pasting in File Explorer on the ForenSeq Universal Analysis Software server.

NOTE
Due to file sizes, the physical export files (*.ZIP) must be accessed from the ForenSeq Universal Analysis Software server. Log on to the server and browse to the export path to place files for import or to retrieve data export files. All data export files are stored on the ForenSeq Universal Analysis Software server at D:\Illumina\ForenSeq\UAS\Data Files.

Use the export data creation wizard to facilitate data export creation.

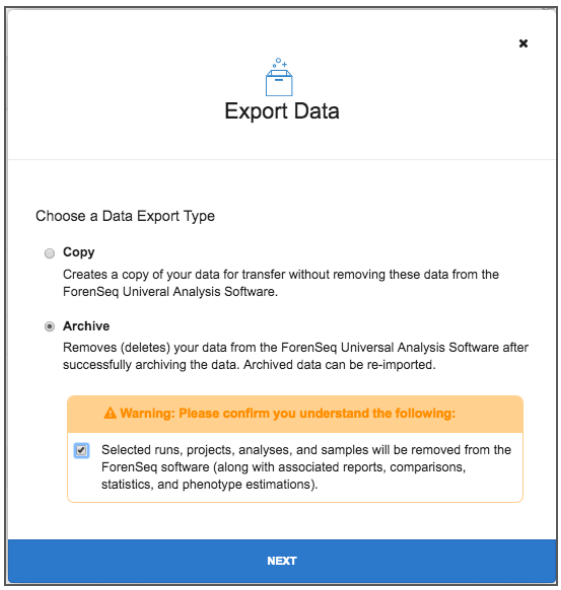
1 Click **EXPORT DATA** on the Data Management screen.

2 Select the type of data export to be created: **Copy** or **Archive**.

Table 30 Export Type Descriptions

Data Export Type	Description
Copy	Creates a copy of your data for transfer without removing these data from the ForenSeq Universal Analysis Software.
Archive	Removes (deletes) your data from the ForenSeq Universal Analysis Software after successfully archiving the data. Archived data can be re-imported.

Figure 121 Data Export Type Selection



When creating an Archive data export type, a confirmation displays requiring an acknowledgment that, after successful creation of the data export, the data will be removed from the ForenSeq software and server.

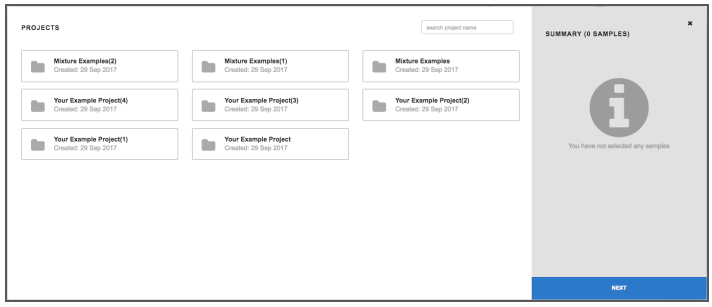
NOTE
When creating an Archive data export type and only a portion of the samples in the run are selected, the run is ***not*** removed from the ForenSeq Universal Analysis Software as it is required to reanalyze the included samples.

3 Select **Next** to proceed.

4 Identify the samples for inclusion in the data export. A list of available projects displays.

- Select a project from the list to display the analyses present in the project. Projects with samples already added to the data export are designated with a green checkmark. Upon selection of a project, a list of available analyses in the project displays.

Figure 122 Export Project Selection

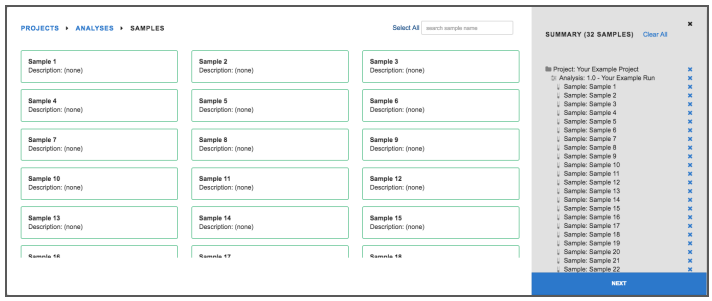


- Select an analysis from the list to view the samples in that analysis. Analyses with at least one sample added to the data export are designated with a green checkmark. Alternatively, the Select All option can be used to select all items (samples and analyses) in the project or to deselect all items. Upon selection of a single analysis, a list of available samples in the analysis displays.

Figure 123 Export Analysis Selection

- Select the samples from the list to add to the data export. Samples added to the data export are bordered in green. Alternatively, the Select All option can be used to select all samples for addition to the data export or to deselect all items.

Figure 124 Export Sample Selection and Summary



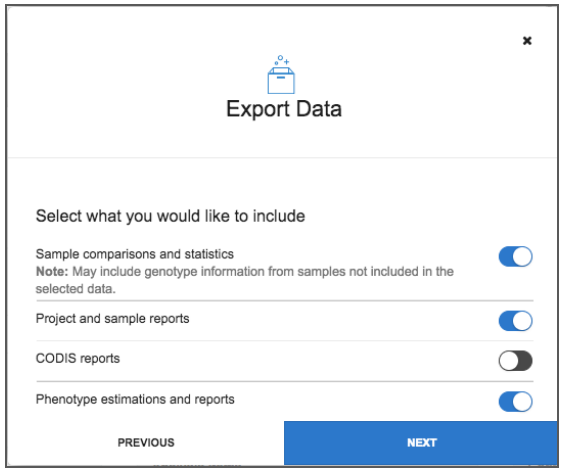
- To remove a project, analysis, or sample from the data export, select the X next to the name of the item. Removing a project from the data export removes all analyses and samples in the selected project. Removing an analysis from the data export removes all samples in the selected analysis. The Clear All section removes all projects, analyses, and samples from the data export.
- As Projects, Analyses, and Samples are added, the summary section on the right displays a summary of the data selected to be included in the data export file. The summary can also be used to clear all added samples or remove individual samples.

5 Select **Next** to proceed.

6 Identify additional analyses and reports for inclusion in the data export.

Each option is presented with a toggle that directs elements being included or excluded.

Figure 125 Export Configuration Options



- When selected (blue toggle display), the Sample comparisons and statistics option includes all comparisons or population statistics calculations performed on any selected samples.
- When selected (blue toggle display), the Project and sample reports option includes all Project Genotype, Sample Details, and Sample Summary reports generated for each selected project or sample. Flanking region reports are not included.
- When selected (blue toggle display), the CODIS reports option includes all CODIS reports generated for each selected sample.
- When selected (blue toggle display), the Phenotype estimations and reports option includes phenotype data and generated phenotype estimation reports for each selected sample.

NOTE
If export configuration items are excluded from the data export, then those items will not appear in the Sample History activity log upon import.

See the following sections for more information on the content of these analyses and reports:

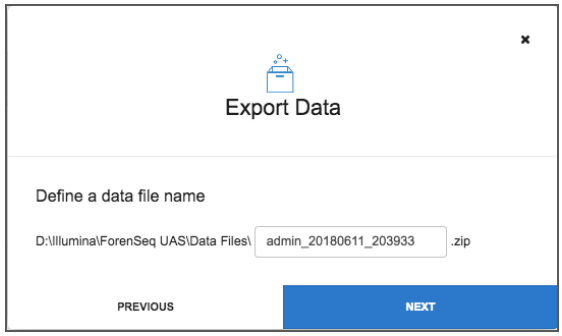
- *Project-Level Genotype Report* on page 71
- *Sample Genotype Report* on page 78
- *Sample-Level CODIS Report* on page 87
- *Project-Level CODIS Report* on page 90
- *Phenotype Estimation Report* on page 83

7 Select **Next** to proceed.

8 Provide a name for the data export file.
A default name is provided that uniquely identifies the data export file. The default name is a concatenation of the user name, date, and time. To modify the name of the data export file, replace the name in the text box.

NOTE
When changing the data export file name, take care to maintain uniqueness of the name. The data export file name must be unique to the ForenSeq Universal Analysis Software installation.

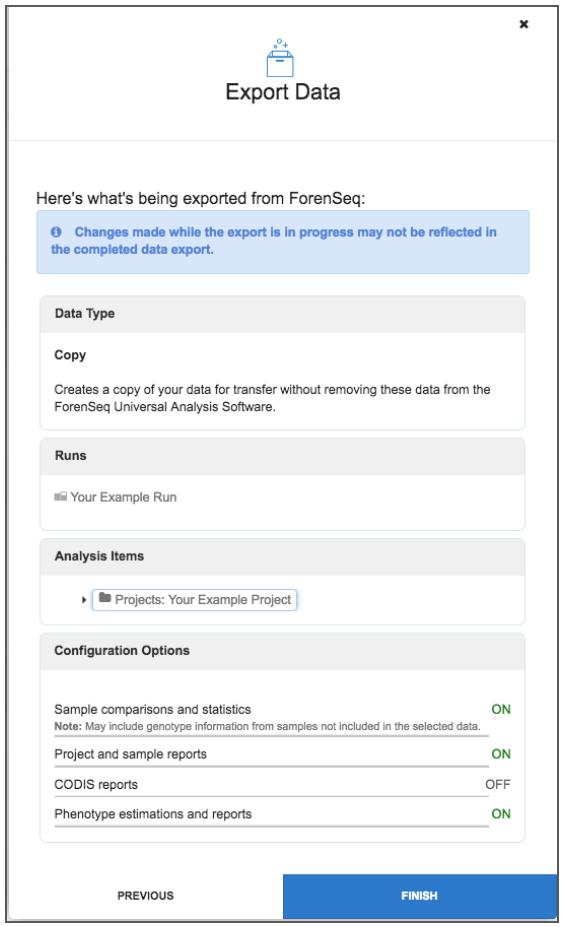
Figure 126 Export File Name



9 Select **Next** to proceed.

10 Review the final summary of data export selections.
The summary includes all selections from the first three steps of the wizard: 1) Data Export Type (Copy or Archive), 2) Runs and Analysis Items, and 3) Configuration Options.

