A Histology-Genomic Integration Analysis Using Machine Learning for Predicting Risk of Recurrence of Breast Cancer

Poster · September 2019

DOI: 10.13140/RG.2.2.29440.58888

3 authors:

Claudia Mazo
University College Dublin

Catherine Mooney
University College Dublin

William M Gallagher
University College Dublin

A Histology–Genomic Integration Analysis Using Machine Learning for Predicting Risk of Recurrence of Breast Cancer

Claudia Mazo[a,b,c,*], Catherine Mooney[a], and William M. Gallagher[c,d]
For further information please contact: claudia.mazovargas@ucd.ie

[a]UCD School of Computer Science, University College Dublin, Dublin, Ireland; [b]CeADAR: Centre for Applied Data Analytics Research, Dublin, Ireland; [c]OncoMark Limited, Dublin, Ireland; [d]UCD School of Biomolecular & Biomedical Science, UCD Conway Institute, Dublin, Ireland

Marie Skłodowska-Curie Actions
This project has received funding from the European Union's Horizon 2020 Research and innovation Programme under the Marie Skłodowska-Curie Co-funding of regional, national and international programmes Grant agreement No: 713654

Abstract – Breast cancer is the most frequently diagnosed cancer in women, with more than 2.1 million new diagnoses worldwide every year. Personalised treatment is critical to optimising outcomes for patients with breast cancer. DNA plays an important role in this way. A large amount of personal genomic data can be generated daily. On the other hand, the tumour microenvironment has been recognised as an integral component of identification, progression, therapy resistance, treatment options, and disease recurrence. We propose a way to face the challenges for managing, analysing, and interpreting these two different input sources using machine learning techniques. One of our main goals is to help doctors and patients to confidently decide on the best course of treatment.

❑Background

Clinical and economic need

- Most early BCs cured by surgery
- 30% recur & require chemotherapy
- Difficult to differentiate the two groups
- Tests required
- Competitors don't meet clinical need

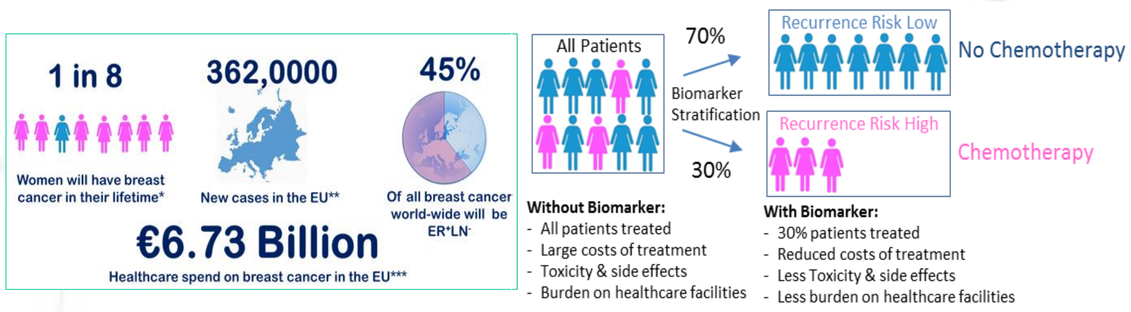

Figure 1. Breast cancer statistics[1]

