Aus dem Institut für Pathologie
der Medizinischen Fakultät Charité – Universitätsmedizin Berlin

DISSERTATION

Integrative analysis of "omics" data and histopathological features in breast and ovarian cancer

zur Erlangung des akademischen Grades
Doctor medicinae (Dr. med.)

vorgelegt der Medizinischen Fakultät
Charité – Universitätsmedizin Berlin

von

Michael Bockmayr
aus München

Datum der Promotion: 10. März 2017

Table of contents

Abbreviations IV

Abstract V

1 Introduction 1

2 Methods 3

3 Results 5

4 Discussion 8

References 11

Affirmation in lieu of an oath/Eidesstattliche Erklärung 16

Declaration of own contributions 17

Publication 1: "New network topology approaches reveal differential correlation patterns in breast cancer" 19

Publication 2: "Classical pathology and mutational load of breast cancer – integration of two worlds" 33

Publication 3: "Prognostic impact of programmed cell death-1 (PD-1) and PD-ligand 1 (PD-L1) expression in cancer cells and tumor-infiltrating lymphocytes in ovarian high-grade serous carcinoma" 47

Curriculum vitae 61

Complete publication list 62

Acknowledgment 64

Abbreviations

ACSM1: Acyl-CoA synthetase medium chain 1
AR: Androgen receptor
BIRC5: Survivin
BRCA1/2: Breast cancer 1/2
CDH1: Cadherin-1
CDKN1B: Cyclin-dependent kinase inhibitor 1B (p27)
CI: Confidence interval
DC: Differential correlation
DE: Differential expression
ER: Estrogen receptor
FDR: False discovery rate
FOXC1: Forkhead box C1
GATA3: Transacting T-cell-specific transcription factor 3
GO: Gene Ontology
HER2: Human epidermal growth factor receptor 2
HR: Hormone receptor
ILC: Invasive lobular carcinoma
KEGG: The Kyoto Encyclopedia of Genes and Genomes
MAP3K1: Mitogen-activated protein kinase kinase kinase 1
MYBL2: Myb-related protein B
NCOR1: Nuclear receptor corepressor 1
NF1: Neurofibromin 1
NST: Invasive breast carcinoma of no special type
OS: Overall survival
PD-1: Programmed cell death-1
PD-L1: Programmed death-ligand 1
PFS: Progression-free survival
PGDH: Hydroxyprostaglandin dehydrogenase
PIK3CA: Phosphatidylinositol-4,5-bisphosphate 3-kinase, catalytic subunit alpha
PR: Progesterone receptor
PTPRD: Receptor-type tyrosine-protein phosphatase delta
RB1: Retinoblastoma 1
TCGA: The Cancer Genome Atlas
TIL: Tumor-infiltrating lymphocyte
TP53: Tumor suppressor 53

Abstract

Histopathological diagnosis of cancer is the basis of oncological therapy. However, during the last two decades, rapidly increasing amounts of high-dimensional molecular profiling data has become available. These data have been used to complement classical morphology- and immunohistology-based methods, but are still rarely used in the clinical practice. The aim of this work is to contribute to bridging the gap between these approaches.

First, we designed two algorithms for differential correlation analysis in gene expression data. Complementary to differential expression analysis, which searches for globally up- or downregulated genes, differential correlation analysis aims to identify groups of genes that exhibit different correlation patterns in two disease states. These algorithms were applied to compare subgroups of breast cancer defined by immunohistochemistry for the estrogen and the HER2 receptor. This permitted the identification of differentially correlated gene groups, which contain known and potentially new prognostic or predictive biomarkers. In particular, our analysis enabled the discovery of subtype specific divergences beyond the results of conventional differential expression analysis. Secondly, we analyzed the relationship between the mutational profile, histological tumor grade and subtypes of breast cancer. Notably, this analysis showed that the mutational load is significantly correlated with the tumor grade and the gene expression of known proliferation markers of breast cancer. Finally, we investigated the prognostic value of programmed cell death-1 (PD-1) and programmed death-ligand 1 (PD-L1) in gene expression data of ovarian high-grade serous carcinoma. High expression levels were associated with favorable prognosis and these results corroborate experimental findings based on immunohistochemistry.

In summary, the different parts of this work, although methodologically distinct, all contribute to an integrative analysis of the complex relations between different data modalities including high-dimensional molecular profiling data and (immuno-)histopathological features. This is one step towards the integration of new omics-based data with classical diagnostic approaches used in breast and ovarian cancer pathology.

