**Genomic alterations in Triple Negative Breast Cancer after chemotherapy**

Enas Helmy, Mai Ahmed, Sahar Mostafa, Nefertiti El-Nikhely

**Link to GitHub**

[**https://github.com/CancerNU/Mutations-in-TNBC-.git**](https://github.com/CancerNU/Mutations-in-TNBC-.git)

1. **Introduction**

Triple-negative breast cancer (TNBC) is an aggressive subtype that frequently develops resistance to chemotherapy. An unresolved question is whether resistance is caused by the selection of rare pre-existing clones or alternatively through the acquisition of new genomic aberrations (Kim *et al*, Cell 2018). To investigate this question, the research group did, among other approaches, exome sequencing on tumor samples before and after chemotherapy. In addition, they sequenced also matching blood samples, which will allow identifying tissue-specific mutations and exclude germline mutations.

*Reference*

Kim et al, Chemoresistance Evolution in Triple-Negative Breast Cancer Delineated by Single Cell Sequencing, Cell 2018 May 03; 173(4): 879–893.

1. **Aim**

* To compare variants in tumor samples before and after chemotherapy in order to identify SNPs altered with chemotherapy that could be good targets responsible for altered drug response.
* To compare variants in blood and tumor samples in order to identify tissue-specific variants.

1. **Methodology**

***Samples:*** Samples were obtained from SRA repository from the BioProject PRJNA396019. Only exome sequencing data were selected from blood samples and matching tumour samples before and after chemotherapy. Samples were sequenced at 100 paired-end cycles on the HiSeq2000 or HiSeq4000 system (Illumina).

***Data Processing:***

FASTQ sequences will be mapped and aligned. PCR duplicates will be removed. Single nucleotide variants will be identified. Tissue-specific somatic variants will be identified by excluding variants common in blood and tumor samples. SNPs will be compared to known databases and data will be filtered. Finally, significant mutation sites will be identified that are specific to tumor tissue and correlated to chemotherapy.

1. **Results and Discussion**
   1. **Sample selection**

Samples were obtained from SRA repository from the BioProject PRJNA396019. This is a big project comprising many data e.g. RNA sequencing, whole genome sequencing (WGS) and whole exome sequencing (WXS). WGS data was selected for variant calling in tumour samples before (PreT\_SRR5969410) and after chemotherapy (PostT\_SRR5969460). Only chromosome 10 was used as reference genome as a genome wide association study on Japanese population has shown increased variant numbers in chromosome 10 in patients with triple negative breast cancer (Low SK et al, 2013).

* 1. **Quality Control of FASTQ files**

Both samples had normal distribution and good quality scores (Fig. 1). These samples were thus used for further analysis.

