Using visual programing in Cervical cancer data of The Cancer Genome Atlas

Thais Talarico Hosokawa
Master’s student in Oncology
Cervical cancer

✓ The forth in women, estimates of 528,000 new cases

✓ In Brazil, the third most common

✓ Highly associated with HPV infection

Figure 1. Worldwide estimates of mortality by cervical cancer (per 100,000 cases). (1) GLOBOCAN Cancer Fact Sheets: Cervical cancer (2016)
We are one of the tissue sources (n=54)
Cervical cancer remains one of the leading causes of cancer-related deaths worldwide. Here we report the extensive molecular characterization of 228 primary cervical cancers, the largest comprehensive genomic study of cervical cancer to date. We observed striking APOBEC mutagenesis patterns and identified SHKBPI, ERBB3, CASP8, HLA-A, and TGFBR2 as novel significantly mutated genes in cervical cancer. We also discovered novel amplifications in immune targets CD274/PD-L1 and PDCDILG2/PD-L2, and the BCAR4 IncRNA that has been associated with response to lapatinib. HPV integration was observed in all HPV18-related cases and 76% of HPV16-related cases, and was associated with structural aberrations and increased target gene expression. We identified a unique set of endometrial-like cervical cancers, comprised predominantly of HPV-negative tumors with high frequencies of KRAS, ARID1A, and PTEN mutations. Integrative clustering of 178 samples identified Keratin-low Squamous, Keratin-high Squamous, and Adenocarcinoma-rich subgroups. These molecular analyses reveal new potential therapeutic targets for cervical cancers.
TCGA results

Our samples: Cervical squamous cell carcinoma (42, 77.8%) Mucinous adenocarcinoma of endocervical type (12, 22.2%)

Figure 4: Somatic alterations. Integrated genomic and molecular characterization of cervical cancer. Nature http://dx.doi.org/10.1038/nature21386 (2017).
Problems in Bioinformatics

- Hugh amount of data
- Deficiency in computacional programming
Purpose

✓ Using workflows to standardize analysis

✓ Controlled environment that keeps metadata
To be achieved

✓ To analyse sequencing, methylation and gene expression (RNA-seq) data;

✓ To possibly identify biomarkers that allow characterization of people at risk for cervical cancer in Brazil;

✓ To compare molecular data with clinical-pathological data such as linfonode methastasis, staging, radiotherapy and prognosis
What do we have so far?

✓ We requested the raw data, but we’re still waiting

✓ We did a data tidying and have some data (open access)

✓ The algorithms that are being used to develop on Galaxy
Clinical data

- RTCGA.clinical R package
- Filter and data tidying
- Fisher test
Some about the clinical data

Patient age in pathologic diagnosis

From 20 to 38 years old (22, 40.7%), From 39 to 46 years old (0, 0%), From 47 to 57 years old (13, 24.1%), From 58 to 88 years old (19, 35.2%)

Race

American Indian or Alaska Native (0, 0%), Asian (1, 2.7%), Black or African American (4, 10.8%), Black or African American (0, 0%), White (32, 86.5%)
RNASeq

Input dataset

FASTQ Groomer

Trimming/ manipulation

FASTQC

Parameters modified to paired-end

Tophat2

Differentially expressed genes

edgeR and DESeq

Visual representation of data

RGalaxy

GRCh37
Methylation

![Flowchart of methylation process](image-url)
Integrating
Acknowledgements

Dr. Adriane Feijó Evangelista

Dr. Rui Manuel Vieira Reis

Dr. José Humberto Tavares Guerreiro Fregnani

Dr. Leandro Mattos

Barretos Cancer Hospital

John’s Hopkins University and their team (Dave Clements and Olga)
Thank you