HALOH (“Halo”) LTE Manuel

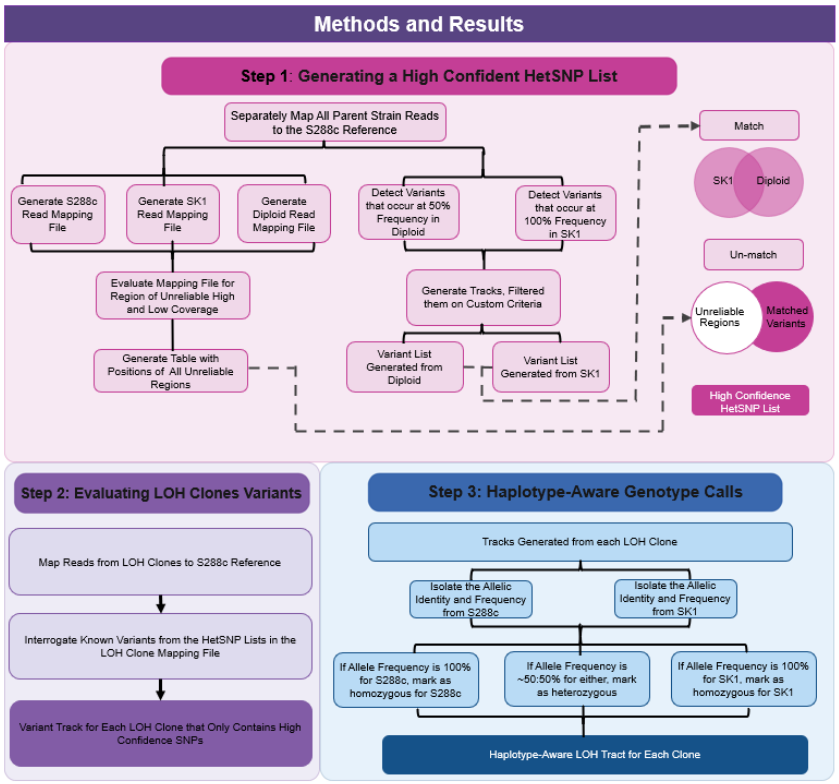
“A tool to call haplotype-aware loss of heterozygosity”