Abstrakt

Die histopathologische Diagnose einer Krebserkrankung ist die Grundlage für onkologische Therapien. Gleichwohl sind in den letzten zwei Jahrzehnten eine stark zunehmende Menge an hochdimensionalen molekularen Profilierungsdaten verfügbar geworden. Diese Daten wurden dazu benutzt, um konventionelle auf Morphologie und Immunhistologie basierende Methoden zu ergänzen. In der klinischen Routinearbeit werden sie aber weiterhin selten eingesetzt. Ziel dieser Arbeit ist es dazu beizutragen, die Lücke zwischen diesen Ansätzen zu verringern.

In dem ersten Teil entwickelten wir zwei Algorithmen zur differentiellen Korrelationsanalyse von Genexpressionsdaten. Komplementär zur differentiellen Expressionsanalyse, welche nach global hoch- bzw. herunterregulierten Genen sucht, zielt die differentielle Korrelationsanalyse auf die Identifikation von Gengruppen, die in zwei Krankheitszuständen unterschiedliche Korrelationsmuster zeigen. Diese Algorithmen wurden anschließend angewendet, um die immunhistochemischen Subgruppen des Mammakarzinoms (eingeteilt nach Östrogenrezeptor- und HER2-Status) zu vergleichen. Dies ermöglichte die Identifizierung von Gruppen differentiell korrelierender Gene, welche bekannte und möglicherweise neue prognostische oder prädiktive Biomarker enthalten. Insbesondere ermöglichte dieser Ansatz, subtypenspezifische Unterschiede zu erkennen, die über die Ergebnisse konventioneller differentieller Expressionsanalysen hinausgehen. Im zweiten Teil untersuchten wir die Zusammenhänge zwischen Mutationsprofil, histologischem Tumorgrad und den Subtypen des Mammakarzinoms. Diese Analyse zeigte insbesondere, dass die Mutationslast signifikant mit dem Tumorgrad und der Genexpression bekannter Proliferationsmarker des Mammakarzinoms korreliert. Schließlich untersuchten wir den prognostischen Wert der Genexpression von *programmed cell death-1* (PD-1) und *programmed death-ligand 1* (PD-L1) im hochgradigen serösen Ovarialkarzinom. Hohe Expressionsniveaus waren mit einer günstigen Prognose assoziiert. Diese Resultate bekräftigen experimentelle immunhistologische Untersuchungen.

Zusammenfassend tragen die verschiedenen Teile dieser Arbeit, trotz unterschiedlicher Methodik, alle zu einer integrativen Untersuchung der komplexen Relationen zwischen verschiedenen Datentypen bei, insbesondere zwischen hochdimensionalen molekularen Profilierungsdaten und (immun-)histologischen Merkmalen. Dies ist ein Schritt hin zur Integration der neuen omik-basierten Daten mit klassischen diagnostischen Ansätzen in der Pathologie des Mamma- und Ovarialkarzinoms.

1 Introduction

The lifetime risk of developing cancer is currently around 51 % for men and 43 % for women in Germany [1]. Despite significant advances in cancer research over the last decades, oncological therapy remains very challenging and around 40 % of the patients die from their disease within 10 years of the initial diagnosis. Breast cancer, which is the most common cancer in females and main focus of this work, accounted for 30.8 % (69,550) of the newly diagnosed cases and 17.5 % (17,728) of the cancer-related deaths for women in Germany in 2012 [1].

In almost all cases, the final cancer diagnosis relies on the histopathological examination of bioptic or surgical tumor samples. The tumor classification depends primarily on the analysis of the morphological features on a standard (hematoxylin and eosin stained) tissue section. Tumor type and other properties such as growth patterns or the differentiation of the tumor cells (tumor grading) are determined based on these sections. Frequently, immunohistochemical stainings are necessary to obtain additional information. For instance, in breast cancer the hormone receptor (HR) status (i.e., the estrogen receptor (ER) and the progesterone receptor (PR) status), the human epidermal growth factor receptor 2 (HER2) status and the cellular growth rate (estimated by the percentage of tumor cells positive for the marker Ki67) may be determined. The HR and the HER2 status have direct clinical implications since the corresponding signaling pathways can be blocked by targeted therapies in patients with HR+ or HER2+ tumors improving their outcome significantly. Accordingly, the current classification of cancers as implemented by the WHO (cf. [2, 3] for an overview of the breast and ovarian cancer classification), primarily depends on morphological features and is only complemented by molecular data.