Figure 127 Export Summary



11 Select **Finish** to create the data export file.
The dialog closes and the data export creation begins. During export file creation, the Data Management page displays the status Processing.

NOTE
During data export creation, other activities can be performed within the ForenSeq software. However, if changes are made to the data included in the export file while an export file is being created, those changes may not be reflected in the data export.

Depending on the contents, the data export creation can take several minutes to complete. Upon creation of the data export, the status updates to Complete on the Data Management page.
The data export file can be accessed on the ForenSeq Universal Analysis Software server at D:\Illumina\ForenSeq\UAS\Data Files.

NOTE
Open the *.zip file to view the contents of the data export file. View the exportContents.txt file in a text viewer.

NOTE
In the data export file, all user names are converted to exportuser@forenseq.uas.

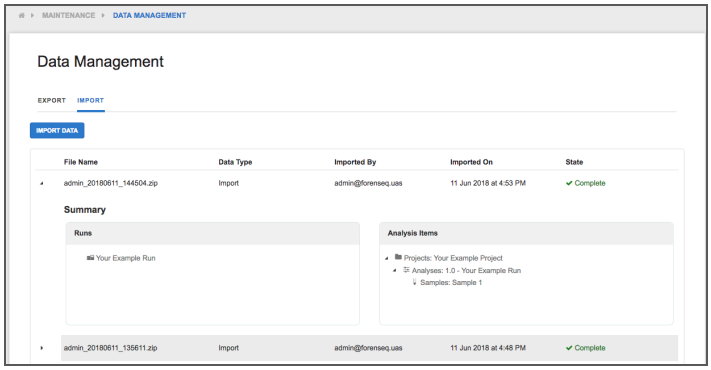
Importing Data

The Data Management page has an IMPORT tab that shows a list of previously imported data by File Name, Data Type, the user responsible for the import (Imported By), the date and time of the import (Imported On), Status (State), and a Summary of contents of the data file.

To view the contents of a previously imported data file, expand the details by clicking on the carat to the left of the data file name. A Summary pane provides details on the Runs and Analysis Items.

- The Runs section displays each run that was included in the data file.
- The Analysis Items section displays a hierarchical view to the samples included.

Figure 128 Data Management Page

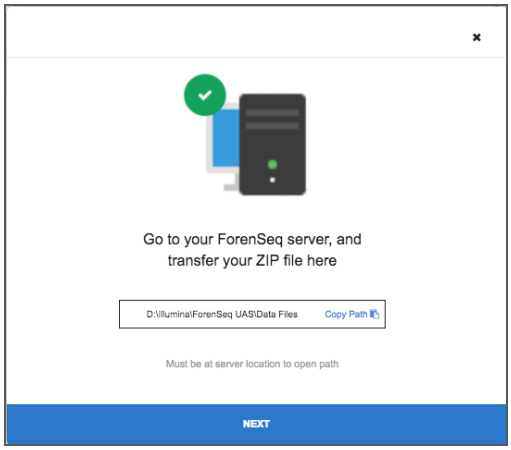


Use the data import wizard to facilitate data import.

1 Click **IMPORT DATA** on the Data Management screen.

2 Determine a location on the ForenSeq Universal Analysis Software server to place the data file upon import.
A dialog opens to define the location for the data import. Import to the location D:\Illumina\ForenSeq\UAS\Data Files.

Figure 129 Import File Location

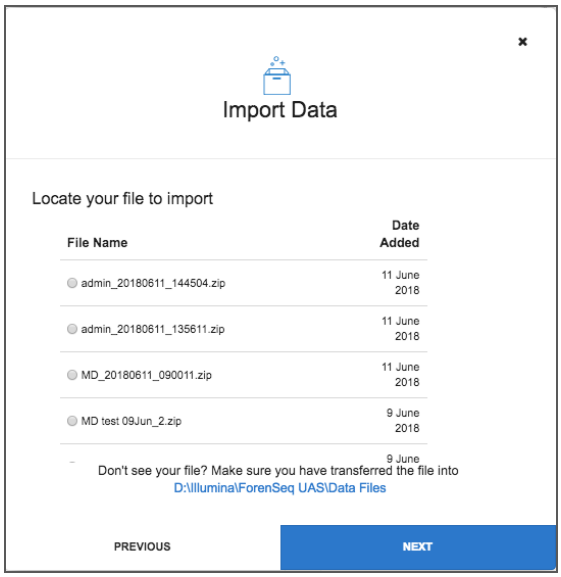


3 Select **Next** to proceed.

4 Identify the data file for import.

- A list of data files at the specified location displays. Select the data file to be imported.
- In the event that a project name already exists in the system, the imported data is added to a new project with the same project name, but with a sequential, numeric identifier appended to the name, e.g., Your Example Project(1).

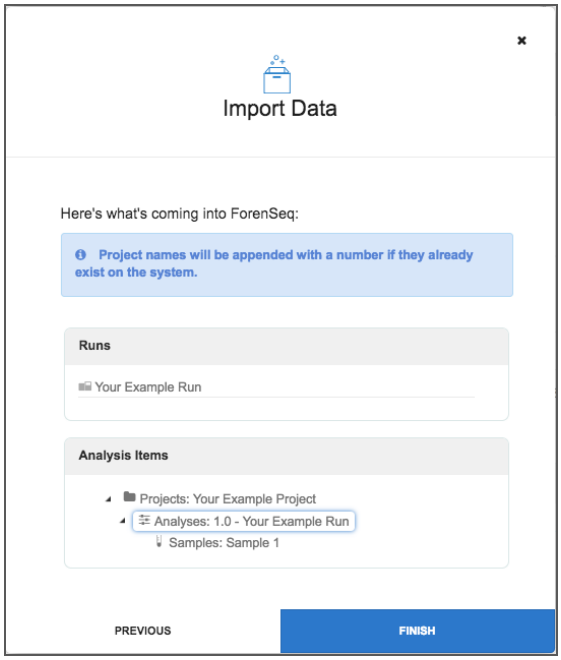
Figure 130 Import File Selection



5 Select **Next** to proceed.

6 Confirm the contents of the selected data file before import.

Figure 131 Import Summary



7 Select **Finish** to import the data file.
The dialog closes and the data import initiates. During import, the Data Management page displays the status Processing.
During the import, other activities may be performed within the ForenSeq software. Depending on the contents, the import can take several minutes to complete. Upon completion, the status updates to Complete on the Data Management page.

Analysis Metrics and Procedures

Chapter 8

Introduction

ForenSeq Universal Analysis Software performs a series of analysis steps on data from the RTA software to obtain STR and SNP information from samples sequenced on the MiSeq FGx instrument.

Analysis Metrics

During the sequencing run, RTA generates data files that include analysis metrics used by ForenSeq Universal Analysis Software. Metrics that appear in the ForenSeq Universal Analysis Software are clusters passing filters, base call quality scores, and phasing and prephasing values.

Clusters Passing Filter

During analysis, RTA filters raw data to remove any reads that do not meet the overall quality as measured by the Illumina chastity filter. The chastity of a base call is calculated as the ratio of the brightest intensity divided by the sum of the brightest and second brightest intensities.

Clusters pass filter (PF) when no more than 1 base call in the first 25 cycles has a chastity of < 0.6.

Quality Scores

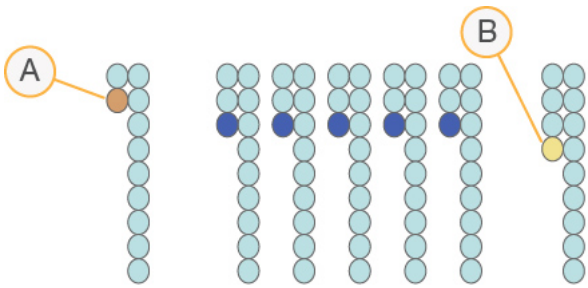
For each read, the quality scores are compared against benchmark values for the assay. If the average quality of the read is lower than the benchmark value, an orange indicator is provided.

During the sequencing run, base call quality scores are calculated during acquisition.

Phasing and Prephasing

During the sequencing reaction, each DNA strand in a cluster extends by 1 base per cycle. A small portion of strands might become out of phase with the current incorporation cycle, either falling a base behind (phasing) or jumping a base ahead (prephasing). Phasing and prephasing rates indicate an estimate of the fraction of molecules that became phased or prephased in each cycle.

Figure 132 Phasing and Prephasing



A Read with a base that is phasing
B Read with a base that is prephasing

The number of cycles performed in a read is 1 more cycle than the number of cycles analyzed.

Analysis Procedures

ForenSeq Universal Analysis Software performs analysis on sequencing data generated on the MiSeq FGx from sample libraries prepared with the ForenSeq DNA Signature Prep Kit. Analysis procedures include demultiplexing, sequence alignment, allele counting, and genotype calling. Quality control indicators address aspects such as DNA mixture detection and genotype call quality.

The ForenSeq Universal Analysis Software addresses STR and SNP loci amplified in either of the two ForenSeq DNA primer mixes in the ForenSeq DNA Signature Prep Kit. Analysis is performed on amplified and sequenced DNA for STR and SNP loci, including the following:

- Autosomal STRs
- X-STRs
- Y-STRs
- iSNPs
- pSNPs
- aSNPs

Analysis is geared toward single-source DNA samples as well as challenging forensic samples. The target type is determined to be a SNP or a STR for sequence alignment purposes. Each read is then aligned to the hg19 human reference sequences that correspond to the loci evaluated in the primer pool to determine the locus and the repeat length.

