



# Assessment of Residual Breast Cancer Cellularity after Neoadjuvant Chemotherapy using Digital Pathology (Post-NAT-BRCA)

## Summary

Breast cancer (BC) is the second most commonly diagnosed cancer in the U.S. with more than 250,000 new cases of invasive breast cancers reported in 2017. The majority of women with locally advanced and a subset of patients with operable breast cancer will undergo systemic therapy prior to their surgery (neoadjuvant therapy/ NAT) to reduce the size of tumor(s) and possibly further undergo breast conserving surgery. The Post-NAT-BRCA dataset is a collection of representative sections from breast resections in patients with residual invasive BC following NAT. Histologic sections were prepared and digitized to produce high resolution, microscopic images of treated BC tumors. Also included, are clinical features and expert pathology annotations of tumor cellularity and cell types. [The Residual Cancer Burden Index \(RCBi\)](#), is a clinically validated tool for assessment of response to NAT associated with prognosis. Tumor cellularity is one of the parameters used for calculating the RCBi. In this dataset, tumor cellularity refers to a measure of residual disease after NAT, in the form of proportion of malignant tumor inside the tumor bed region; also annotated. (See [M D Anderson RCB Calculator](#) for a detailed description of tumor cellularity.) Malignant, healthy, lymphocyte and other labels were also provided for individual cells to aid development of cell segmentation algorithms.

## Acknowledgements

We would like to acknowledge the individuals and institutions:

- Canadian Cancer Society, grant #705772
- National Cancer institute of the National Institutes of Health under #U24CA199374
- Sunnybrook Health Sciences Centre, Toronto, Ontario, Canada

## Data Access

### Data Access

Data Type	Download all or Query/Filter
Images (SVS, 43.2 GB)	
Annotations (XLSX)	

## Detailed Description

### Detailed Description

Image Statistics	
Modalities	Pathology
Number of Participants	64
Number of Images	96

Images Size (GB)	43.2
------------------	------

## Content

The Post-NAT-BRCA dataset is composed of:

- 96 whole slide images stored in an uncompressed .svs file format, standard for pathology slides. Slides were scanned at 20x objective on an Aperio slide scanner at Sunnybrook Health Sciences Centre.
- An Excel (.xlsx) file containing clinical features for each patient including age, treatment, ER/PR/HER2 status etc. A key and detailed description of each column is provided in a separate tab titled "Definitions". Each row in the spreadsheet corresponds to a single (.svs) slide and anonymised patient ID's are provided in a separate column.
- Manual annotations of tumor cellularity and cell labels, provided as Sedeen annotation files (.xml). Annotations are given in two directories, where the "WSI\_train" folder contains WSIs annotated by a single rater and "WSI\_test" was annotated by two raters.

## Recommended Software

To browse whole slide images and annotations, we highly recommend you use Pathcore's Sedeen Viewer which is available for free: <https://pathcore.com/sedeen/>

Please ensure that the "sedeen" folder is unzipped and placed into the same folder containing the .svs files. Sedeen Viewer loads annotations from this folder automatically when images are opened.

Upon opening WSIs in Sedeen, you will notice that annotations have been color-coded according to the following key:

- Pink: Healthy (0% tumor cellularity)
- Blue: Low tumor cellularity (0 - 30%)
- Yellow: Medium tumor cellularity (31 - 70%)
- Green: High tumor cellularity (70 - 100%)
- White: Contains annotations at the cell level and labeled as:
  - Lymphocyte: TIL-E, TIL-S
  - Normal Epithelial: normal, UDH, ADH,
  - Malignant Epithelial: IDC, ILC, Muc C, DCIS 1, DCIS 2, DCIS 3, MC- E, MC - C, MC - M

### Citations & Data Usage Policy

## Citations & Data Usage Policy

Add any special restrictions in here.

Users of this data must abide by the [TCIA Data Usage Policy](#) and the [Creative Commons Attribution 3.0 Unported License](#) under which it has been published. Attribution should include references to the following citations:

### **i** Data Citation

"Martel, A. L., Nofech-Mozes, S., Salama, S., Akbar, S., & Peikari, M. (2019). Assessment of Residual Breast Cancer Cellularity after Neoadjuvant Chemotherapy using Digital Pathology [Data set]. The Cancer Imaging Archive. <https://doi.org/10.7937/TCIA.2019.4YIBTJNO>"

### **i** Publication Citation

Peikari, M., Salama, S., Nofech-Mozes, S. and Martel, A.L., 2017. Automatic cellularity assessment from post-treated breast surgical specimens. *Cytometry Part A*, 91(11), pp.1078-1087. DOI: <https://doi.org/10.1002/cyto.a.23244>

### **i** TCIA Citation

Clark K, Vendt B, Smith K, Freymann J, Kirby J, Koppel P, Moore S, Phillips S, Maffitt D, Pringle M, Tarbox L, Prior F. **The Cancer Imaging Archive (TCIA): Maintaining and Operating a Public Information Repository**, *Journal of Digital Imaging*, Volume 26, Number 6, December, 2013, pp 1045-1057. DOI: [10.1007/s10278-013-9622-7](https://doi.org/10.1007/s10278-013-9622-7)

## Other Publications Using This Data

TCIA maintains [a list of publications](#) which leverage TCIA data. If you have a manuscript you'd like to add please [contact the TCIA Helpdesk](#).

1. Akbar, S., Peikari, M., Salama, S., Panah, A. Y., Nofech-Mozes, S., Martel, A. L. (2019) Automated and Manual Quantification of Tumour Cellularity in Digital Slides for Tumour Burden Assessment. *Sci Rep*, 9, 14099. <https://doi.org/10.1038/s41598-019-50568-4>

### Versions

#### **Version 1 (Current): Updated 2019/10/01**

Data Type	Download all or Query/Filter
Images (SVS, 43.2 GB)	
Annotations (XLSX )	