Title: RNA-seq data of invasive ductal carcinoma and adjacent normal tissues from a Korean patient with breast cancer

Purpose:
It is well established that gene expression changes drive tumorigenesis, and these can be effectively identified using transcriptomic analyses such as RNA-seq. Ten samples of invasive ductal carcinoma from a Korean patient exhibiting Luminal-B subtype of breast cancer were sequenced for gene expression changes. Three samples of adjacent normal tissue were used as matched normal controls. Differentially expressed genes (FDR-adjusted p-value cutoff of 0.05) were identified and used to perform pathway analysis using Gene Ontology. The study found several pathways associated with the onset or progression of breast cancer, using this workflow.

Analysis Steps

1. Trim Galore (v0.4.2) with Cutadapt (v1.1.2) - Used to remove the adapter sequences from the reads before alignment
2. STAR (v2.5.2b) - Used to align trimmed reads to reference human genome (hg38)
3. Sambamba (v0.6.5) - Used to remove PCR-duplicates of mapped reads
4. RSeQC (v2.6.4) - Used to determine the quality of RNA-seq data with the TIN (Transcript Integrity Number) score
5. Cufflinks with Cuffnorm function (v2.2.1) - Used to identify differentially expressed genes
6. Gene Ontology (with Metascape) - Used to identify up and down-regulated pathways in invasive ductal carcinoma

Tool	Input	Output
Trim Galore	Fasta or Fastq	Trimmed Fasta, Fastq
STAR	Fasta or Fastq	SAM, BAM
Sambamba	SAM, BAM	SAM, BAM
RSeQC	BAM	BAM, txt
Cufflinks	SAM, BAM	GTF
Metascape	CSV, txt, Xls,	Tables, figures

Hong, J. H., Ko, Y. H., & Kang, K. (2018). RNA-seq data of invasive ductal carcinoma and adjacent normal tissues from a Korean patient with breast cancer. *Data in brief*, *18*, 736-739.