Introduction
We experiment on automated high-throughput pipelines for single nucleotide polymorphism (SNP) calling and discovery from whole-genome RNA-Seq data.

To assess possible SNP signatures of breast cancer, the pipelines are tested on short reads of breast tissue and BT474 & MCF7 cell lines transcriptomes.

The pipelines
The short reads are first aligned on the UCSC human genome reference, build HG18, by using two open source alignment modules, [based on]:
- Bowtie/TopHat (http://bowtie-bio.sourceforge.net) [2]
- BWA (http://bio-bwa.sourceforge.net) [3]

The alignment modules can be run in sequential or parallel mode on a high performance computing (HPC) facility. They are tested here on a distributed memory Linux cluster, using up to 312 computing cores.

SNP calling
Consensus building and downstream analysis is performed by SAMTools (http://samtools.sourceforge.net) [4], a set of utilities for the manipulation of mapped reads. SNPs are called from the consensus sequence and then filtered in order to retain only high-quality SNPs, based on the following inclusion criteria:
- Mapping quality > 25 (PHRED-based)
- Read depth > 3
- SNPs do not fall within 10bp from a gap

REFERENCES
2. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biology; 2009