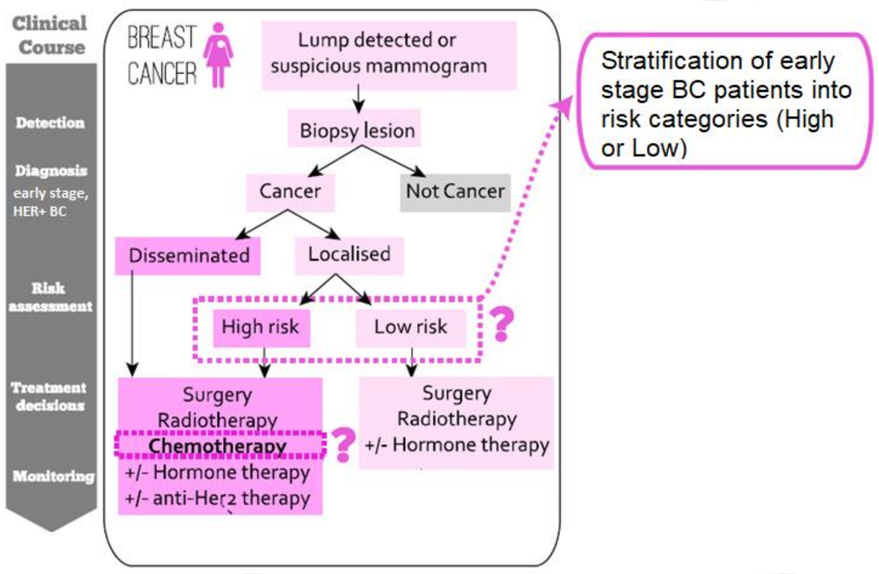

Figure 2. Overview of the clinical course of the disease

OncoMasTR (OM) is a multigene prognostic signature to predict recurrence of early stage breast cancer. OM was discovered using a novel transcriptional network analysis that identified genes – Master Transcriptional Regulators (MTRs) – that regulate previously known prognostic genes[2,3]. OM has been analytically[4] and clinically validated[5,6].

Figure 3. OncoMasTR web page

❑Method

The goal of this research is to use machine learning to develop a CDSS, called OncoEngine+, that stratifies early stage BC patients into risk categories

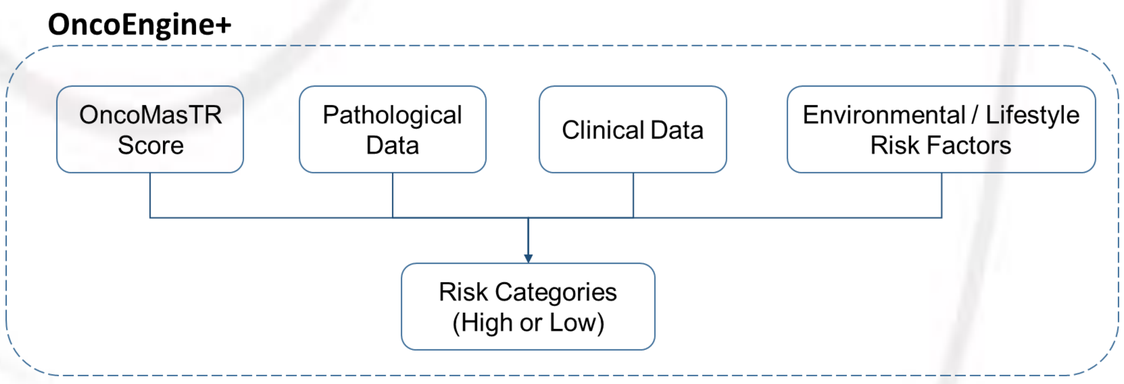

Figure 4. Input sources

❑Dataset

The Trial Assigning IndividuaLized Options for Treatment (Rx), or TAILORx, and The Cancer Genome Atlas (TCGA) are two very useful datasets which include genomic, epigenomic, transcriptomic, and proteomic data, clinical, and imaging data. TAILORx is sponsored by the National Cancer Institute (NCI) and TCGA joint effort between the National Cancer Institute and the National Human Genome Research Institute. These datasets are based on women recently diagnosed with Estrogen-Receptor positive, Her2/neu-negative breast cancer that had not yet spread to the lymph nodes.

