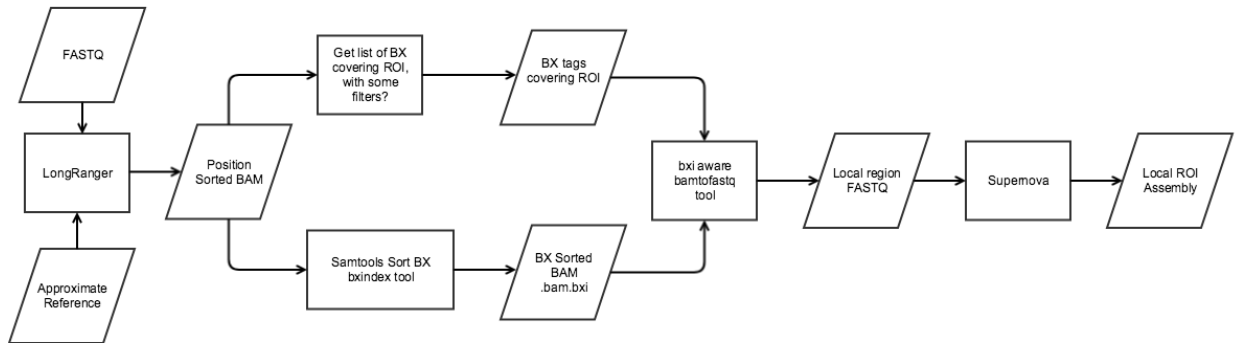


Local Reassembly Workflow

Introduction

This workflow describes how to perform barcode-guided local de-novo assembly of linked-read data. The set of barcodes mapped to a locus is used to select a group of reads that are highly enriched for that locus, but subject to much less mapping bias than would be the case with individually mapped short reads, especially if the true sequence contains large deviations from the reference (especially insertions). The Long Ranger 'lariat' alignments to the reference genome are used to determine the set of barcodes covering a locus. All reads from these barcodes are passed to the Supernova assembler. A BAM file that is sorted and indexed by the BX (linked-read barcode) tag is used to efficiently fetch the set of reads belonging to each barcode.



Here's the procedure to go from a Long Ranger BAM to local assemblies with Supernova. Inputs required are a Long Ranger BAM and a locus of interest. bindex, bamtofastq and samtools should all be on the LR 2.1.4 PATH.

<input_bam_file> is any Long Ranger phased_possorted_bam.bam file.

E.g. /mnt/analysis/marsoc/pipestances/CAU5LANXX/PHASER_SVCALLER_EXOME_PD/37366/1015.11.13-0/PHASER_SVCALLER_EXOME_PD/PHASER_SVCALLER_EXOME/_SNPINDEL_PHASER/ATT ACH_PHASING/fork0/files/phased_possorted_bam.bam

Output File Description	Output Filename	Command
BX sorted BAM file	bx-37366.bam	samtools sort -o bx-37366.bam -t BX -@ 8 -m 4G <input_bam_file>
BX index	bx-37366.bam.bxi	bindex bx-37366.bam
BX list in chosen locus (PMS2 gene in this example)	pms2-bx-list-37366.txt	locus_bx_list chr7:5012870-7791232 <input_bam_file> > pms2-bx-list-37366.txt
FASTQs path from BX list	fastqs_37366_pms2	bamtofastq --bx-list=pms2-bx-list-37366.txt bx-37366.bam fastqs_37366_pms2
SN assembly from FASTQs	asm_37366	supernova run --fastqs=fastqs_37366_pms2/37366_MissingLibrary_1_CAU5LANXX/ --id=asm37366 --desc=pms2-37366 --localcores=7 --nopreflight
SN pseudohap fasta	sm37366-pseudohap.fasta.gz	supernova mkoutput --asmdir=asm37366/outs/assembly/ --outprefix=asm37366-pseudohap --style=pseudohap