After alignment, potential alleles are counted from the numbers of reads. Counts are used to detect if the sample is a mixture, call genotypes, and assign quality control indicators. Quality control indicators can help in mixture detection and sample interpretation.

NOTE
For more information on hg19 sequences, see http://genome.ucsc.edu/cgi-bin/hgGateway.

Demultiplexing

Demultiplexing is the first step in post-sequencing analysis. Demultiplexing separates data among pooled samples based on short index sequences that tag samples from different libraries. Each Index Read sequence is compared to the index sequences specified in sample information entered on the Create a New Run page. The ForenSeq kit enables pooling of many i7 and i5 index pairs into the same run.

Alignment of STRs

Each read is evaluated against all the possible loci in the primer pool to determine its source. If a read can be aligned to multiple loci, a scoring mechanism is employed and the read is attributed to the highest scoring locus.

Evaluation of the read is done by identifying the start and stop of the repeat on both sides of the read as compared to the reference sequence. In the case of STRs, the read is then further evaluated to identify the repeated region of the read for additional allele calling refinement.

Alignment of SNPs

Each read is evaluated against all the possible loci in the primer pool to determine its source. If a read can be aligned to multiple loci, a scoring mechanism is employed and the

read is attributed to the highest scoring locus. Sequence-based variation in SNPs is not based on repetitive elements as they are for STRs.

Allele Counting

The software determines the length of the STR sequence between the forward and reverse PCR primers, and the number of repeats within that sequence. In addition to determining the length of the sequence, an analysis algorithm accounts for the potential that STR alleles of the same length can vary in sequence, therefore representing 2 different alleles.

STR Genotype Calling

Genotype calling of an STR occurs as the ForenSeq Universal Analysis Software considers signal at a target. The analysis algorithm uses allele counts and several parameters to call (type) an allele.

A default baseline value, or fixed detection threshold, of 10 reads per locus is used in the ForenSeq Universal Analysis Software. When an STR or SNP allele sequence is present at 10 or fewer total reads, the sequence is not reported as the baseline value was not exceeded; when a sequence is present at 11 or more total reads, the sequence is reported as a possible allele.

NOTE
Low-level, background signal was assessed on the MiSeq FGx using the ForenSeq DNA Signature Prep Kit to estimate the lower limit of reads required to assign a sequence as an allele. No-template controls (NTCs) provided an indicator of the default baseline value in the ForenSeq Universal Analysis Software. Ninety-six NTC reactions (water only) were prepared using ForenSeq DNA Primer Mix B (DPMB) with 230 STR and SNP loci (plus Amelogenin) evaluated. Data indicated a mean number of reads per locus in NTCs of 0.15, with a standard deviation of three reads across the 22,176 loci evaluated. The baseline value was then determined as the mean number of reads plus three standard deviations (9.15) and rounding up to the nearest integer (10).

First, analytical and interpretation thresholds are determined through the application of the supplied threshold percentage values. The user-defined percentage value is multiplied by the sum of read counts at a locus, respectively. In cases of low coverage, a minimum number of reads is used (650) as the number of reads for the locus in the determination of thresholds.

After automated STR allele calling, the gender of the sample is determined based on the signal at the X and Y loci. The software uses gender designation at the Y STR loci and X STR loci to dictate the expected copy number at these loci. The number of alleles typed at a locus is important in setting genotype quality control indicators such as allele count. Additionally, based on the gender-informed copy number expectation, a single typed allele at an X STR locus will be depicted as a homozygote (e.g., 15,15) for an XX sample, and as a hemizygote (e.g., 15) for an XY sample. By default, negative amplification control samples and reagent blank samples are not assessed for gender, displaying Inconclusive in place of XX or XY.

Table 31 Gender Determination

Inconclusive	XX	XY
Negative control samples are always assigned a gender of inconclusive. A sample is also assigned a gender of inconclusive if it meets either of the two conditions: • Of the X STR loci in the sample, less than 3 loci have signal above the analytical threshold. • Less than 3 Y STR loci are active in the analysis.	For a sample to receive the designation of gender XX, all of the following conditions must apply: • The sample is not a negative control. • Of the X STR loci in the sample, at least 3 loci have signal above the analytical threshold. • Less than 3 Y STR loci have signal greater than the analytical threshold, but at least 3 loci are active in the analysis.	For a sample to receive the designation of gender XY, all of the following conditions must apply: • The sample is not a negative control. • Of the X STR loci in the sample, at least 3 loci have signal above the analytical threshold. • Of the Y STR loci in the sample, at least 3 loci have signal above the analytical threshold.

After gender-calling, the software checks the read counts for the following quality considerations:

- **Low Coverage**—If the allele with the highest reads does not meet the effective interpretation threshold, the locus is not processed further. The low coverage indicator can be seen in results from version 1.1 or earlier of the ForenSeq Universal Analysis Software.
- **Stutter Filtering**—Stutter is computed as sequence stutter, where the software checks the repeat units in the STR sequence. This computation can enable the separation of a minor contributor allele from stutter when the sequences differ.

  A sequence is considered stutter, with offset *k* repeat units if:
  - The sequences differ in length by *k* times the reference repeat unit length
  - The number of repeat units in the potential stutter are identical to an adjacent unit of length of the reference repeat unit
  - The are no (other) mismatches between the sequences

  In the case of tetra- and penta-nucleotide repeats, the reference repeat unit lengths are 4 and 5 respectively. The software assesses differences in the sequences for *k* = -2, -1 and +1 repeat units. That is, with n the length in base pairs of a given allele, the software assesses differences for sequences at the n-8 or n-10, n-4 or n-5, and n+4 or n+5 positions for tetra- and penta-nucleotide repeats, respectively.

  The stutter filter setting depends on the differences in repeat units. The stutter filter setting is applied as is for the n-4 and n-5 positions. The stutter filter setting is squared it at the n-8 or n-10 and the n+4 or n+5 positions.

  The expected stutter intensity is the product of the stutter filter setting and the called allele intensity at a repeat length. For example, a stutter filter setting of 10% (0.1) is squared (0.1 x 0.1 = 0.01 or 1%) before multiplying by the intensity of the called parent allele.

  A stutter quality control indicator for elevated stutter is displayed when both of these conditions exist:
  - Uncalled read intensity, with sequence of a potential stutter of a called allele
  - Uncalled read intensity is greater than the maximum expected, user-defined stutter % of the potential parent allele
- **Interpretation Threshold**—If the read count of any DNA sequence has a read count of a non-stutter DNA sequence between the analytical and interpretation threshold, an interpretation threshold quality indicator icon appears.

- **Allele Count**—Indicated when the signal from possible non-stutter alleles is greater than the analytical threshold, and is greater than the copy number of the locus. The copy number at the locus is defined by the expected zygosity of the locus.
- **Analytical Threshold**—When a locus has signal below the analytical threshold and no alleles exceed the interpretation threshold, an analytical threshold quality indicator icon appears.
- **Not Detected**—When no signal is present at the locus, a not detected quality control indicator quality indicator icon appears.
- **Imbalance**—If the calculated intralocus balance is below the intralocus balance threshold, and the locus is not homozygous, an imbalance quality control indicator is triggered.
- **Mixture Detection**—A sample is indicated as a possible mixture when the total number of loci that are assigned the many alleles quality control indicator is > 5.

SNP Genotype Calling

Genotype calling of a SNP occurs as the ForenSeq Universal Analysis Software determines that DNA sequence at a targeted locus passes quality control thresholds. The analysis algorithm uses allele counts and several parameters to call (type) an allele.

Then the software checks the read counts for the following quality considerations:

- **Per Locus QC**—A quality check for each locus is performed. Based on the assessment, a preliminary genotype is assigned based on the signal intensity for each allele.
- **Interpretation Threshold**—If the read count of either allele present is between the analytical and interpretation thresholds, an interpretation quality indicator icon appears. The interpretation threshold percent is defined in the analysis settings.
- **Imbalance**—If the count of the reference allele divided by the count of the alternate allele is less that the user-defined imbalance ratio, and the locus is not homozygous, then it is indicated as imbalanced.
- **Not Detected**—When no signal is present at the locus, a not detected quality control indicator quality indicator icon appears.
- **Mixture Detection**—A sample is indicated as a possible mixture when the total number of loci that are imbalanced is > 10.
- **Analytical Threshold**—When a locus has signal below the analytical threshold and no alleles exceed the interpretation threshold, an analytical threshold quality indicator icon appears

Troubleshooting

Introduction

This chapter features troubleshooting tips for ForenSeq sequencing runs, analysis errors, and population group file uploads. Most of the troubleshooting recommendations for ForenSeq sequencing runs include library prep steps, which are detailed in the *ForenSeq DNA Signature Prep Reference Guide (document # VD2018005)*.

Troubleshooting ForenSeq Runs

Table 32 ForenSeq Run Issues

Problem	Resolution
Sample Intensity (Number of Reads) Low	If sample intensity for a sample is low, this condition can indicate a problem with sample quantity or sample quality. 1 Make sure that DNA concentrations are not below the recommended amount. 2 Depending on the source, purify the DNA or dilute, and then sequence the sample again. 3 Load fewer samples and sequence the samples again. Make sure to maintain a minimum of 8 samples in the run.

