CRAM (version 3) is a reference based compression file of sequence data (1). CRAM files can be viewed through samtools version 1.3 (more information can be found at the samtools website and example below) providing an additional flag of a reference genome (fasta file) that has been indexed with samtools faidx (2). Both of these files have been provided for you at ./Analysis_Pipeline_Files/human_g1k_v37_decoy.fasta). If you need to convert back to a bam file; this file can be converted back through with the script, cram_to_bam_conversion.sh (example command line below). This conversion retains all of the original information in the BAM file. If FASTQ files are desired, the cram_fastq_picard.sh script can create these FASTQ files directly from the cram file without creating a BAM intermediate using picard-1.141 or newer.

##View a CRAM File

Command line ex: samtools-1.3 view -h path/to/file.cram -T path/to/ref.fasta

A simple command line example to view the header and records within the cram file. The parameters are:

- Path to cram file with its accompanying index (or just name if file is located in current working directory)
- Path to reference genome with its accompanying index (or just name if file is located in current working directory)

##Conversion of CRAM to BAM

Command line ex: cram_to_bam_conversion.sh path/to/file.cram /path/to/bam/directory/ path/to/ref.fasta

This cram_to_bam_conversion.sh script will take these parameters in the following order:

- Path to cram file and its accompanying index (or just name if file is located in current working directory)
- Full path of desired output directory for the bam file and its index
- Path to reference genome and its accompanying index (or just name if file is located in current working directory)

Using samtools version 1.3 this script will convert whatever cram file you have specified to a bam file in the directory defined in the second parameter.

##Conversion of CRAM to FASTQ

Command Line ex: cram_fastq_picard.sh path/to/file.cram path/to/output/directory/ path/to/ref.fasta

The cram_fastq_picard.sh script will take these parameters in the following order:

1. Cram File ($CRAM_FILE)
2. Full Output Directory ($OUTPUT_DIRECTORY)
3. The Reference Genome ($REFERENCE_GENOME)

The script uses picard-1.141 SamToFastq to convert a valid cram file into its fastq components. The output of this command will place the fastq files, separated by read group, into the output directory specified in the second argument. The reference genome used should be the same version of the one used to create the cram file and is passed through as the third argument.

Conversion of BAM to CRAM

Command line ex: cram_conversion.sh path/to/file.bam path/to/output/directory/ path/to/ref.fasta

The cram_conversion.sh script will take these parameters in the following order:

- Path to bam file (or just name if file is located in current working directory)
- Full path of desired output directory
- Path to reference genome (or just name if file is located in current working directory)

This script uses samtools-1.3 to convert the bam file to a cram file and create an index file (.crai) in the specific $OUTPUT_DIRECTORY parameter. The cram file will have the same name as the bam file with the exception the extension will be .cram.

Reference Materials

1) **Efficient storage of high throughput DNA sequencing data using reference-based compression**
Markus Hsi-Yang Fritz, Rasko Leinonen, Guy Cochrane, and Ewan Birney
Genome Res. May 2011 21: 734-740; Published in Advance January 18, 2011,
doi:10.1101/gr.114819.110

2) CRAM format specification (version 3.0) : http://samtools.github.io/hts-specs/CRAMv3.pdf

3) Samtools-1.3 documentation: http://www.htslib.org/doc/samtools.html