By Joy Love

Table of Contents

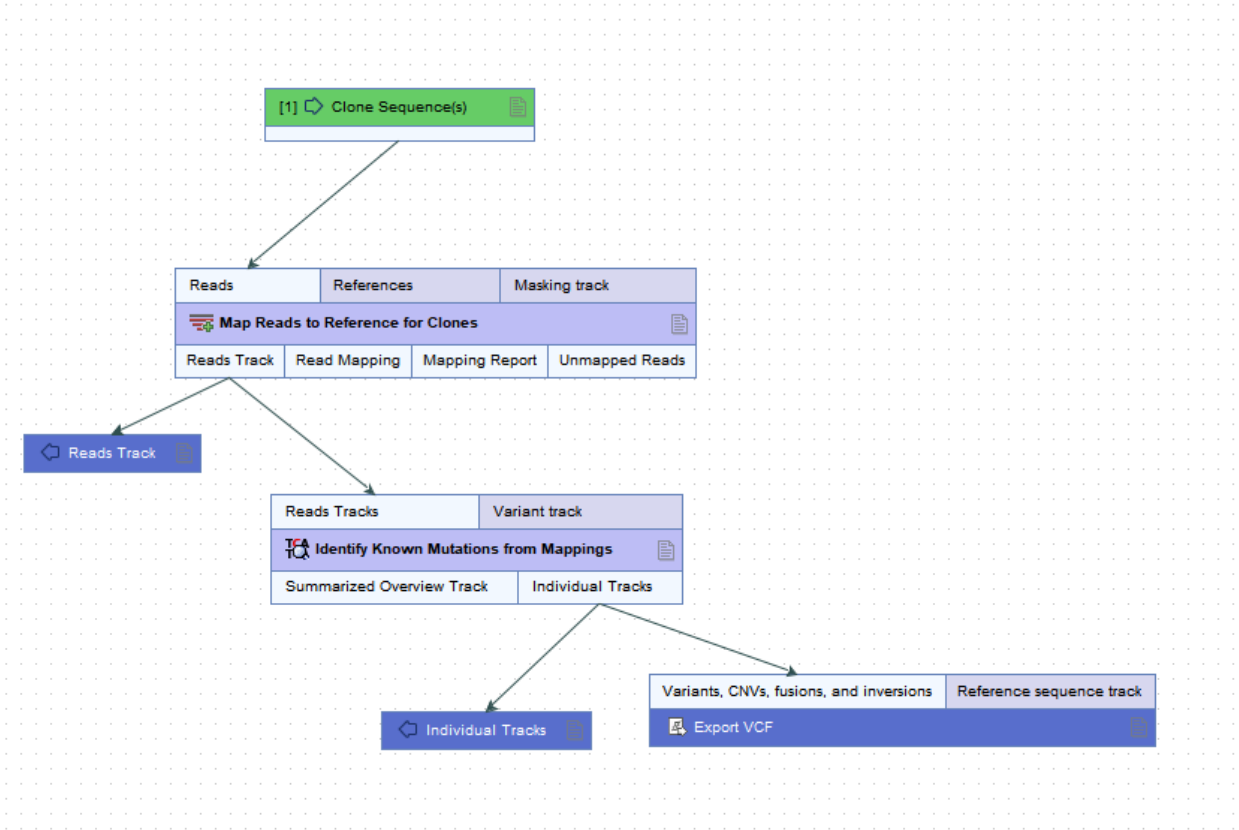
Overview

This manual outlines the Haplotype Aware Loss of Heterozygosity (HALOH) caller. As indicated by its name, this tool is a haplotype-aware and cost-effective LOH calling solution that leverages Python and CLC Genomics Workbench. This tool was developed with the assumption that the experiments to obtain the LOH tracts were in a hybrid heterozygous diploid strain background SK1 and S288c. To summarize this methodology, first, we created a complete list of heterozygous single nucleotide polymorphisms (HetSNPs) in the parent strain. We validated this list by mapping reads from the fully heterozygous parent diploid to one of the haploid parents (S288c, reference genome). We then derived a list of variants shared by the two sets and removed variants present at repetitive regions of the genome to arrive at a high-confidence HetSNP list. This enabled us to generate a custom variant tract in CLC. Subsequently, we mapped WGS reads from the known LOH clones above to the reference, creating a variant list for each LOH clone. These mapping variants were interrogated against the high-confidence HetSNP variant tract. CLC provided allelic identity and frequency outputs at all HetSNP positions, which were used in a custom Python code to make haplotype-aware phased genotyping calls.



This manual will summarize how to navigate CLC Genomic Workbench and Haloh LTE User Interface to generate haplotype-aware LOH tracts. To see this information and find the original code go to https://github.com/JoyLove0/HALOH-LTE.git.

Step 1: Running CLC Genomic Workbench Workflow called "Evaluation LOH Clone Variant"



- Open CLC Genomic Workbench
- Import Illumina
- Add Files
- Next and save to designated folder
- Finish
- Find "Evaluation LOH Clone Variant" Workflow and Click Run
- Select Reads and transfer
- Finish

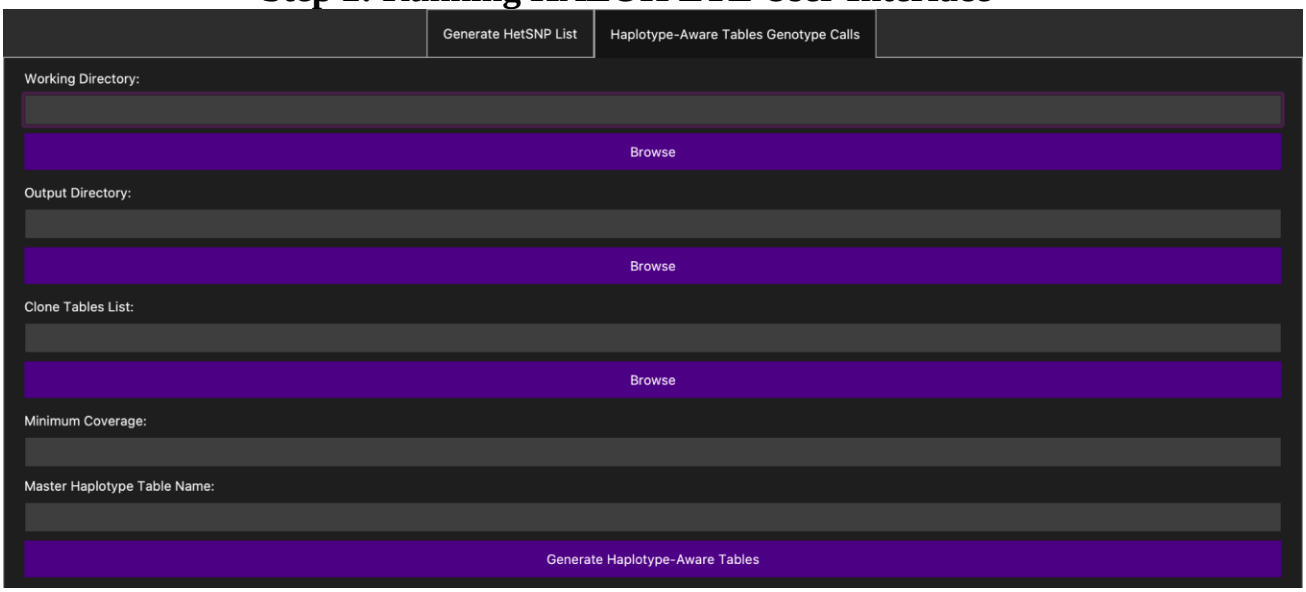
Helpful Manual and Links for CLC Genomic Workbench