❑Proposal

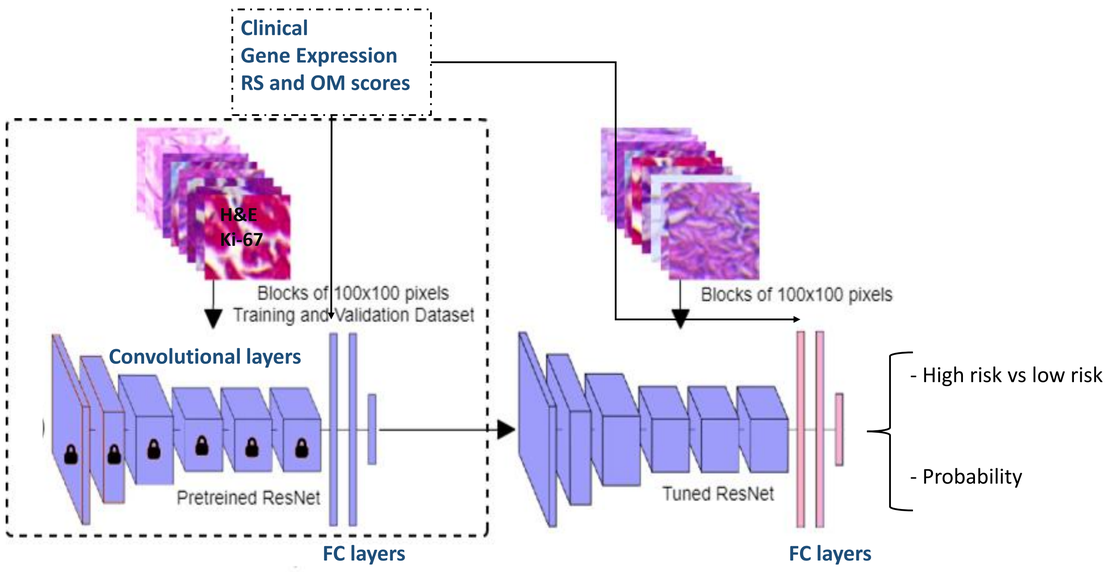

Figure 5. Histology-genomic integration proposal using transfer learning techniques

- Microenvironmental heterogeneity evaluation by ROIs:
  - Colour
  - Shape
  - Texture
  - Spatial
  - Morphometric
- Additional measurements from images: proliferation, total nuclei, +/- nuclei
- Use transfer learning to adapt state-of-the-art networks to our specific task
- Evaluate at two stages:
  - Compare all transferred networks to each other
  - Compare transferred networks vs individual approaches (genetic, clinical or histopathological)

❑Acknowledgement

This publication has emanated from research supported in part by a research grant from Science Foundation Ireland (SFI) under Grant Number 15/IA/3104. This material is based upon works supported by the Irish Cancer Society Collaborative Cancer Research Centre BREAST-PREDICT Grant CCRC13GAL.

❑Bibliography

1. National Cancer Institute: Surveillance, Epidemiology and End Results Program 2010-2012, **Incidence, GLOBOCAN 2012; ***healthcare spend on breast cancer in the EU in 20091.
2. Lanigan F, Brien GL, Fan Y, Madden SF, Jerman E, et al. Delineating transcriptional networks of prognostic gene signatures refines treatment recommendations for lymph node-negative breast cancer patients. FEBS J 2015; 282: 3455–3473
3. Moran B, Rahman, A, Palonen K, Lanigan F, Gallagher WM. Master transcriptional regulators in cancer: discovery via reverse engineering approaches and subsequent validation. CANCER RES 2017; 77(9): 2186-2190
4. Loughman T, Wang CJ, Dynoodt P, Fender B, Lopez-Ruiz C, et al. Analytical validation of OncoMasTR, a multigene test for predicting risk of distant recurrence in hormone receptor-positive early stage breast cancer. ANN ONCOL 2018, 29 (suppl; abstr 204P)
5. Sestak I, Buus R, Cuzick J, Barron S, Loughman T, et al. Evaluation of the OncoMasTR prognostic signature in postmenopausal women with ER-positive breast cancer. J CLIN ONCOL 2018, 36 (suppl; abstr 553)
6. O'Connor D, Kelly C, Crown J, Russell N, Barron S, et al. Additional prognostic value of OncoMasTR multigene prognostic signature to clinicopathological information in patients with HR-positive, HER2-negative, lymph node-negative breast cancer from the TAILORx Tissue Bank Ireland. ; Journal of Clinical Oncology 2019 37:15_suppl, 535.