In silico strategies to improve insight in breast cancer

Bense, Rico

DOI: 10.33612/diss.101935267

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date: 2019

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):

Copyright
Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

Take-down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): http://www.rug.nl/research/portal. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.
CHAPTER 1

General introduction
CHAPTER 1

BACKGROUND

Worldwide, breast cancer is the most commonly diagnosed type of cancer and the largest cause of cancer-specific death in women. Breast cancer is a heterogeneous disease in which multiple subtypes are recognized by means of immunohistochemical staining or microarray-profiling. These subtypes have specific biological features, distinct treatment targets, and varying prognoses.

Although these distinctions have been instrumental for improved breast cancer insight in the last decades, the awareness is increasing that there is even more variability between breast cancers than previously thought. This variability poses various challenges in the management of breast cancer. For instance, 5 years of adjuvant endocrine therapy has dramatically reduced the recurrence risk in patients with nonmetastatic estrogen receptor (ER)-positive breast cancer, but still a subset of patients experiences a recurrence more than 5 years after diagnosis. Also, not all patients benefit from standard systemic adjuvant treatment regimens, resulting in considerable undertreatment and overtreatment. Furthermore, for patients with the aggressive basal-like, usually triple-negative breast cancer, thus far no targeted treatment is available. This makes drug target finding for this subgroup of utmost importance. Therefore, although clear progression has been made in breast cancer treatment, patient outcome should still be improved by increasing the understanding of this disease and its subtypes.

Trials to assess treatment efficacy, as well as to validate diagnostics, are large scale and very long-term enterprises, excessively costly and difficult to be repeated. To reduce these debilitating factors, convenient low-cost tools for generating hypotheses in the context of clinical data would be helpful. Numerous gene expression data have been publicly stored in online databases such as the Gene Expression Omnibus and The Cancer Genome Atlas for all sorts of tissues since the emergence of microarray and RNA-Seq technology. Collecting these publicly available gene expression profiles and assembling them into a database could be used as a platform to generate strong hypotheses, which can ultimately facilitate insight into breast cancer.

In light of the above, we have compiled a large database of publicly available gene expression profiles from primary breast tumors and their corresponding clinicopathological data. Using this database, in this thesis we aim to gain more insight into how to select the best treatment for the right patients and to find potential new therapeutic strategies for difficult to treat subtypes of breast cancer.
THESIS OUTLINE

In ER-positive breast cancer, 20-25% of patients experience breast cancer recurrence despite 5 years of endocrine therapy.\(^4,10\) Half of these recurrences are late recurrences occurring more than 5 years after diagnosis.\(^7\) Reducing late recurrence risk by means of extending endocrine therapy beyond 5 years only benefits a subset of patients, and enhanced understanding of late recurrence might improve patient selection in this setting. Therefore, in \textit{chapter 2} we summarized the current knowledge of late recurrence biology, clinical trials on extended endocrine therapy, and tools for predicting late recurrence and benefit from treatment extension. Additionally, we performed a pooled analysis using 2,231 mRNA profiles of primary ER-positive/human epidermal growth factor receptor 2-negative breast cancers to identify biological pathways associated with an increased early or late recurrence risk.

Advances in neoadjuvant and adjuvant treatment in breast cancer have clearly improved patient outcome in recent decades. However, not all patients benefit from standard treatment regimens.\(^5,6\) New insights into the role of tumor-infiltrating immune cells suggest that their composition as well as their functionality might be relevant for breast cancer management.\(^8-11\) Therefore, in \textit{chapter 3} we assessed the independent predictive and prognostic value of 22 immune cell type fractions and immune gene signatures in breast cancer subtypes by performing several complementary \textit{in silico} analyses on gene expression profiles of 7,270 tumor samples of nonmetastatic breast cancer patients.

The androgen receptor (AR) is overexpressed in the majority of breast cancer and is considered to be an interesting drug target.\(^12\) In \textit{chapter 4} we summarized the role of the androgen receptor in breast cancer based on preclinical and clinical data. To gain insight in the role of AR in breast cancer, we additionally performed a retrospective pooled analysis using 7,270 mRNA profiles from primary breast tumors to explore the prognostic value of the AR in breast cancer subtypes.

Many cancers are characterized by genomic instability, which can result in the accumulation of structural chromosomal aberrancies such as copy number alterations (CNAs). CNAs are very common in many cancer types, including breast cancer,\(^13\) and promote tumor evolution by altering gene expression levels. Due to transcriptional adaptive mechanisms, however, changes in gene copy number at the genomic level do not always translate proportionally into altered gene expression levels. Despite current efforts, the degree of transcriptional adaptation to CNAs remains unclear for most genes. In \textit{chapter 5}, we reanalyzed >34,000 publicly available gene expression profiles from patient-derived
healthy tissue and tumor samples, and tumor cell line samples, which revealed the degree of transcriptional adaptation to CNAs in a genome-wide fashion. Based on our analyses and findings, we developed a platform-independent method – ‘transcriptional adaptation to CNA profiling’ (TACNA profiling) – that extracts the transcriptional effects of CNAs from gene expression profiles without requiring paired CNA profiles.

Reanalyzing publicly available gene expression profiles may also be useful for possible support of particular therapeutic targets. Mesothelin is a membrane-bound glycoprotein with limited expression in normal tissues. However, it has been shown to be overexpressed in several cancer types and is therefore considered an interesting potential drug target. Data on immunohistochemical assessment of mesothelin overexpression in cancer are mainly based on studies using small sample sizes, various assays or different definitions of positivity. In chapter 6 we determined overexpression of mesothelin using gene expression data of 19,746 unrelated, patient-derived tumor samples representing 41 tumor types. We applied functional genomic mRNA profiling, a method that corrects mRNA expression data for major, non-genetic factors (e.g. physiological, metabolic, and experimental factors), on these gene expression profiles to gain more detailed information about the potential of mesothelin as a generalizable drug target.

Finally, in chapter 7 the main findings of this thesis are summarized. This is followed by a discussion of the interpretations of these findings and future perspectives.

References