During the last two decades, there has been a rapidly increasing amount of newly available molecular data of cancer. These data are frequently referred to as “omics” data, where the suffix -omics summarizes the different fields from which the data arise, e.g., genomics (mutational data and copy number variations), transcriptomics (DNA-microarray gene expression data or RNA-seq data), proteomics and metabolomics. In 2001, Sørlie et al. used clustering of breast cancer gene expression data in their seminal work [4] to define new molecular subtypes, which were shown to be associated with distinct clinical outcomes. In the following years, advances in experimental techniques, in particular next-generation sequencing, allowed for the generation of huge amounts of comprehensive genetic profiling data from various cancers. Most notably, The Cancer Genome Atlas (TCGA) project, started in 2005, characterized around 11,000 tumors from 33 different entities and made the resulting genomic, transcriptomic and proteomic data publicly available [5].

These huge amounts of high-dimensional molecular profiling data are undoubtedly offering very promising opportunities for cancer research. Extraction of the relevant information and their interpretation can, however, be very difficult. Indeed, the analysis of

the data cannot be performed manually, and requires novel computational methods and algorithms. Furthermore, statistical analysis of these data does not necessarily provide new insights in the pathogenetic disease mechanisms, in particular if it is not associated with prior biological knowledge [6]. The aim of this work is to contribute to bridging the gap between the new omics-based research and classical morphology- and immunohistology-based pathology. To this end, we developed novel and applied existing quantitative methods for the analysis of "omics" data integrating histopathological features related to cancer pathology.

In the first part of this thesis, we designed algorithms to identify differentially correlated genes between two disease states. We applied these methods to determine subtype specific gene expression patterns of the breast cancer subgroups as defined by immunohistochemistry (ER+ vs. ER-, HER2+ vs. HER2-). Differential expression analysis (DE) [7] has widely been used for this purpose [8–10]. Differential correlation (DC) analysis is a complementary approach (see for example [11–18]), which aims for a deeper understanding of disease-specific gene expression patterns. Indeed, differentially activated signaling pathways resulting in highly coordinated gene expression patterns are not necessarily associated with global up- or downregulation of gene sets. These expression patterns can be identified by DC analysis, but may be overlooked by DE analysis. Previous approaches developed different strategies to compare correlation networks of two different disease conditions [14–16]. In general, building these correlation networks requires setting a threshold. The genes under investigation are represented by vertices and they are connected by edges if the Pearson correlation of their gene expression profiles exceeds the given threshold. However, the network topology can differ significantly between two thresholds. As a novelty, our algorithms do not only compare networks built for a unique threshold, but investigate a comprehensive series of networks built for 100 or 200 thresholds. This allows for a robust detection of different kinds of correlation changes, strong changes of a small number of genes and moderate changes of many genes.

In the second part, we analyzed the relationship between the mutational profile, the histological tumor grade, the molecular subtypes and gene expression data in breast cancer. As mentioned previously, daily routine diagnosis of breast cancer is still based on clinical, morphology- and immunohistology-based features. To refine this approach, multigene signatures have been proposed to classify the tumors according to their prognosis (such as the PAM50 classifier [19]) or to assess the risk of recurrence after surgery without chemotherapy [20–22]. However, mutational data is not used in routine diagnostics with the exception of the *BRCA1/2* screening. Therefore, our analysis aims at integrating mutational profiling data and classical pathology of breast cancer.

Finally, we analyzed the prognostic value of programmed cell death-1 (PD-1) and programmed death-ligand 1 (PD-L1) in high-grade serous ovarian carcinoma. PD-1 is an inhibitory receptor frequently expressed on tumor-infiltrating lymphocytes (TILs). TILs

were shown to be associated with increased survival in this tumor entity [23–25]. PD-L1, which is commonly present on tumor cells, is a ligand of this receptor and its expression inhibits antitumoral T cell response in different mouse models (reviewed in [26]). Several clinical trials have investigated immune checkpoint inhibitors, i.e., molecules that target PD-1 or PD-L1 and block their interaction, in various cancers (reviewed in [26]). These molecules have been approved for the treatment of metastatic melanoma and advanced non-small cell lung cancer. However, the precise functioning of the PD-1/PD-L1 pathway in ovarian cancer, especially in presence of other mediators, remains unclear. In our paper, PD-1 and PD-L1 expression patterns determined by immunohistochemistry in cancer cells and TILs were systematically investigated. Our contribution to this work was focused on the analysis of PD-1 and PD-L1 gene expression in data from TCGA, which we related to the immunohistological findings.