Problem	Resolution
Cluster Density High	Review the Run Metrics tab. If cluster density for the run is high, but the reads are above the sample read count threshold in the Sample Representation tab, proceed with analysis. If there is insufficient data for analysis, it can be due to 1 of the following reasons. ▸ Sample from the Purified Library Plate (PLP) or the Library Normalization Beads 1 (LNB1) reagent was over-pipetted during normalization. a If the sample has a high number of alleles with quality control indicators, then rerun the sample on a new sequencing run, and confirm that correct volumes and reagents are used for normalization. b If the volume of sample used is correct, consider further diluting the PCR product in the Pooled Normalization Libraries (PNL) tube. c If normalization volumes or reagents were not correct, redo the *Normalize Libraries* process and resequence. ▸ Too much adapter dimer in library prep. a Run 1–2 µl of the PCR product in the PLP plate on a Bioanalyzer or Fragment Analyzer and check the amount of primer dimer. Contact Technical Support with any questions. b If there is more than 5 % adapter dimer in the PCR product you check, redo the *Purify Libraries* process on the remaining PCR product in the PLP plate using the 30 µl remaining in each well. Adjust the *Purify Libraries* process to use 30 µl SPB. Add 32.5 µl Resuspension Buffer (RSB) and transfer 30 µl to a new 96-well PCR plate. Proceed with the *Normalize Libraries* process. c [Optional] If SPB are limited, purify the PCR product in the PLP plate using a standard column cleanup method. Use manufacturer instructions. ▸ Not enough Hybridization Buffer (HT1) added, or too much normalized library added to HT1 when preparing libraries. If the sample has a high number of alleles with quality control indicators, then rerun the sample, on a new sequencing run, making sure the normalized library is sufficiently diluted.

Problem	Resolution
Cluster Density Low	Review the Run Metrics tab. If cluster density for the run is low, but the reads are above the sample read count threshold in the Sample Representation tab, proceed with analysis. If the sample has an insufficient number of reads due to low density, it is likely due to 1 of the following reasons: ▸ During library prep, DNA input was too low or overly degraded, or the DNA dilution was incorrect. If possible, repeat library prep with more DNA. ▸ During library prep, a critical reagent was not added. Repeat library prep. ▸ LNB1 was not pipetted sufficiently during normalization. Consider diluting the PCR product in the PNL tube less. ▸ HT1 was over-pipetted when preparing libraries. Repeat sequencing using the correct volume. ▸ Excessive sample was lost during SPB purification. Repeat library prep using bead-handling best practices. ▸ Sample not denatured properly. a Make sure that the final library pool is denatured with HP3. b Make sure that the HP3 pH is above 12.5. c Make sure that the final library pool is heat denatured. ▸ Instrument blockage. a Perform a maintenance wash and repeat sequencing. b Make sure to refill the wash tray and bottle for every wash. c Make all MiSeq FGx wash solutions daily.
Percentage of Clusters Passing Filter Low	If the percentage of clusters passing filter is too low, it is likely due to 1 of the following reasons: ▸ Cluster density is too high. See troubleshooting for Cluster Density High. ▸ Phasing or prephasing is too high. a Perform a maintenance wash and repeat sequencing. b Make sure to refill the wash tray and bottle for every wash. c Make all MiSeq FGx wash solutions daily. ▸ If the problem is still not resolved, it could be that reagents are not performing as expected. Contact Technical Support.
Phasing or Prephasing High	If phasing or prephasing is too high, it is likely due to 1 of the following reasons: ▸ Cluster density is too high. See troubleshooting for Cluster Density High. ▸ Instrument blockage. a Perform a maintenance wash and repeat sequencing. b Make sure to refill the wash tray and bottle for every wash. c Make all MiSeq FGx wash solutions daily. ▸ If the problem is still not resolved, it could be that reagents are not performing as expected. Contact Technical Support.

Problem	Resolution
Run Does Not Complete	If the sequencing run stops before completion, it is likely due to 1 of the following reasons: ▸ When preparing libraries, HSC was not added. Add HSC according to the *Denature and Dilute Libraries* process and repeat the sequencing run. ▸ Cluster density is too high. See troubleshooting for Cluster Density High. ▸ If an error message is displayed, contact Technical Support.
Read 1 or Read 2 Quality Scores Low	If quality scores for Read 1 or Read 2 are too low, it is likely due to 1 of the following reasons: ▸ Phasing and prephasing are too high. See troubleshooting for Phasing and Prephasing High. ▸ Cluster density is too high. See troubleshooting for Cluster Density High. ▸ If the problem is still not resolved, it could be that reagents are not performing as expected. Contact Technical Support.
Index 1 or Index 2 Quality Scores Low	If quality scores for Index 1 or Index 2 are too low, it is likely due to 1 of the following reasons: ▸ Phasing and prephasing are too high. See troubleshooting for Phasing and Prephasing High. ▸ Cluster density is too high. See troubleshooting for Cluster Density High. ▸ There was low diversity in the Index Read because not enough samples were sequenced. Repeat sequencing with a minimum of 8 samples. ▸ If the problem is still not resolved, it could be that reagents are not performing as expected. Contact Technical Support.
Reads Per Sample (Intensity) Low	Review the Sample Representation tab. If the samples are above the sample read count threshold, then proceed with analysis. If reads per sample are below the sample read count threshold, it is likely due to 1 of the following reasons: ▸ During library prep, DNA input was too low, or the DNA dilution was incorrect. If possible, repeat library prep with more DNA. ▸ During library prep, a critical reagent was not added. Repeat library prep. ▸ LNB1 was not pipetted sufficiently during normalization. a If the sample has a high number of alleles with quality control indicators, then rerun the sample on a new sequencing run. b If the sample has a low number of alleles with quality control indicators, consider diluting the PNL less and rerun the samples on a new sequencing run. c If the run does not pass, redo the *Normalize Libraries* process. ▸ HT1 was over-pipetted when preparing libraries. Redo the *Normalize Libraries* process. ▸ The PCR product was not heat-denatured when preparing libraries. Perform the denature heating step on the Diluted Normalized Libraries (DNL) tube and repeat the sequencing run.

Problem	Resolution
Reads Per Sample (Intensity) for Human Sequencing Control Low	Review the Sample Representation tab. If the samples are above the sample read count threshold, then proceed with analysis. If reads per sample are below the sample read count threshold, it is likely due to 1 of the following reasons: ▸ When preparing libraries, HSC was not added. Add HSC according to the *Denature and Dilute Libraries* process and repeat the sequencing run. ▸ HT1 was over-pipetted when preparing libraries. Repeat the sequencing using the correct volume. ▸ The sample was not heat-denatured when preparing the sample for sequencing. Perform the denature heating step and repeat the sequencing run. ▸ HSC was not denatured with HP3 when preparing the sample for sequencing. Make sure that the HSC is denatured with HP3 before adding it to the library, and repeat the sequencing run.

Troubleshooting Analysis Errors

Table 33 Analysis Errors

Problem	Resolution
Analysis Error Message	When you click Create New Analysis, and you receive an analysis error message, it is likely due to 1 of the following reasons: ▸ The ForenSeq Analysis service is not running. a On the ForenSeq Universal Analysis Software server desktop, click the Windows icon in the lower-left corner of the screen. b Enter **Services** to find the Services application. c Open the Services application. d Locate ForenSeq Analysis in the list of services. e If the status of the service is not Running, right-click on the service and select **Start**. f After the service has started, try to the Create New Analysis command again. ▸ The ForenSeq Analysis service is running, but the system cannot locate the run data to perform an analysis. a On the ForenSeq Universal Analysis Software server, open the log file for the analysis service at (C:\Illumina\Forenseq UAS\Analysis\logs\application.log). b Find the message, `EXCEPTION OCCURRED:System.Exception` in the log file to confirm the problem is that the system cannot locate the run data. c If the run repository is stored on a networked location, confirm the network is running. d Navigate to the run folder listed in the log file as the RunStoragePath to confirm that the data exists in the expected location. e Try the Create New Analysis command again.

Problem	Resolution
All Q Icons on Dashboard are Gray	When all Q icons on the dashboard are gray, and no information is available when you hover over the icons, this condition can indicate the following: ▸ The system cannot locate the run data. a Click the Windows icon in the lower-left corner of the screen. b Enter **Services** to find the Services application. c Open the Services application. d Locate ForenSeq Analysis in the list of services. e If the status of the service is not Running, right-click on the service and select **Start**. ▸ The run repository is stored on a networked location, and the network is not connected. a Navigate to the run folder described in the configuration file. b Confirm that data exists in the expected location. c Confirm that no permission changes were made to access the run folder location.

Troubleshooting Population Group File Uploads

Table 34 Population Group File Upload Errors

Error Message	Item Affected	Interpretation and Possible Resolution
File format is unrecognized by the system	File	Unable to proceed because the system cannot recognize the provided data. Make sure that the file is text tab-delimited (*.txt).
Population group definition does not meet minimum or maximum size requirements	File	Unable to proceed because the population group is too small or too large for the system. Make sure that the population group is > 3 and ≤ 2500.
Locus does not match an expected locus name	Locus	The locus is not defined in the population group because system results cannot be correlated to the locus. Make sure that loci names are consistent with loci names in the software.
Number of alleles defined for the locus is not consistent with expectations	Locus	The locus is not defined in the population group because the number of alleles is not equal to 2. Review allele entries.
Invalid chromosome type	Locus	The locus is not defined in the population group because it is not an autosomal STR. Remove any X STRs and Y STRs from the population group.
Allele name is not recognized	Subject	The subject results are not included in the population group because the allele name is inconsistent with expectations. If your file contains STRs, make sure that entries contain repeat numbers and no letters. If your file contains SNPs, make sure that entries contain letters, and no repeat numbers.
Data has an incomplete profile	Subject	The subject results are not included in the population group because results are missing for at least 1 locus. Make sure that each subject has a genotype for each locus in the file.
Subject name is not unique to the population group	Subject	The subject results are not included in the population group because there are 2 or more rows with the same sample identifier. Remove duplicate sample identifiers.

Troubleshooting Data Import and Export

Table 35 Data Import and Export Errors

Problem	Resolution
Data import or export state displays error message	Retry the import or export. If trying to export a large amount of run data, consider a smaller file size. If an additional error message displays, contact Technical Support.

Supporting Information

Appendix A

Human Sequencing Control Loci

The following is a list of loci that are contained in human sequencing control with allele lengths.

