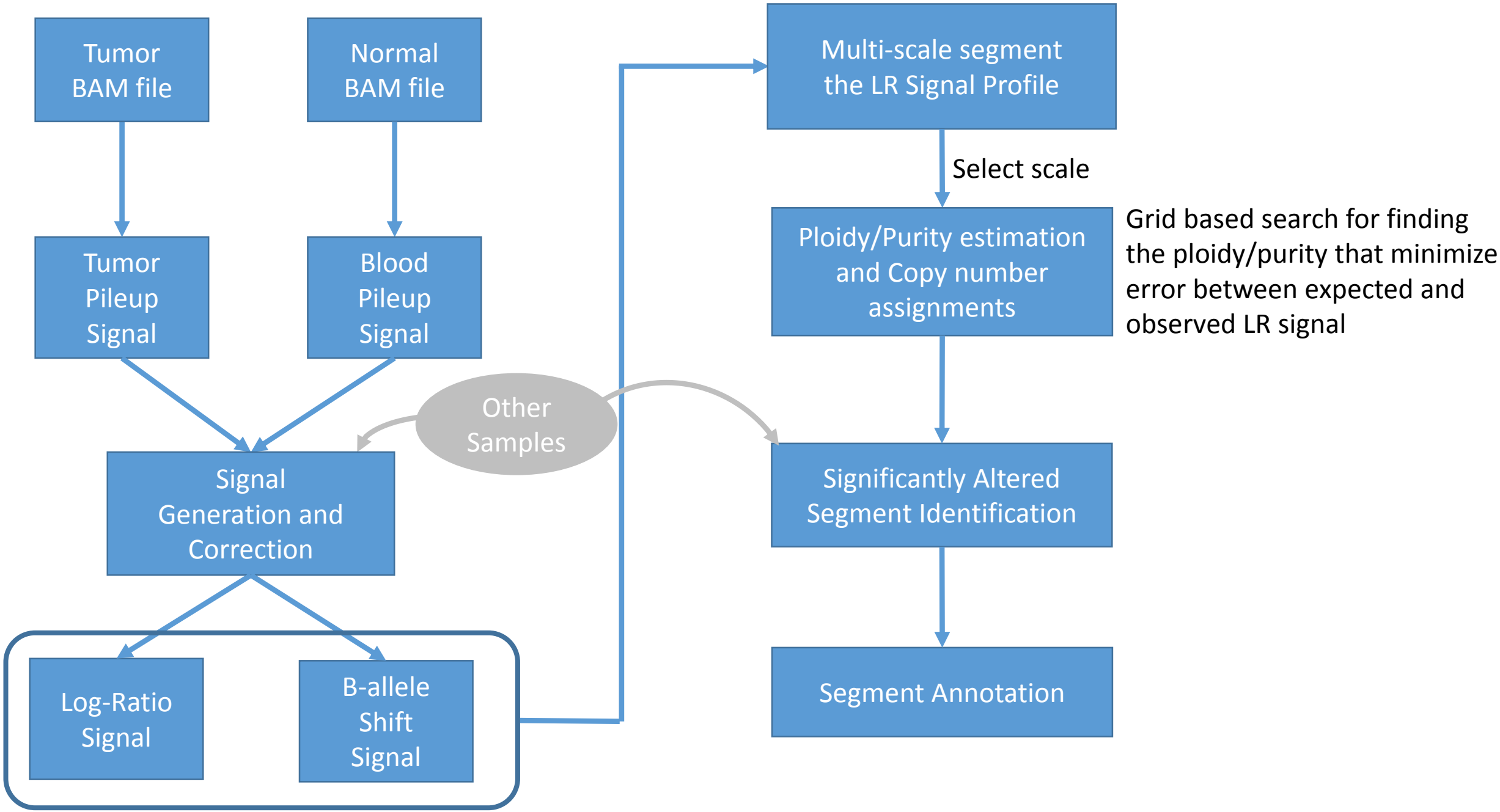
Somatator

12-7-2015

Motivation: Why a new tool?

- We are starting to perform large scale cancer sequencing data analysis in the lab
- CNV identification is an important step
- CNV analysis tools are not very convenient since different steps are performed with different tools
  1. Process reads and generate log ratio and LOH (B-allele shift) signals
  2. Noise/Bias reduction
  3. Identify segments
     a. Focal and large scale events
  4. Genotype segments
     a. Account for unknown ploidy and purity of the tumor
  5. Identify the significantly altered segments among all samples and annotate the segments
- Somatator aims at performing these steps in one run.
  - Exome/Whole genome

Somatator Workflow



Somatator Output: BED file with several extra columns

- Ploidy/purity estimates for each sample
- Reliability score for each segments' coordinates
- Copy number estimates for each segment
- Significantly altered segments among all samples
- Segments with significant loss-of-heterozygosity
- Observed LOH and LR signal levels for each segment
- Clonality estimate for each segment
- Annotations for each segment
- ***In progress:***
  - Local phasing information, visualization
  - Testing on a large set of cancer genomes
- Any suggestions?