Guideline for Manual Reference Chromosome Selection

The selection of reference diploid chromosome is an important step in CNV prediction. While it is preferable to select reference chromosomes on independent experiments (such as karyotype), they can also be manually selected from WGS coverage data for samples not suspected to have aneuploidy.

Goal: Select a set of reference autosomes (one or more) that are largely diploid.

Step 1: Create the histograms of coverage distribution for each chromosome in both germline and tumor samples. Figure 1 shows a sample histogram.

Step 2: A largely diploid chromosome has a roughly symmetrical bell shape curve centered around the genome-wide median coverage. In reference selection, chromosomes significantly deviating from the expected distribution should be excluded. For example:

> **2.1.** Chromosomes with a heavy left tail/shoulder (chr2 and chr12 in the figure) in the tumor sample suggest a large scale deletion while a heavy right tail/shoulder (chr9) suggests a large scale gain. In the example, chr9 also shows a significantly higher peak around 0, which suggests a significant fraction of chr9 is homozygously deleted. At this step, chr2, chr9 and chr12 should be excluded from the reference selection.
>
> **2.2.** A chromosome with a bell shape curve shifted to the left or right of the genome-wide median coverage indicates a whole chromosome loss or gain. The example coverage plot suggests a whole chromosome gain for chr19 and chr22, which should be excluded from the reference chromosome selection.

After these steps, the initial reference chromosome set for this sample consists of chr1,3-8,10,11,13-18,20, and 21.

Step 3 (optional): The reference selection could be further refined based on the CONSERTING output using the initial reference selection to exclude chromosomes with medium size CNVs.

Please note that this procedure will selected an incorrect set of reference chromosomes if the tumor sample showed aneuploidy. In this situation, karyotyping is the preferred choice for reference chromosome selection.