Mixed effect modeling of human and mouse data in xenograft studies corrects for missed contamination or transcriptomic spillover

Xenograft models are widely used as a surrogate for better molecular characterization of patient disease. While xenograft models provide excellent material for analysis, there is always the concern of mouse contamination of the patient sample and even in strains with no immune compartment and extensive purification of the sample, contamination persists, especially in assays dependent on next-generation sequencing technologies. The standard practice to deal with this contamination is to use alignment tools to identify reads likely to be mouse contamination and filter it out. However, because of the high degree of homology between the mouse and human transcriptomes, there is still the risk of reads being misassigned. We have developed a method using standard NGS methods that incorporate the mouse data as a batch effect into the modeling of the human data, and have found it very effective for understanding the extent of contamination as well as effectively identifying spurious interpretations of the data.   
  
Required data  
  
Childhood Cancer Data Initiative (CCDI): CCDI Pediatric In Vivo Testing Program - Leukemia

Pediatric Preclinical Testing Consortium (PPTC)

Study of Leukemia Stem Cells in B-ALL

TARGET: Cancer Model Systems (MDLS): Cell Lines and Xenografts (including PPTP)

TARGET: Neuroblastoma (NBL)

Computational resources

For the purposes of the project, we would need the ability to align the data to mouse and human genomes prior to the jamboree, and generate count data for each sample. At the jamboree, the actual resources would be a moderate cloud instance, as the actual modeling is accomplished using the Bioconductor stack.