In summary, the different parts of this work, although methodologically distinct, have in common that they all allow for the analysis of complex relationships between different data modalities, including molecular profiling and histopathological data.

2 Methods

All the statistical analyses, the implementation of the algorithms to investigate differential correlation and most of the data visualization were performed using the statistical programming language `R` [27].

Datasets

The dataset used for differential correlation analysis was obtained by fusion of 6 publicly available microarray datasets of breast cancer (GSE1456, GSE2034, GSE4922, GSE6532, GSE7390 and GSE11121) from Gene Expression Omnibus [28]. The ER and the HER2 status of the samples were determined from the gene expression level of the corresponding genes. The clinical information, the mutational data and the RNA-seq data of breast cancer used in the second part of this work were obtained from TCGA [29]. The clinical data and the transcriptomics data of ovarian cancer used in the third part for the *in silico* validation of the experimental results were also obtained from TCGA [30]. All these data were previously published and available without limitations.

Differential network analysis

We developed two algorithms to compare the global (`DCglob`) and the local (`DCloc`) topology of correlation networks. In such a network, genes are represented by vertices and they are connected by an edge if the Pearson correlation of the corresponding gene expression profiles exceeds a given threshold t. Let us assume that we compare two disease conditions

A and B. Since the network topology strongly depends on the choice of the cutoff, we did, as a novelty, not only compare two fixed networks, but a series of k (typically 100 or 200) networks $(N_A^t), (N_B^t)$, $t \in 1, \ldots, k$. `DCglob` compares the evolution of connected components, i.e., parts of the network in which any two vertices are connected by a path. For each threshold t, the algorithm computes the connected components in the networks N_A^t and N_B^t. A gene is considered to be differentially correlated for the threshold t if it is member of a connected component with at least 3 vertices in the network constructed for one of the disease conditions but not the other. This yields (potentially empty) sets of intervals containing threshold values for which a gene is differentially correlated. The length of the longest interval in this set is then converted into a score characterizing the strength of differential correlation (DC) for every gene. `DCloc` focuses on the evolution of the local gene neighborhood for the different thresholds. For every gene and for every threshold, the number of common next neighbors in both networks N_A^t and N_B^t divided by the total number of next neighbors is computed. This score is high if the neighborhoods are similar and low if the neighborhoods are dissimilar. The results for the different thresholds are averaged yielding a measure of differential correlation for every gene. Finally, both algorithms return lists of differentially correlated genes showing higher correlation in one of the disease states compared to the other, which can be ordered by the strength of DC. The algorithms are untargeted in the sense that they do not rely on any prior biological knowledge. In particular, they do not compare predefined gene modules but they are able to detect DC in general situations. For a more precise description of the algorithms, we refer to the methods section of the corresponding paper [31].

Association of clinico-pathological parameters and “omics” data

To identify associations of clinico-pathological parameters and “omics” data or associations between different types of “omics” data (e.g., mutational load and RNA-seq), we used several statistical methods and tests, including the Spearman correlation, Wilcoxon's test, Welch's test and the Jonckheere-Terpstra test. In general, p-values < 0.05 were considered statistically significant. Multiple testing correction using the Bonferroni or the Benjamini-Hochberg method was applied whenever appropriate. More details are provided in the methods section of the corresponding papers [31, 32].

Data visualization and functional analysis

The statistical analysis of the first two papers in this thesis frequently resulted in lists of candidate genes, which were, for instance, differentially correlated [31] or associated to the mutational load [32]. For the visualization and interpretation of the differentially correlated genes, we used heatmaps and hierarchical clustering as implemented in `R` as well as network representations prepared with `Cytoscape` [33]. We applied gene set enrichment

analysis for a functional description of the resulting genes in both papers. To this end, the overlap of the resulting genes and categories from established databases, like the Gene Ontology (GO)[34], the Kyoto Encyclopedia of Genes and Genomes (KEGG) [35] or the Reactome pathway [36] was computed and assessed for significance using Fisher's exact test.