Table 36 Human Sequencing Control Loci and Lengths

Locus	Allele Length
D3S1358	136
D5S818	68
D7S820	93
D8S1179	61
D13S317	118
D16S539	118
D18S51	131
D21S11	178
FGA	126
PentaE	340
TPOX	45
DYF387S1	185
DYS391	88
DYS392	297
Y-GATA-H4	120
DXS7423	145
DXS10074	193
DXS10103	129
PentaD	220
DYS448	311
DYS460	303

Autosomal, Y, and X STR Filters and Thresholds

The following is a list of autosomal, Y, and X STRs with values for stutter filter, analytical, and interpretation thresholds.

Table 37 STR Locus Thresholds

Loci	% Stutter	% Analytical	% Interpretation
Amelogenin	0	> 1.5	> 4.5
CSF1PO	< 10	> 1.5	> 4.5
D1S1656	< 25	> 1.5	> 4.5
D2S441	< 7.5	> 1.5	> 4.5
D2S1338	< 20	> 1.5	> 4.5
D3S1358	< 15	> 1.5	> 4.5
D4S2408	< 7.5	> 1.5	> 4.5
D5S818	< 12.5	> 1.5	> 4.5
D6S1043	< 12.5	> 1.5	> 4.5
D7S820	< 10	> 1.5	> 4.5
D8S1179	< 25	> 1.5	> 4.5
D9S1122	< 12.5	> 1.5	> 4.5
D10S1248	< 20	> 1.5	> 4.5
D12S391	< 33	> 1.5	> 4.5
D13S317	< 12.5	> 1.5	> 4.5
D16S539	< 20	> 1.5	> 4.5
D17S1301	< 20	> 1.5	> 4.5
D18S51	< 22	> 1.5	> 4.5
D19S433	< 12.5	> 1.5	> 4.5
D20S482	< 15	> 1.5	> 4.5
D21S11	< 10	> 1.5	> 4.5
D22S1045*	< 20	> 1.5	> 4.5
FGA	< 25	> 1.5	> 4.5
PentaD	< 7.5	> 1.5	> 4.5
PentaE	< 10	> 1.5	> 4.5
TH01	< 10	> 1.5	> 4.5
TPOX	< 10	> 1.5	> 4.5

Loci	% Stutter	% Analytical	% Interpretation
vWA	< 22	> 1.5	> 4.5
DYS19	< 15	> 1.5	> 4.5
DYS385a-b	< 20	> 1.5	> 4.5
DYF387S1	< 20	> 1.5	> 4.5
DYS389I	< 20	> 1.5	> 4.5
DYS389II	< 35	> 5	> 15
DYS390	< 15	> 1.5	> 4.5
DYS391	< 20	> 1.5	> 4.5
DYS392*	< 30	> 1.5	> 4.5
DYS437	< 45	> 1.5	> 4.5
DYS438	< 15	> 1.5	> 4.5
DYS439	< 15	> 1.5	> 4.5
DYS448	< 15	> 3.3	> 10
DYS460	< 15	> 1.5	> 4.5
DYS481	< 50	> 1.5	> 4.5
DYS505	< 15	> 1.5	> 4.5
DYS522	< 15	> 1.5	> 4.5
DYS533	< 15	> 1.5	> 4.5
DYS549	< 22	> 1.5	> 4.5
DYS570	< 22	> 1.5	> 4.5
DYS576	< 15	> 1.5	> 4.5
DYS612[1]	< 35	> 1.5	> 4.5
DYS635	< 15	> 3.3	> 10
DYS643	< 20	> 1.5	> 4.5
Y-GATA-H4[2,3]	< 35	> 1.5	> 4.5
HPRTB	< 15	> 1.5	> 4.5
DXS7132	< 22	> 1.5	> 4.5
DXS7423	< 15	> 1.5	> 4.5
DXS8378	< 15	> 1.5	> 4.5
DXS10074	< 25	> 1.5	> 4.5
DXS10103	< 22	> 1.5	> 4.5
DXS10135	< 22	> 1.5	> 4.5

* Interpret loci D22S1045 and DYS392 with caution. Elevated n-1 repeat stutter might be observed, particularly with decreased marker coverage. Heterozygote imbalance might be observed at locus D22S1045 regardless of marker coverage. Consider multilocus genotype when determining the presence of a DNA mixture. See the *ForenSeq DNA Signature Prep Reference Guide (document # VD2018005)* for additional information and interpretation examples.

Resources

1. D'Amato ME, Ehrenreich L, Cloete K, Benjeddou M, Davison S. Characterization of the highly discriminatory loci DYS449, DYS481, DYS518, DYS612, DYS626, DYS644 and DYS710. Forensic Sci Int Genet. 2010 Feb;4(2):104-10.

2. Butler JM, Schoske R, Vallone PM, Kline MC, Redd AJ, Hammer MF. A novel multiplex for simultaneous amplification of 20 Y chromosome STR markers. Forensic Sci Int. 2002 Sep 10;129:10-24.

3. Short Tandem Repeat DNA Internet Database. SRM 2395-Human Y-Chromosome DNA Profiling Standard. National Institute of Standards and Technology (NIST). http://www.cstl.nist.gov/strbase/srm2395.htm. Updated December 15, 2009. Accessed June 9, 2015.

aSNP Thresholds

The following is a list of aSNPs with analytical, and interpretation thresholds.

Table 38 aSNP Thresholds

Loci	% Analytical	% Interpretation
rs3737576	> 1.5	> 4.5
rs7554936	> 1.5	> 4.5
rs2814778	> 1.5	> 4.5
rs798443	> 1.5	> 4.5
rs1876482	> 1.5	> 4.5
rs1834619	> 1.5	> 4.5
rs3827760	> 1.5	> 4.5
rs260690	> 1.5	> 4.5
rs6754311	> 1.5	> 4.5
rs10497191	> 1.5	> 4.5
rs1919550	> 1.5	> 4.5
rs12498138	> 1.5	> 4.5
rs4833103	> 1.5	> 4.5
rs1229984	> 1.5	> 4.5
rs3811801	> 1.5	> 4.5
rs7657799	> 1.5	> 4.5
rs870347	> 1.5	> 4.5
rs7722456	> 1.5	> 4.5
rs192655	> 1.5	> 4.5
rs3823159	> 1.5	> 4.5
rs917115	> 1.5	> 4.5
rs1462906	> 1.5	> 4.5
rs6990312	> 1.5	> 4.5
rs2196051	> 1.5	> 4.5
rs1871534	> 1.5	> 4.5
rs3814134	> 1.5	> 4.5
rs4918664	> 1.5	> 4.5
rs174570	> 1.5	> 4.5

Loci	% Analytical	% Interpretation
rs1079597	> 1.5	> 4.5
rs2238151	> 1.5	> 4.5
rs671	> 1.5	> 4.5
rs7997709	> 1.5	> 4.5
rs1572018	> 1.5	> 4.5
rs2166624	> 1.5	> 4.5
rs7326934	> 1.5	> 4.5
rs9522149	> 1.5	> 4.5
rs200354	> 1.5	> 4.5
rs1800414	> 1.5	> 4.5
rs12439433	> 1.5	> 4.5
rs735480	> 1.5	> 4.5
rs1426654	> 1.5	> 4.5
rs459920	> 1.5	> 4.5
rs4411548	> 1.5	> 4.5
rs2593595	> 1.5	> 4.5
rs17642714	> 1.5	> 4.5
rs4471745	> 1.5	> 4.5
rs11652805	> 1.5	> 4.5
rs2042762	> 1.5	> 4.5
rs7226659	> 1.5	> 4.5
rs3916235	> 1.5	> 4.5
rs4891825	> 1.5	> 4.5
rs7251928	> 1.5	> 4.5
rs310644	> 1.5	> 4.5
rs2024566	> 1.5	> 4.5
rs1689198	> 1.5	> 4.5
rs1291383	> 1.5	> 4.5
rs16891982	> 1.5	> 4.5
rs12913832	> 1.5	> 4.5

iSNP Thresholds

The following is a list of iSNPs with analytical, and interpretation thresholds.

Table 39 iSNP Thresholds

Loci	% Analytical	% Interpretation
rs1490413	> 1.5	> 4.5
rs560681	> 1.5	> 4.5
rs1294331	> 1.5	> 4.5
rs10495407	> 1.5	> 4.5
rs891700	> 1.5	> 4.5
rs1413212	> 1.5	> 4.5
rs876724	> 1.5	> 4.5
rs1109037	> 1.5	> 4.5
rs993934	> 1.5	> 4.5
rs12997453	> 1.5	> 4.5
rs907100	> 1.5	> 4.5
rs1357617	> 1.5	> 4.5
rs4364205	> 1.5	> 4.5
rs2399332	> 1.5	> 4.5
rs1355366	> 1.5	> 4.5
rs6444724	> 1.5	> 4.5
rs2046361	> 1.5	> 4.5
rs279844	> 1.5	> 4.5
rs6811238	> 1.5	> 4.5
rs1979255	> 1.5	> 4.5
rs717302	> 1.5	> 4.5
rs159606	> 1.5	> 4.5
rs13182883	> 1.5	> 4.5
rs251934	> 1.5	> 4.5
rs338882	> 1.5	> 4.5
rs13218440	> 1.5	> 4.5
rs1336071	> 1.5	> 4.5
rs214955	> 1.5	> 4.5