The tools used in this step are "Evaluating LOH Clone Variants" are Map Reads to Reference, Identify Known Mutations from Mapping, and Export as VCF.

For more information on each these tools, see below:

Program	Manuals and/or Useful Links
All Manuals	https://resources.qiagenbioinformatics.com/manuals/clcgenomicsworkbench/current/index.php?manual=Introduction_CLC_Genomics_Workbench.html
Export in VCF	https://resources.qiagenbioinformatics.com/manuals/clcgenomicsworkbench/current/index.php?manual=Export_in_VCF_format.html
Identify Known Mutations from Mapping	https://resources.qiagenbioinformatics.com/manuals/clcgenomicsworkbench/950/index.php?manual=Identify_Known_Mutations_from_Sample_Mappings.html

Step 2: Running HALOH LTE User Interface



NOTE: All user inputs are explained in more detail in the next section

➢ Find the folder that contains the VCF files made in the last step

- ➢ In this folder, make a new text file. This file should contain a list of clone table names you wish to make haplotype aware tables for. See below for an example:

```
JS1R1_S1_L004-1 (Reads, Individual Track).csv
JS3R3_S49_L004 (Reads, Individual Track).csv
JS4R1_S10_L004 (Reads, Individual Track).csv
JS6R1_S60_L004 (Reads, Individual Track).csv
```

- ➢ Make a new folder within that folder containing the VCF files. This new folder will contain the haplotype-aware table (Recommended that you make a folder called "Output")
- ➢ Open HALOH LTE
- ➢ Using the browse button, select the folder (called working directory) where the VCF files are located
- ➢ Using the browse button, select the folder where you want to put the haplotype-aware tables
- ➢ Using the browse button, select the file that contains your file names
- ➢ Set the minimum coverage needed to make a genotype call. (Recommended you start with 25)
- ➢ Provide the name for the Master Output Table (.csv).
- ➢ Hit "Generate Haplotype-Aware Tables"

Inputs for HALOH LTE User Interface

Below is a more details description of all inputs for HALOH LTE:

Inputs	Descriptions
Working Directory	The working directory is the folder that contains the input and output for the tool. For reference, this folder must contain all the VCF files made with CLC. This is also the folder where the tools will do all the heavy lifting in.
Output Directory	The output directory is a folder you must make and will be the place where all the output files will be stored. For ease, it is recommended that this folder be inside the working directory.

Clone Table List	The clone table list is a text file containing the list of clone table names that should be analyzed. An example for the file name is "Clone_Variant_Tables.txt". Note that the tables in this file are the read files from CLC.
Minimum Coverage	You get to set the minimum coverage needed to a reliable call region in your LOH tracks heterozygous or homozygous. If a position falls underneath this coverage, it will be marked "No Call Due to Low Coverage". Suggest starting with 25 and making this number higher to make this input stricter.
Master Haplotype Aware Table Name	Provide the name for the file that will be made by this program. The table named here will contain all genotype calls for all selected clones, but individual tables will also be generated. This file can be useful in further analysis like SNP map plotting.