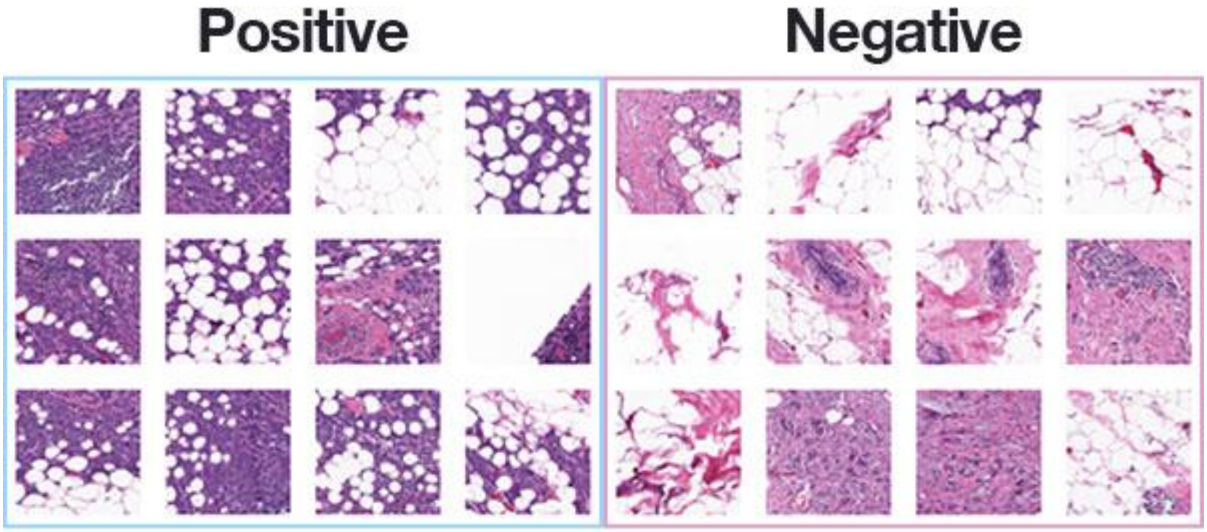
1. Describe “your data” from the final project. What is the data about and how is the data obtained experimentally and computationally? At the minimum, follow the steps taken in this notebook to show various attributes, dimensions, and meta data.

In my project I want to build deep network to analyze breast histology images for cancer risk factors. Dataset contains histology images (obtained experimentally) for Invasive Ductal Carcinoma that is type of breast cancer. The original dataset has only 162 images. Deep network requires a lot of data and only 162 images are not enough to build well suited network. Because of that (and also because of size of the images) patch extraction is needed to build proper dataset. **277,524** patches of *50 by 50* pixels were extracted from this 162 images. New dataset now contains 198,738 IDC negative and 78,786 IDC positive. That numer is big enough to build deep network.

The original papers are available here: https://www.ncbi.nlm.nih.gov/pubmed/27563488 and http://spie.org/Publications/Proceedings/Paper/10.1117/12.2043872

Example of patches:



2. Normalize your data. Follow the prescribed normalization process in the original paper. Critically discuss (e.g., provide a comment) whether their normalization process can be improved with ComBat or SVA. Apply additional normalization steps. How do they affect your data (e.g., cleaned data from SVA)?
3. Apply the jackstraw tests for PCA or related methods. Make sure how you decide to choose the number of significant PCs (`r` in the algorithm). Discuss the results. Create the histograms of p-values, q-values, null statistics, and observed statistics.

I think that tasks number 2 and 3 does not fit to my project. I do not work with normal, raw, numeric data but with pictures, so specific of my project is quite different. I am going to build convolutional network that can clarify whenever histology images contain cancer cells.