Loci	% Analytical	% Interpretation
rs727811	> 1.5	> 4.5
rs6955448	> 1.5	> 4.5
rs917118	> 1.5	> 4.5
rs321198	> 1.5	> 4.5
rs737681	> 1.5	> 4.5
rs763869	> 1.5	> 4.5
rs10092491	> 1.5	> 4.5
rs2056277	> 1.5	> 4.5
rs4606077	> 1.5	> 4.5
rs1015250	> 1.5	> 4.5
rs7041158	> 1.5	> 4.5
rs1463729	> 1.5	> 4.5
rs1360288	> 1.5	> 4.5
rs10776839	> 1.5	> 4.5
rs826472	> 1.5	> 4.5
rs735155	> 1.5	> 4.5
rs3780962	> 1.5	> 4.5
rs740598	> 1.5	> 4.5
rs964681	> 1.5	> 4.5
rs1498553	> 1.5	> 4.5
rs901398	> 1.5	> 4.5
rs10488710	> 1.5	> 4.5
rs2076848	> 1.5	> 4.5
rs2107612	> 1.5	> 4.5
rs2269355	> 1.5	> 4.5
rs2920816	> 1.5	> 4.5
rs2111980	> 1.5	> 4.5
rs10773760	> 1.5	> 4.5
rs1335873	> 1.5	> 4.5
rs1886510	> 1.5	> 4.5
rs1058083	> 1.5	> 4.5

Loci	% Analytical	% Interpretation
rs354439	> 1.5	> 4.5
rs1454361	> 1.5	> 4.5
rs722290	> 1.5	> 4.5
rs873196	> 1.5	> 4.5
rs4530059	> 1.5	> 4.5
rs1821380	> 1.5	> 4.5
rs8037429	> 1.5	> 4.5
rs1528460	> 1.5	> 4.5
rs729172	> 1.5	> 4.5
rs2342747	> 1.5	> 4.5
rs430046	> 1.5	> 4.5
rs1382387	> 1.5	> 4.5
rs9905977	> 1.5	> 4.5
rs740910	> 1.5	> 4.5
rs938283	> 1.5	> 4.5
rs8078417	> 1.5	> 4.5
rs1493232	> 1.5	> 4.5
rs9951171	> 1.5	> 4.5
rs1736442	> 1.5	> 4.5
rs1024116	> 1.5	> 4.5
rs719366	> 1.5	> 4.5
rs576261	> 1.5	> 4.5
rs1031825	> 1.5	> 4.5
rs445251	> 1.5	> 4.5
rs1005533	> 1.5	> 4.5
rs1523537	> 1.5	> 4.5
rs722098	> 1.5	> 4.5
rs2830795	> 1.5	> 4.5
rs2831700	> 1.5	> 4.5
rs914165	> 1.5	> 4.5
rs221956	> 1.5	> 4.5

Loci	% Analytical	% Interpretation
rs733164	> 1.5	> 4.5
rs987640	> 1.5	> 4.5
rs2040411	> 1.5	> 4.5
rs1028528	> 1.5	> 4.5

pSNP Thresholds

The following is a list of pSNPs with analytical, and interpretation thresholds.

Table 40 pSNP Thresholds

Loci	% Analytical	% Interpretation
rs28777	> 1.5	> 4.5
rs12203592	> 1.5	> 4.5
rs4959270	> 1.5	> 4.5
rs683	> 1.5	> 4.5
rs1042602	> 1.5	> 4.5
rs1393350	> 1.5	> 4.5
rs12821256	> 1.5	> 4.5
rs12896399	> 1.5	> 4.5
rs2402130	> 1.5	> 4.5
rs1800407	> 1.5	> 4.5
rs312262906_N29insA	> 1.5	> 4.5
rs1805005	> 1.5	> 4.5
rs1805006	> 1.5	> 4.5
rs2228479	> 1.5	> 4.5
rs11547464	> 1.5	> 4.5
rs1805007	> 1.5	> 4.5
rs201326893_Y152OCH	> 1.5	> 4.5
rs1110400	> 1.5	> 4.5
rs1805008	> 1.5	> 4.5
rs885479	> 1.5	> 4.5
rs1805009	> 1.5	> 4.5
rs2378249	> 1.5	> 4.5
rs16891982	> 1.5	> 4.5
rs12913832	> 1.5	> 4.5

Loci

- The amplicon length does not include 120 bp for adapter sequences. The amplicon start and end positions are the 1-based endpoints of the entire amplicon including the sequence matching primers on the hg19 human reference genome.
- All loci in DNA Primer Mix A are also included in DNA Primer Mix B.
- SNP alleles are reported as described in dbSNP build 141.
- **Amelogenin**—A genetic marker that confirms the gender of the donor of the biological sample. Its size range is 106–112 bp and the control DNA is male.

Autosomal STRs

The following loci are detected using DNA Primer Mix A or DNA Primer Mix B.

Table 41 Autosomal STRs

Locus	Repeat Range (repeats)	Amplicon Length Range (bp)	Chromosome	2800M Control Alleles
D1S1656	7–21.3	133–192	1	12,13
TPOX	4–16	61–109	2	11,11
D2S441	7–17	137–177	2	10,14
D2S1338	10–33.1	110–203	3	22,25
D3S1358	8–22	138–194	3	17,18
D4S2408	8–13	98–118	4	9,9
FGA	12.2–53	150–312	4	20,23
D5S818	4–20	98–162	5	12,12
CSF1PO	5–17	72–120	5	12,12
D6S1043	8–26	154–226	6	12,20
D7S820**	5–21.1	118–183	7	8,11
D8S1179	6–20	82–138	8	14,15
D9S1122	8–15	104–132	9	12,12
D10S1248	7–20	124–176	10	13,15
TH01	3–14	96–140	11	6,9.3
vWA	11–26	135–195	12	16,19
D12S391	13–28	229–289	12	18,23
D13S317	5–17	138–186	13	9,11
PentaE	5–28.4	362–481	15	7,14
D16S539	4–17	132–184	16	9,13
D17S1301	9–15	130–154	17	11,12
D18S51	6–40	136–272	18	16,18
D19S433	4–27	148–240	19	13,14
D20S482	9–17	125–157	20	14,15
D21S11	12–41.2	147–265	21	29,31.2
PentaD	1.1–19	209–298	21	12,13
D22S1045*	8–19	201–245	22	16,16

* Interpret locus D22S1045 with caution. Elevated n-1 repeat stutter might be observed, particularly with decreased marker coverage. Heterozygote imbalance might be observed regardless of marker coverage. Consider multilocus genotype when determining the presence of a DNA mixture.

** A low-level plus .1 base pair artifact might be observed at locus D7S820 with a single T addition at the end of the STR repeat sequence of the parent allele (e.g., 8,8.1 or 11,11.1).

Identity Informative SNPs

The following loci are detected using DNA Primer Mix A or DNA Primer Mix B.

Table 42 Identity Informative SNPs

Locus	Amplicon Length (bp)	Chromosome	Amplicon Start Position	Amplicon End Position	2800M Control Alleles
rs10495407	109	1	238439234	238439342	G
rs1294331	85	1	233448359	233448443	GA
rs1413212	64	1	242806767	242806830	G
rs1490413	98	1	4367256	4367353	A
rs560681	90	1	160786641	160786730	AG
rs891700	115	1	239881850	239881964	AG
rs1109037	118	2	10085691	10085808	G
rs12997453	100	2	182413195	182413294	A
rs876724	119	2	114945	115063	C
rs907100	115	2	239563542	239563656	CG
rs993934	120	2	124109120	124109239	C
rs1355366	119	3	190806041	190806159	AG
rs1357617	120	3	961696	961815	AT
rs2399332	157	3	110300999	110301155	AC
rs4364205	98	3	32417576	32417673	G
rs6444724	120	3	193207306	193207425	T
rs1979255	102	4	190318007	190318108	G
rs2046361	120	4	10968994	10969113	A
rs279844	167	4	46329584	46329750	AT
rs6811238	120	4	169663541	169663660	G
rs13182883	169	5	136633252	136633420	AG
rs159606	104	5	17374845	17374948	A
rs251934	97	5	174778619	174778715	T
rs338882	157	5	178690599	178690755	C
rs717302	110	5	2879333	2879442	G
rs13218440	170	6	12059928	12060097	AG
rs1336071	120	6	94537182	94537301	G
rs214955	120	6	152697629	152697748	G
rs727811	115	6	165045254	165045368	A
rs321198	165	7	137029715	137029879	T
rs6955448	120	7	4310285	4310404	CT
rs737681	120	7	155990742	155990861	T
rs917118	109	7	4456953	4457061	C
rs10092491	116	8	28411037	28411152	CT
rs2056277	104	8	139399038	139399141	C
rs4606077	151	8	144656710	144656860	CT
rs763869	85	8	1375576	1375660	CT
rs1015250	117	9	1823702	1823818	G
rs10776839	103	9	137417271	137417373	G
rs1360288	119	9	128967994	128968112	C
rs1463729	99	9	126881396	126881494	GA
rs7041158	115	9	27985907	27986021	C
rs3780962	94	10	17193284	17193377	T