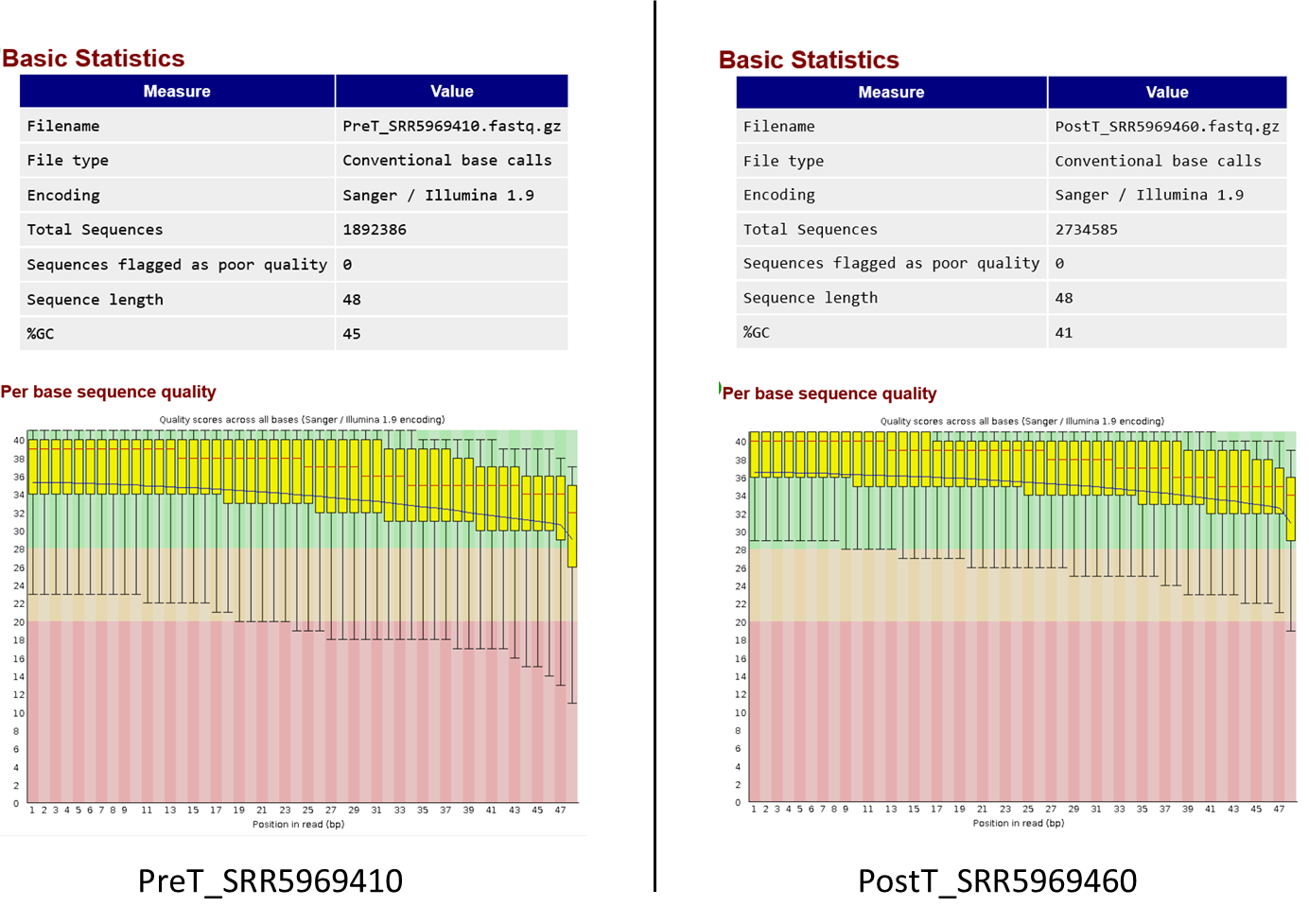


Fig.1 Quality control of FASTQ files of selected samples

* 1. **Sequence Alignment using bwa**

Fastq files were aligned against chromosome 10 after indexing and the produced SAM files were checked and visualized using SAM tools (Fig. 2).