Survival Analysis

We used the Cox proportional hazards model as implemented in the `R` package `survival` to assess the continuous influence of covariates on the outcome [27, 37]. For dichotomous analyses, the Cutoff Finder software [38] was used to find optimal cutoff points for biomarkers. The statistical significance was assessed using the logrank test.

3 Results

Differential correlation in breast cancer

Reference: M. Bockmayr, F. Klauschen, B. Györffy, C. Denkert and J. Budczies: New network topology approaches reveal differential correlation patterns in breast cancer; BMC Syst Biol. 2013 Aug 15;7(1):78

We developed two untargeted algorithms (`DCloc` and `DCglob`) that are capable of identifying differential correlation patterns from microarray data for two disease conditions. The construction of correlation networks, which serve as input for the algorithms, requires the selection of a correlation threshold above which the vertices (genes) are connected by edges. As a novelty, our algorithms do not only investigate a single network constructed for a fixed threshold but systematically analyze networks constructed for a comprehensive series of 100 to 200 thresholds covering the full range of positive correlations. This allows for the detection of different kinds of correlation changes at the same level of significance: strong changes of a few genes and moderate changes of many genes. We applied the algorithms on a large breast cancer microarray dataset (1317 samples) obtained by fusion of 6 publicly available datasets and compared the ER+ vs. ER- and the HER2+ vs. HER2- subtypes. The false discovery rate (FDR) was estimated using a repeated random subsampling analysis. Using `DCglob`, 630 differentially correlated genes (FDR = 12.1 %) were detected between the ER subtypes and 804 (FDR = 9.5 %) between the HER2 subtypes. Using `DCloc`, 770 differentially correlated genes (FDR = 12.8 %) were detected between the ER subtypes and 1027 (FDR = 9.6 %) between the HER2 subtypes. We performed a two-fold cross-validation to assess the reproducibility of our results. The overlap of the top 5 % differentially correlated genes comparing distinct sets of 140 ER- tumors and 140 ER+ tumors was 49 % for `DCloc` and 33 % for `DCglob`. Gene

set enrichment analysis was executed on the resulting gene lists and revealed numerous significantly enriched gene sets, in particular cell cycle genes, for both analyses. The resulting genes were also visualized using heatmaps and a network representation. The clusters of genes showing higher correlation in ER- compared to ER+ breast cancer were shown to be associated with marker genes of previously described breast cancer subtypes, including invasive apocrine carcinomas (IAC) [39], the HER2+ subtype [40], an androgen receptor (AR) responsive subtype [41], and the *FOXC1* subtype [42]. Remarkably, our algorithms detected several significantly differentially correlated genes (between 23 % and 53 % for the different analyses) that were not differentially expressed, including two of the markers for IAC, hydroxyprostaglandin dehydrogenase (*PGDH*) and acyl-CoA synthetase medium chain 1 (*ACSM1*).

Mutational load and classical pathology of breast cancer

Reference: J. Budczies, M. Bockmayr*, C. Denkert, F. Klauschen, J.K. Lennerz, B. Györffy, M. Dietel, S. Loibl, W. Weichert and A. Stenzinger: Classical pathology and mutational load of breast cancer – integration of two worlds; J Path: Clin Res 2015 Oct;1(4):225–238 (* J. Budczies and M. Bockmayr contributed equally to this work.)*

The goal of this work was to link the two worlds of histopathology and multi-layered molecular profiling in breast cancer. The most relevant histopathological characteristics that influence clinical decision-making are the tumor type [2], the ER and HER2 status, which is generally determined by immunohistochemistry, as well as the tumor grading, which is based on the nuclear morphology, the mitotic rate and the presence of tubule formation. Here, we performed an integrated analysis to elucidate the relationships between molecular data (somatic mutations and RNA-seq) and the aforementioned histopathological features. To this end, we evaluated the number of genes with non-silent somatic mutations in a cohort of 687 primary breast cancer patients from TCGA. The number of mutated genes was strongly associated with the tumor grade, increasing from a median of 23 mutated genes in G1 tumors via 27 in G2 tumors to 43 in G3 tumors (p = 1.4e-14). It was also associated with the immunohistochemical subtype with a median number of mutations increasing from 27 in ER+/HER2- via 39.5 in ER+/HER2+ via 41 in ER-/HER2+ to 49 in ER-/HER2- (p = 1.4e-10) and the molecular subtype as determined by the PAM50 classifier [19] (p = 4.3e-10). The two main histological subtypes, i.e., invasive breast carcinoma of no special type (NST) and invasive lobular carcinoma (ILC) were not associated with a significantly different number of mutated genes. Interestingly, nodal positive tumors had a slightly lower median number of mutations than nodal negative tumors (34 vs. 30; p = 0.0048). Second, evaluating the relationship between the mutational load and recurrently mutated genes, we found that a high number of mutated genes was significantly associated with mutations in