Locus	Amplicon Length (bp)	Chromosome	Amplicon Start Position	Amplicon End Position	2800M Control Alleles
rs735155	170	10	3374133	3374302	A
rs740598	120	10	118506839	118506958	AG
rs826472	153	10	2406511	2406663	T
rs964681	105	10	132698394	132698498	CT
rs10488710	118	11	115207134	115207251	CG
rs1498553	111	11	5708981	5709091	CT
rs2076848	118	11	134667502	134667619	AT
rs901398	90	11	11096173	11096262	T
rs10773760	99	12	130761623	130761721	AG
rs2107612	103	12	888262	888364	AG
rs2111980	94	12	106328186	106328279	G
rs2269355	65	12	6945881	6945945	C
rs2920816	157	12	40862976	40863132	T
rs1058083	76	13	100038193	100038268	AG
rs1335873	109	13	20901665	20901773	T
rs1886510	116	13	22374646	22374761	CT
rs354439	170	13	106938320	106938489	T
rs1454361	118	14	25850765	25850882	AT
rs4530059	170	14	104769099	104769268	G
rs722290	101	14	53216686	53216786	G
rs873196	114	14	98845506	98845619	CT
rs1528460	115	15	55210664	55210778	T
rs1821380	118	15	39313343	39313460	G
rs8037429	63	15	53616876	53616938	T
rs1382387	89	16	80106318	80106406	GT
rs2342747	104	16	5868645	5868748	AG
rs430046	119	16	78016980	78017098	C
rs729172	104	16	5606153	5606256	C
rs740910	113	17	5706552	5706664	A
rs8078417	143	17	80461847	80461989	CT
rs938283	98	17	77468433	77468530	T
rs9905977	170	17	2919324	2919493	G
rs1024116	98	18	75432317	75432414	A
rs1493232	75	18	1127945	1128019	A
rs1736442	153	18	55225698	55225850	G
rs9951171	119	18	9749789	9749907	G
rs576261	76	19	39559780	39559855	AC
rs719366	170	19	28463281	28463450	T
rs1005533	158	20	39487066	39487223	A
rs1031825	126	20	4447416	4447541	C
rs1523537	117	20	51296076	51296192	C
rs445251	119	20	15124865	15124983	CG
rs221956	97	21	43606933	43607029	C
rs2830795	114	21	28608089	28608202	A
rs2831700	79	21	29679639	29679717	A
rs722098	101	21	16685561	16685661	AG
rs914165	156	21	42415865	42416020	AG
rs1028528	78	22	48362256	48362333	AG

Locus	Amplicon Length (bp)	Chromosome	Amplicon Start Position	Amplicon End Position	2800M Control Alleles
rs2040411	68	22	47836378	47836445	A
rs733164	120	22	27816711	27816830	AG
rs987640	120	22	33559450	33559569	AT

Y Haplotype Markers

The following loci are detected using DNA Primer Mix A or DNA Primer Mix B.

Table 43 Y Haplotype Markers

Locus	Repeat Range (repeats)	Amplicon Length Range (bp)	Chromosome	2800M Control Alleles
DYF387S1	30–44	207–263	Y	37,38
DYS19	9–19	269–309	Y	14
DYS385a-b	7–28	232–316	Y	13,16
DYS389I	9–17	236–268	Y	14
DYS389II	24–34	283–323	Y	31
DYS390	17–28	290–334	Y	24
DYS391	5–16	119–163	Y	10
DYS392*	6–17	318–362	Y	13
DYS437	10–18	194–226	Y	14
DYS438	6–16	129–179	Y	9
DYS439	6–17	167–211	Y	12
DYS448	14–26	330–402	Y	19
DYS460	7–14	348–376	Y	11
DYS481	17–32	129–174	Y	22
DYS505	9–15	162–186	Y	11
DYS522	8–17	298–334	Y	12
DYS533	7–17	186–226	Y	12
DYS549	10–14	210–226	Y	13
DYS570	10–26	142–206	Y	17
DYS576	10–25	163–223	Y	18
DYS612	26–33	275–296	Y	29
DYS635	15–30	242–302	Y	21
DYS643	7–15	141–181	Y	10
Y-GATA-H4	8–15	159–187	Y	11

* Interpret the locus DYS392 with caution. Elevated n-1 repeat stutter might be observed, particularly with decreased marker coverage. Consider multilocus genotype when determining the presence of a DNA mixture.

X Haplotype Markers

The following loci are detected using DNA Primer Mix A or DNA Primer Mix B.

Table 44 X Haplotype Markers

Locus	Repeat Range (repeats)	Amplicon Length Range (bp)	Chromosome	2800M Control Alleles
DXS10074	7–22	184–244	X	21
DXS10103	14–21	157–185	X	18
DXS10135	15.3–34	239–312	X	28

Locus	Repeat Range (repeats)	Amplicon Length Range (bp)	Chromosome	2800M Control Alleles
DXS7132	11–20	175–211	X	13
DXS7423	10–18	188–220	X	15
DXS8378	8–14	434–458	X	12
HPRTB	8–17	193–229	X	12

Phenotypic Informative SNPs

The following loci are detected when using DNA Primer Mix B. These loci are not present when using DNA Primer Mix A.

Table 45 Phenotypic Informative SNPs

Locus	Amplicon Length (bp)	Chromosome	Amplicon Start Position	Amplicon End Position	2800M Control Alleles
rs28777	92	5	33958916	33959007	A
rs12203592	110	6	396273	396382	C
rs4959270	161	6	457655	457815	AC
rs683	120	9	12709246	12709365	AC
rs1042602	113	11	88911659	88911771	AC
rs1393350	99	11	89010977	89011075	G
rs12821256	119	12	89328278	89328396	CT
rs12896399	73	14	92773627	92773699	G
rs2402130	120	14	92801169	92801288	A
rs1800407	119	15	28230246	28230364	G
N29insA	112	16	89985688	89985799	C
rs1110400	173	16	89986044	89986216	T
rs11547464	173	16	89986044	89986216	G
rs1805005	213	16	89985774	89985986	G
rs1805006	213	16	89985774	89985986	C
rs1805007	173	16	89986044	89986216	C
rs1805008	173	16	89986044	89986216	C
rs1805009	227	16	89986484	89986710	G
rs201326893_Y152OCH	173	16	89986044	89986216	C
rs2228479	213	16	89985774	89985986	G
rs885479	173	16	89986044	89986216	G
rs2378249	118	20	33218028	33218145	A

Ancestry Informative SNPs

The following loci are detected when using DNA Primer Mix B. These loci are not present when using DNA Primer Mix A.

Table 46 Ancestry Informative SNPs

Locus	Amplicon Length (bp)	Chromosome	Amplicon Start Position	Amplicon End Position	2800M Control Alleles
rs2814778	120	1	159174650	159174769	A
rs3737576	98	1	101709521	101709618	A
rs7554936	106	1	151122413	151122518	CT
rs10497191	101	2	158667153	158667253	C
rs1834619	84	2	17901444	17901527	G

Locus	Amplicon Length (bp)	Chromosome	Amplicon Start Position	Amplicon End Position	2800M Control Alleles
rs1876482	120	2	17362526	17362645	C
rs260690	115	2	109579681	109579795	A
rs3827760	108	2	109513546	109513653	T
rs6754311	98	2	136707920	136708017	CT
rs798443	84	2	7968221	7968304	A
rs12498138	119	3	121459545	121459663	G
rs1919550	117	3	121364112	121364228	A
rs1229984	120	4	100239288	100239407	G
rs3811801	114	4	100244261	100244374	C
rs4833103	95	4	38815462	38815556	AC
rs7657799	116	4	105375396	105375511	T
rs7722456	114	5	170202901	170203014	T
rs870347	119	5	6844995	6845113	T
rs16891982*	108	5	33951621	33951728	G
rs192655	70	6	90518235	90518304	AG
rs3823159	119	6	136482701	136482819	A
rs917115	71	7	28172543	28172613	T
rs1462906	84	8	31896545	31896628	C
rs1871534	71	8	145639652	145639722	C
rs2196051	120	8	122124216	122124335	T
rs6990312	111	8	110602270	110602380	G
rs3814134	104	9	127267664	127267767	T
rs4918664	168	10	94920962	94921129	A
rs1079597	167	11	113296227	113296393	G
rs174570	120	11	61597179	61597298	C
rs2238151	113	12	112211753	112211865	CT
rs671	136	12	112241658	112241793	G
rs1572018	116	13	41715225	41715340	AG
rs2166624	71	13	42579949	42580019	AG
rs7326934	96	13	49070482	49070577	G
rs7997709	85	13	34847693	34847777	T
rs9522149	119	13	111827125	111827243	C
rs200354	165	14	99375246	99375410	G
rs12439433	100	15	36219979	36220078	G
rs1426654	92	15	48426457	48426548	A
rs1800414	116	15	28196969	28197084	A
rs735480	108	15	45152321	45152428	T
rs12913832*	119	15	28365523	28365641	AG
rs459920	78	16	89730800	89730877	T
rs11652805	119	17	62987113	62987231	T
rs17642714	118	17	48726060	48726177	AT
rs2593595	102	17	41056210	41056311	TC
rs4411548	158	17	40658440	40658597	G
rs4471745	67	17	53568849	53568915	G
rs2042762	83	18	35277568	35277650	A
rs3916235	120	18	67578894	67579013	AG
rs4891825	106	18	67867615	67867720	AG
rs7226659	149	18	40488180	40488328	G

Locus	Amplicon Length (bp)	Chromosome	Amplicon Start Position	Amplicon End Position	2800M Control Alleles
rs7251928	200	19	4077044	4077243	A
rs310644	89	20	62159472	62159560	A
rs2024566	88	22	41697312	41697399	A

* Also used for phenotype prediction.

STR and iSNP Flanking Region Reporting

Data from amplicon flanking regions may be used as a tool for further interrogation and classification of sequence data from the ForenSeq DNA Signature Prep Kit. Flanking region sequence data are not depicted in the user interface of the ForenSeq Universal Analysis Software and do not affect genotyping results in the software. These data are provided for informational use; the ForenSeq Universal Analysis Software does not declare a genotype from data within an amplicon's flanking sequences.

Analysis setting thresholds such as analytical thresholds, interpretation thresholds, and stutter filters, are applied by the ForenSeq Universal Analysis Software only during genotyping of ForenSeq loci and are not applied for flanking region data analysis.

The following are applied when reporting flanking region sequence data in Excel:

- The reported amplicon sequence includes the sequence data from Read 1 for nucleotides that exist between the ForenSeq PCR primers.
- Unique sequences for a locus present at greater than 10 reads are reported.
- Sequence data downstream of the flanking regions are not reported for DYS389II, DYS439, or DYS570.
- Flanking regions for nine amplicons might be truncated to ensure sequencing data integrity. See Table 47 for the maximum number of bases that will be displayed.

NOTE
As additional data are available, flanking region sequence reports may be updated as needed for analysis of forensically relevant loci.

