	Intermountain Forensics	**SOP #**	IAC-206
		Revision #	01
Forensic DNA Technical Leader Approval		**Issue Date**	
(signature)		04/02/2024	

NovaSeq 6000 Data Analysis

1. Purpose

This document describes the bioinformatic pipeline used for the analysis of data generated through whole genome sequencing on the NovaSeq 6000 or an equivalent platform.

2. Summary

The analysis of whole genome sequencing data is used by Intermountain Forensics to generate SNP files suitable for upload to genetic genealogy databases for uses in the identification of human remains and crime investigation by law enforcement, as well as determination of characteristics of the source of the DNA including haplogroups and sex.

3. Procedure

Documentation
1. The data is generated for all samples documented on the DOC-337 Library Preparation and Pooling Workbook, in cases where bioinformatics has been requested.
2. A DOC-335 Bioinformatics Data Analysis Worksheet is generated for each sample and will be included in the case file.
3. The MultiQC file/files will be saved for the data analysis within a batch.
4. SNP output files, if the data is suitable.

Data Transfer from BaseSpace
1. Download the fastq files for the batch using the BaseSpace CLI to a secure server.

Analyze data with nf-core/eager fork
1. Read pre-processing
2. Mapping
3. Deduplication
4. Sex Determination
5. Y-DNA Haplogroup Classification
6. mtDNA Haplogroup Classification and Contamination Assessment
7. Generation of sequencing data metrics
   a. Endogenous percentage
   b. Duplicate rate
   c. Sequence length
   d. Coverage depth and breadth (% of genome with 1X coverage)

Assessment of Autosomal Contamination
1. If the mean coverage is greater than 4X, an additional assessment of contamination will be performed based on the autosomal data using VerifyBAMID2.

Genotyping Imputation with GLIMPSE
1. The 1000 Genomes data set is used as a reference panel.

SNP Filtering
1. Completed with bcftools and custom script to remove runs of heterozygosity
2. Two SNP files are output
   a. GEDMatch
   b. FamilyTreeDNA

Archive FastQ Files to Azure Storage
1. Completed with the Azure CLI

4. References

N/A

5. Definitions

N/A