TP53, *NCOR1*, *NF1*, *PTPRD* and *RB1*, but not with mutated *PIK3CA*. Furthermore, we assessed the correlation between the mutational burden and gene expression. We observed significant associations ($|R| > 0.4$) between the abundance of mutated genes and expression levels of genes related to proliferation in the overall and the ER+ cohort, including the Recurrence Score gene signature [20] (e.g., *MYBL2* and *BIRC5*). Specific genes, including *TP53*, *GATA3*, *CDKN1B*, *PIK3CA*, *CDH1*, *MAP3K1* showed characteristic associations with tumor grade, immunohistochemical and PAM50 subtype. Finally, in a dichotomized multivariate analysis of overall survival using Cutoff Finder [38], a larger number of mutations (> 21) was associated with worse overall survival (hazard ratio = 4.6, 95 % CI: 1.0 – 20.0, $p = 0.044$). To sum up, we provided evidence that specific mutational patterns underlie different morphological and biological phenotypes in breast cancer.

Prognostic impact of programmed cell death-1 (PD-1) and PD-ligand 1 (PD-L1) expression in cancer

Reference: S. Darb-Esfahani, C.A. Kunze*, H. Kulbe, J. Sehouli, S. Wienert, J. Lindner, J. Budczies, M. Bockmayr, M. Dietel, C. Denkert, I. Braicu and K. Jöhrens: Prognostic impact of programmed cell death-1 (PD-1) and PD-ligand 1 (PD-L1) expression in cancer cells and tumor-infiltrating lymphocytes in ovarian high-grade serous carcinoma; Oncotarget 2015 Nov;7(2):1486-1499 (* S. Darb-Esfahani and C.A. Kunze contributed equally to this work.)*

In this paper, the prognostic value of PD-1 and PD-L1 expression in high-grade serous ovarian carcinoma tumor cells and tumor-infiltrating lymphocytes (TILs) was investigated. We mainly contributed to the analysis of gene expression data from TCGA. Immunohistochemistry was used to detect the expression of PD-1 and PD-L1 in cancer cells (201 and 202 available cases, respectively) and the expression of PD-1, PD-L1 and CD3 in TILs (200 cases). Furthermore, mRNA of PD-1 and PD-L1 was measured using quantitative reverse transcription PCR (200 and 204 available cases, respectively). PD-1 and PD-L1 expression in cancer cells, CD3+, PD-1+, and PD-L1+ TILs densities as well as PD-1 and PD-L1 mRNA levels were positive prognostic factors for progression-free (PFS) and overall survival (OS), with all factors being significant for PFS ($p < 0.035$ each), and most being significant for OS. Furthermore, tumors with high PD-1+ TILs or PD-L1+ TILs density in addition to high CD3+ TILs had a better prognosis (both PFS and OS) than tumors with low PD-1+ or PD-L1+ TIL counts despite of a high CD3 infiltration (significance for PFS: PD-1+: $p = 0.002$; PD-L1+: $p = 0.002$). Finally, an *in silico* validation using the high-grade serous ovarian carcinoma gene expression dataset from TCGA [30] was performed. The prognostic value as to OS was assessed in three platforms (Affymetrix, Agilent, RNA-seq) for PD-1 and two platforms (Agilent, RNA-seq) for PD-L1. Cutoff Finder was used for the determination of cutoff points [38]. PD-L1

expression was a robust positive prognostic factor in the total study cohort (Agilent: 113 out of 444 cutoffs significant (25.5 %); optimal cutoff $p < 0.0001$, RNA-seq: 117 out of 380 cutoffs significant (30.8 %), optimal cutoff: $p < 0.0001$). PD-1 expression was also a positive prognostic factor for the total cohort. However, its prognostic value was of reduced robustness as only few cutoffs were significant: Agilent: 14 out of 460 cutoffs significant (3.0 %), optimal cutoff $p = 0.02$, which was also seen in Affymetrix data (36 out of 445 cutoffs significant (7.9 %), optimal cutoff $p = 0.013$), however missed significance in RNA-seq data ($p = 0.065$).