Table 47 STR Loci with Potential Truncated Flanking Sequence

STR Locus	Maximum Reported Amplicon Length (Bases)
Penta E	197
DXS10135	258
DXS8378	168
DYS19	237
DYS390	254
DYS392	177
DYS448	199
DYS460	233
DYS522	253

Set Up an External Data Repository

This section describes how to set up an external data repository on your network or on an external data storage device.

Important Notes

- For the new location on the network, all services and files are accessed by the same user, typically, a dedicated service user or group. When set up is complete, verify that the service user or group can read and write files at the remote location and that all web and windows services are running as that user.
- In the following steps, domain and service_user are placeholders for the names of the network domain and the network account, respectively, linking the ForenSeq Universal Analysis Software server box.
- In the following steps, the parent to the Runs folder is the folder that contains the Runs folder in the external repository.
- During the entire move operation, do not create or start new runs or new analyses. Doing so results in an unstable system state.
- In the config files, add .<domain>.com at the end of the machine name in the new run storage location entry. For example, if the machine is Nas01, the domain is Verogen, and the runs are in UAS001\Runs on that machine. The config entry for the management service is \\nas01.verogen.com\UAS001\Runs.

  NOTE
  The new location should have a parent folder for the Runs folder. In the example, the parent folder is UAS001.

- The steps listed below occur on either the network or the ForenSeq Universal Analysis Software server. It is not necessary to do any steps on the MiSeq FGx.

Overview of Steps

1. Request a new network user.
2. Stop the ForenSeq Universal Analysis Software Management service.
3. Create the parent folder.
4. Copy the existing runs folder.
5. Validate the service_user's accessibility.
6. Configure the ForenSeq Universal Analysis Software Management and ForenSeq Universal Analysis Software Web services.
7. Set up the configuration file for the management service.
8. Configure the Analysis service.
9. Set up the configuration file for the Analysis service.
10. Start the Analysis service.
11. Start the ForenSeq Universal Analysis Software service.

Detailed Steps

Request a New Network User

Contact your technology support department to create a new network user with access to all services and file locations, including read/write permissions.

Stop the ForenSeq Universal Analysis Software Management Service

1 Open the Windows Start menu and enter IIS.

2 Select Internet Information Services (IIS) Manager in the search box.

3 In the Connections pane, expand **domain\service_user > Sites** to view ForenSeq Universal Analysis Software sites.

4 Select **ForenSeq UAS Management**.

5 In the Actions pane of IIS under Manage Website, select **Stop** to enable the Start button.

Create the Parent Folder

1 Create and name the parent folder to the Runs folder.

2 Right-click on the new folder and select Properties.

3 Select the Sharing tab.

4 Select **Share**.

5 Enter the service_user account in the Choose people on your network to share with dialog box.

6 Select **Add** and verify that the user name appears in the list box near the bottom of the screen.

7 Click the arrow under Permission Level for the service_user and check **Read/Write**.

8 Select the Share tab.

9 Select **Done** in the File Sharing dialog box.

10 Click **OK** in the Properties dialog.

11 Close the window.

NOTE
Do not create the Runs folder.

Copy the Existing Runs Folder

NOTE
Depending on the number of runs, this portion of the operation may take several hours. Consider scheduling this step to run overnight.

1 In File Explorer, browse to C:\Illumina\Forenseq UAS.

2 Right-click the **Runs** folder and Select **Copy**.

3 In File Explorer, navigate to the parent folder of the new Runs folder.

4 Open the **parent folder**, in the folder contents, right-click and select **Paste**.

5 Open some run files to verify that all of the content in the run files was copied to the new Runs folder.

6 On the ForenSeq Universal Analysis server, rename the Runs folder at C:\Illumina\Forenseq UAS\Runs to RunsOld.

CAUTION
Do not delete the old Runs folder from the ForenSeq Universal Analysis Software server.

Validate the Service_User's Accessibility

NOTE
Skip this step if installing the run files on a local drive.

1 In File Explorer, right-click **Network** and select **Map network drive**.

2 Record the drive letter displayed in the Drive field.

3 In the Address Bar, paste the path to the new network location including the Runs folder.
For example: \\nas01.verogen.com\UAS001\Runs

4 Uncheck **Reconnect at sign-in**.

5 Check **Connect using different credentials**.

6 Click **Finish**.

7 Enter the service_user name and password in the Windows Security dialog.
The domain is pre-selected and has the same name as the ForenSeq Universal Analysis Software machine domain.

8 Click **OK**.

9 Open several run files to verify that they have content.

10 From File Explorer, open any Run folder .

11 In the contents of the Runs folder, right-click and select **New > Text Document**.

12 Verify the new text document is created in the Runs folder.

13 Delete the new text document by right-clicking it and selecting **Delete**.

Configure the ForenSeq Universal Analysis Software Management Service

1 In IIS, select **Application Pools** in the Connections pane.

2 In the Application Pools pane, right-click **ForenSeq UAS Management**.

3 Select **Advanced Settings**.

4 In the Advanced Settings dialog box, scroll down to the Process Model section, select **Identity Entry**, and click the **ellipses (…)**.

5 Select **Custom Account**.

6 Click **Set**.

7 Enter the User name as domain\service_user in the Set Credentials dialog box.
- domain is the network domain
- service_user is the new user

8 Enter and confirm the Password for service_user

9 Click **OK**.

10 In the Advanced Settings dialog box, click **OK** .

11 In the Application Pools pane, right-click **ForenSeq UAS Management**.

12 Select **Recycle** or **Start**, whichever is enabled.

Configure the ForenSeq Universal Analysis Software Web Service

1 In IIS, select **Application Pools** in the Connections pane.

2 In the Application Pools pane, right-click **ForenSeq UAS Web**.

3 Select **Advanced Settings**.

4 Scroll down to the Process Model, select the **Identity Entry** click the **ellipses (…)**.

5 In the Application Pool Identity dialog box, select **Custom Account**.

6 Click **Set**.

7 Enter the User name as domain\service_user.
- domain is the network domain
- service_user is the new user

8 In the Set Credentials dialog box, enter and confirm the Password for service_user.

9 Click **OK**.

10 In the Advanced Settings dialog box, click **OK** .

11 In the Application Pools pane, right-click the **ForenSeq UAS Web**.

12 Select **Recycle** or **Start**, whichever is enabled.

Set Up the Configuration File for the Management Service

1 In File Explorer, browse to C:\Illumina\Forenseq UAS\Management.

2 Right-click the **Web** file and select **Edit** to open the file in Windows Notepad.
The Web file shows as Web.config if extensions are shown.
Write permissions are required to edit the file.

3 From the Edit menu, select **Find**.

4 In the Find what box, enter "**RunFileRootPath**" (including the quotes).

5 Click **Find Next**.

6 Update the entry with the path to the remote location including the Runs folder.
In this example, the updated entry is in **bold**:
<add key="RunFileRootPath" value="**\\nas01.verogen.com\UAS001\Runs**" />

7 Save and close the configuration file.

Configure the Analysis Service

1 Open the Windows Start Menu, enter Services, select **Services**, and scroll down to ForenSeq Analysis.

2 Right-click **ForenSeq Analysis** and select **Stop**.

3 When successfully stopped, right-click **ForenSeq Analysis** again and select **Properties**.

4 Select the **Log On** tab.

5 In the Log on as section, select **This Account**.

6 In the This account box, enter the domain\service_user.
- domain is the network domain
- service_user is the new user

7 Enter the Password for service_user and click **Confirm**.

8 In the Properties dialog box, click **OK** to close it.
You should see a confirmation of the Log on as a Service permission grant.

Set Up the Configuration File for the Analysis Service

1 In File Explorer, browse to C:\Illumina\Forenseq UAS\Analysis.

2 Right-click **Matchbox.Analysis.ServiceHost.exe.config** and select **Edit** to open the file in Windows Notepad.
- The Web file shows as Web.config if extensions are shown.
- Write permissions are required to edit the file.

3 From the Edit menu, select **Find**.

4 In the Find what box, enter "**RootDirectory**" (including the quotes).

5 Click **Find**.

6 Update the RootDirectory to be the remote location to store the Runs folder. Do not include the Runs folder at the end of the path.
In this example, the updated entry is in **bold**:
<add key="RootDirectory" value="**\\nas01.verogen.com\UAS001**">

7 Save and close the configuration file.

Start the Analysis Service

1 From the Search field in the Windows Start Menu, enter Services, and select **Services**.

2 Scroll down the Services(Local) list to ForenSeq Analysis.

3 Right-click **ForenSeq Analysis** and select **Start**.

4 Verify in the Status column that ForenSeq Analysis is running.

Start the ForenSeq Universal Analysis Software Service

1 In the Connections pane of IIS, expand **domain\service_user > Sites** to view ForenSeq UAS sites.

2 Select **ForenSeq UAS Management**.

3 In the Actions pane, select **Start**.

4 Close the IIS window.

Index

P

Q

R

S

T

U

V

W

Technical Assistance

For technical assistance, contact Verogen Technical Support.

Table 48 General Contact Information

Address	11111 Flintkote Avenue San Diego CA 92121 USA
Website	www.verogen.com
Email	techsupport@verogen.com
Phone	+1.833.837.6436 toll-free (North America) +1.858.285.4101 (outside North America)

Safety data sheets (SDSs)

- For MiSeq FGx sequencing kit safety data sheets, visit www.verogen.com/sds.
- For Research Use Only (RUO) sequencing reagent and Illumina library preparation kit safety data sheets, visit support.illumina.com/sds.

Product documentation—Available for download in PDF from the Verogen website. Go to www.verogen.com/support select the appropriate document.

Verogen
+1.833.837.6436 toll-free (North America)
+1.858.285.4101 (outside North America)
techsupport@verogen.com
www.verogen.com