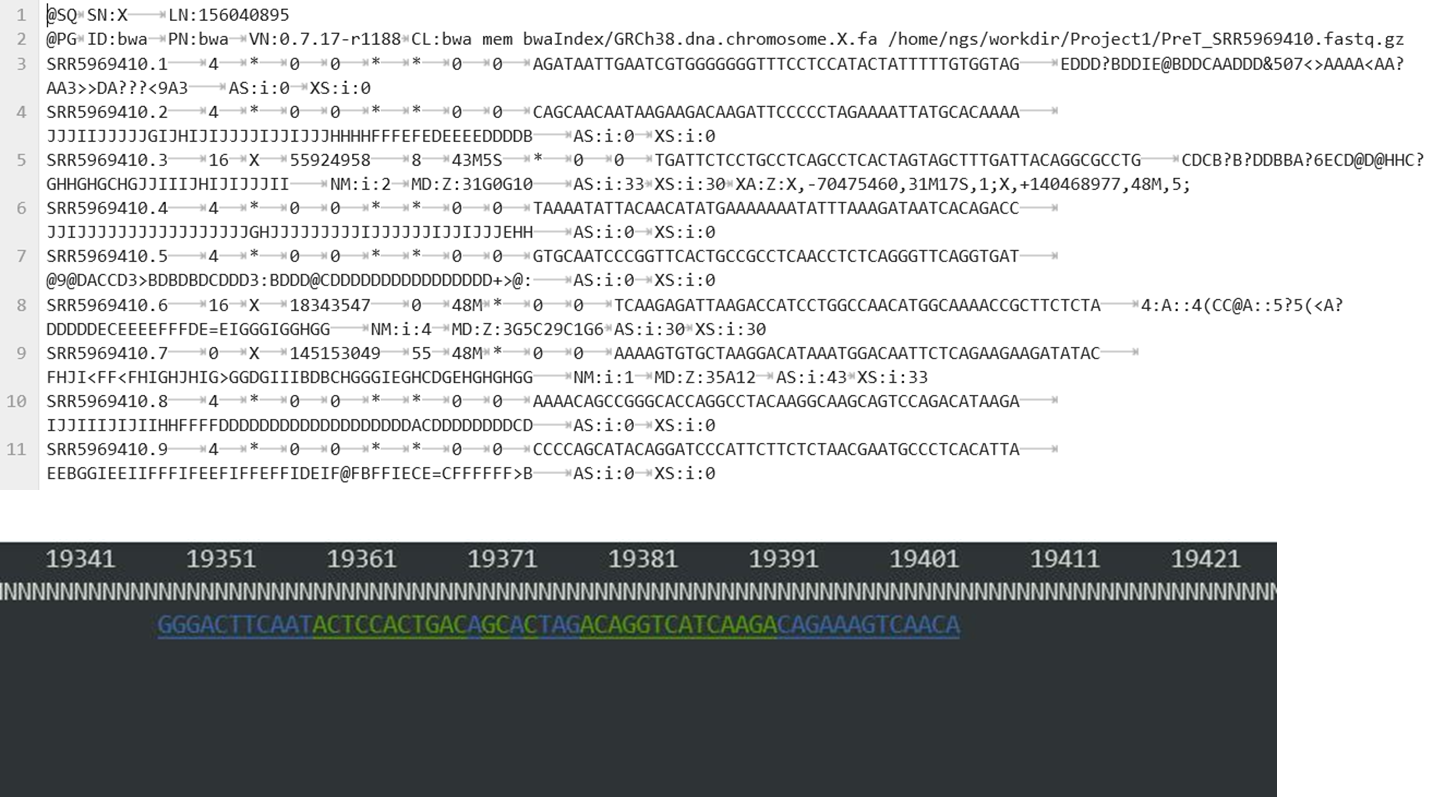


Fig. 2 Visualization of SAM files

* 1. **Variant calling using BCF tools**

Variant calling using BCF tools showed more SNPs after chemotherapy compared to the pretreatment sample. Also number of deletions and insertions is more after treatment (Fig. 3). However, this result need verification using GATK to compare with dbVar database and to identify unique variant. Mutect can also be used to compare 2 samples by variant joint calling.

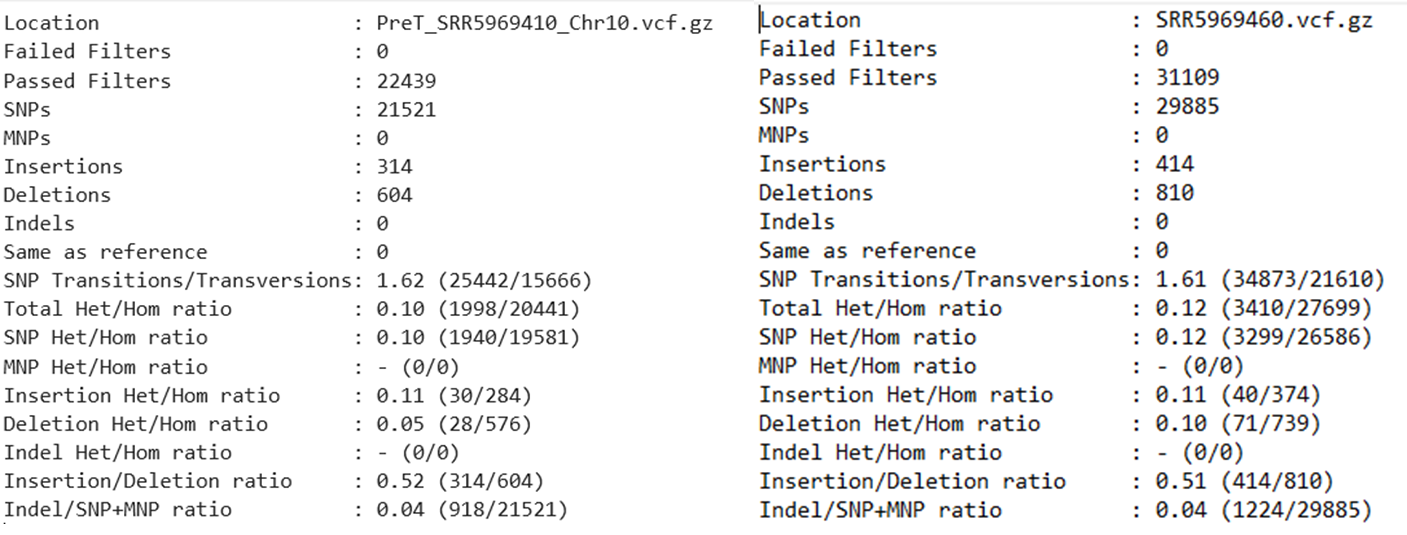


Fig. 3 Variant calling statistics

1. **Recommendations**

Further analysis is required to identify unique variants and to exclude germline variants using blood samples. In order to increase the significance of the identified variants more samples are needed.

1. **List of useful resources**

* <http://www.sixthresearcher.com/list-of-helpful-linux-commands-to-process-fastq-files-from-ngs-experiments/>
* <https://bioinformaticsworkbook.org/dataAcquisition/fileTransfer/sra.html>
* <https://github.com/ncbi/sra-tools/wiki/02.-Installing-SRA-Toolkit>
* <http://www.ensembl.org/info/data/ftp/index.html> (To download reference genome or variantions)
* <https://bioinf.shenwei.me/seqkit/faq/>
* Low SK et. al,Genome-wide association study of breast cancer in the Japanese population. PlosOne 2013