4 Discussion

In this thesis, we present different quantitative methodologies for an integrative analysis of high-dimensional data, which permits a more precise description of tumor subgroups and characteristics defined by histopathological features.

In the first paper, we developed two untargeted algorithms for the detection of differential correlation patterns in microarray data and demonstrated their efficacy on a large breast cancer dataset. Previous approaches to study differential correlation [14–16] compare correlation patterns between two fixed networks. This kind of correlation networks is built by fixing a threshold and connecting with an edge all the genes showing a correlation exceeding the cutoff. These approaches neglect the fact that the topology of a correlation network is heavily influenced by the choice of the threshold. Therefore, as a novelty, our algorithms do not only compare two networks that are constructed for fixed thresholds, but for a comprehensive series of 100 (or 200) thresholds, covering the full range of positive correlations. This yielded more robust results than the classical approaches. While differential expression (DE) of genes between breast cancer subtypes has already been extensively studied (see for example [8–10]), our work was one of the first untargeted attempts to characterize differential correlation in this disease (cf. [17] for a *de novo* partitioning method or [18] for a targeted analysis of KEGG pathways). Our results showed that DC analysis provides insights beyond the results of ordinary DE analysis. We were able to identify several relevant genes that are differentially correlated but not differentially expressed. Indeed, subtype specific changes in the correlation structure, which could be mediated by the activation of a specific signaling pathway or transcription factor, are not necessarily associated with up- or downregulated gene expression in the full subgroup, and might be overlooked by DE analysis. The algorithms identified numerous clusters of genes that show highly significant correlations in one of the subtypes but not in the other. Several of these clusters, especially in the ER- group, could be associated with marker genes of already known breast cancer subtypes [39–42]. Others might contain new prognostic or predictive biomarkers or possible therapeutic targets. However, further studies are needed

to elucidate the clinical role of these genes. The use of the algorithms `DCglob` and `DCloc` is not restricted to the analysis of microarray data. It can easily be translated to other research areas. For example, we used a slightly modified version (with another metric) of the algorithm `DCglob` to identify differentially correlated phosphosites in phosphoproteomic time series data of lung cancer cell lines [43].

In the second paper, we investigated the relationship of histopathological features of breast cancer, including immunohistochemical subtype and grade, and the mutational load. The clinical staging and histopathological features are currently used to tailor specific therapies and estimate the outcome of breast cancer. In addition, molecular data has become more important in clinical oncology over the past decade. As an example, new guidelines [44] are based on the molecular subtype of breast cancer [4], which is in principle determined from gene expression data. Nevertheless, the molecular subtype is not determined in routine diagnostics from gene expression data, but only approximated by the immunohistochemical subtype. Furthermore, several multigene signatures have been proposed to evaluate the risk of recurrence of breast cancer and are increasingly used in the clinical praxis [20–22]. However, with exception of the *BRCA1/2* screening for familiar breast cancer, mutational data has not yet been integrated into the diagnostic, nor the therapeutic process. In this study, we showed that the tumor grade is highly significantly associated with the mutational load, which is itself assumed to be a measure of the genetic heterogeneity of a tumor [45]. Hence, the tumor grade might by a microscopic readout of the tumor's genetic heterogeneity. Many of the genes correlated with high mutational loads (*TP53*, *NCOR1*, *NF1*, *PTPRD* and *RB1*) have been associated with genomic instability previously. However, the most frequent mutation in breast cancer, *PIK3CA*, was not associated with an increased mutational load. Thus, we can hypothesize that these mutations have different functional relevance and occur at different moments of tumor evolution. Interestingly, the number of mutated genes was not positively correlated with the nodal status of the tumor, which is one of the most relevant prognostic factors, suggesting that a large number of mutations is not necessarily linked to a more aggressive metastatic behavior. The subtype-specific mutational profiles of diverse oncogenes, in particular *TP53*, *GATA3*, *CDKN1B*, *PIK3CA*, *CDH1*, *MAP3K1* suggest different mechanisms of tumor evolution in the immunohistochemical subtypes. Therefore, further investigation of the clinical impact and the biological function of these mutations in a histopathological or subtype-specific context is required. Although our results are mostly observational, they are a step towards genomics-informed breast pathology.

