Using the independent CLC workflow for analyzing NGS SNP data

Description: This workflow is designed to align sequences to the included abridged human genome (*Abridged\_genome\_for\_forensic\_SNPs.fasta*) which contains sequences directly flanking all ancestry and identity forensic SNPs (n, 301) targeted by the ForenSeq™ DNA Signature Prep Kit (Illumina) and HID Precision Ancestry and Identity panels (ThermoFisher Scientific). Additionally, the workflow reports the depth of coverage of each nucleotide for each SNP, in an exported tab separated file (\*.tsv).

Dependencies: The Forensic\_SNP\_Extractor-1.0.cpw contains all necessary files to execute the pipeline. (*Abridged\_genome\_for\_forensci\_SNPs, Adapters.clc, and Forensic\_SNP\_locations.bed*)

Preparation: mport the developed CLC workflow by selecting the workflow icon from the upper right corner and selecting ‘*manage workflows*’. Then select the ‘*install from file’* button along the bottom of the pop-up window. Select the *Forensic\_SNP\_Extractor-1.0.cpw* file and then choose where to place it within CLC (Recommended to be in the main CLC\_Data level). \*In some cases, a window will pop-up and inform you that and update needs to be applied to the workflow before it can be run. Full instructions on importing workflows can be found at http://resources.qiagenbioinformatics.com/manuals/clcgenomicsworkbench/752/index.php?manual=Installing\_workflow.html

To set a default output location for the exported coverage (\*.tsv) from the workflow, double click on the box labelled ‘*export mapping coverage*’. A pop up window will appear and click on the ‘*folder/magnifying glass*’ icon to navigate to the desired folder for the output coverage files.

Execution: To access the workflow, look for *Forensic\_SNP\_Extractor-1.0* under workflows in the lower left Toolbox window. With the workflow open in the main window, click the ‘*run*’ button in the lower right corner. If more than one sample is to be run, ensure that the ‘*batch mode box*’ is checked. When all samples that are to be processed are selected and visible in the box on the right hand side, click ‘*next*’. Apply any file filtering options in the next screen (default is no action as files have already been selected) and click ‘*nex*t’. Change the output location if desired, otherwise, click ‘*finish*’. For each file, the workflow will automatically begin by filtering, trimming and post processing all reads within a single file, aligning the remaining high-quality reads to the abridged hg19 genome reference sequence and subsequently exporting the depth of coverage information for the 301 SNPs.

\*\*\*All settings definitions can be found in the CLC user’s manual, found at…

http://resources.qiagenbioinformatics.com/manuals/clcgenomicsworkbench/600/index.php?manual=Introduction\_CLC\_Genomics\_Workbench.html