In the final paper, the prognostic impact of the biomarkers PD-1 and PD-L1 in ovarian high-grade serous carcinoma was studied. We mainly contributed to the *in silico* validation of the experimental results using gene expression data provided by TCGA. Indeed, the publicly accessible repositories of "omics" data do not only offer opportunities to generate hypotheses, as done in the first two papers, but also allow for the validation of already

established biological conclusions. High expression of PD-1 and PD-L1 on cancer cells as well as high densities of PD-1+ and PD-L1+ TILs were positive prognostic factors in ovarian high-grade serous carcinoma. These results were surprising since PD-1 and PD-L1 are both reported to have an immune-inhibitory function. Controversial results might partly be related to the lack of standardization of PD-1 antibodies. Even within our study different antibodies for PD-1 showed distinct staining results. Indeed, PD-1 expression in cancer was observed using one carefully validated antibody, while using another antibody, which produced quite similar staining results for TILs, PD-1 cancer cell expression was not seen. Nevertheless, the results from our study corroborate the potential importance of immune-checkpoints in this tumor entity. Further work, however, is required to characterize the complex and multilayered interactions between cancer cells and the immune system, in this and other tumor entities. In previous work [46], we investigated the spatial relationship of cancer cells and immune infiltrates in breast cancer using spatial statistics methods, in particular Ripley's K-function. On the one hand, we could confirm that the quantity of TILs was a positive prognostic factor in ER- but not in ER+ breast cancer in agreement with previous results (see for example [47, 48]). On the other hand, we could surprisingly not identify any prognostic relevance of the local spatial patterns (e.g., clustering vs. repulsion) of these two cell types, which we expected to be a morphological portrait of their interaction. However, it might be interesting to reevaluate these spatial features in combination with immunohistochemical data, as, for instance, the PD-1 and PD-L1 expression on cancer and immune cells to get a better understanding of this process. Indeed, the functional relevance of the PD-L1/PD-1 pathway is still not well understood and there are contradictory results in other cancer types [49].

A precise characterization of specific tumor types together with the identification of prognostic and predictive biomarkers is of outstanding importance for clinical oncology. Despite the indisputable success of omics-based cancer research in the last decades, this is still mostly effectuated with classical histopathological methods in the clinical routine. We identified new elements that allow for a more precise, molecular characterization of established disease conditions and features that are used in the current diagnostic framework. We determined several clusters of genes that showed specific correlation patterns in the immunohistochemical subtypes of breast cancer. Some of these clusters were composed of known marker genes of clinically relevant subtypes, while others might contain new biomarkers or therapeutic targets. Furthermore, we provided a portrait of the mutational landscape, which was associated with breast cancer grading and molecular subtypes. Finally we validated findings on PD-1 and PD-L1 expression based on immunohistochemistry on an independent gene expression dataset from TCGA. All these results contribute towards a more quantitative and less biased form of omics-informed pathology. However, it is questionable that a purely genetic or transcriptomic profiling might be able to fully elucidate the complex pathogenic mechanisms underlying oncological diseases. Indeed, the biological

impact of well-defined mutations, as for instance BRAFV600E, is different in nevi, malignant melanoma, hyperplastic polyps of the colon, and colorectal cancer [50]. Although there are large similarities between the mutational profiles of many cancers, this is not necessarily reflected by their biological and clinical characteristics [51]. As discussed above, the role of TILs and immune markers like PD-1 is manifestly not the same in different tumor entities. Therefore, the full complexity of cancer can only be understood by an integrated analysis combining omics-based data, macroscopic and microscopic morphological information on the tumor and its environment, and clinical knowledge. Making sense of these huge amounts of multilayered high-dimensional data requires development and application of mathematical and computational methods able to incorporate the full, highly convoluted information. The work presented here is one step in this direction.

References

1. Krebs in Deutschland 2011/2012. *10. Ausgabe. Robert Koch-Institut (Hrsg) und die Gesellschaft der epidemiologischen Krebsregister in Deutschland e.V. (Hrsg). Berlin* (2015).
2. Sinn, H. P. & Kreipe, H. A Brief Overview of the WHO Classification of Breast Tumors, 4th Edition, Focusing on Issues and Updates from the 3rd Edition. *Breast Care (Basel)* **8,** 149–154 (2013).
3. Meinhold-Heerlein, I., Fotopoulou, C., Harter, P., Kurzeder, C., Mustea, A., Wimberger, P., Hauptmann, S. & Sehouli, J. The new WHO classification of ovarian, fallopian tube, and primary peritoneal cancer and its clinical implications. *Arch. Gynecol. Obstet.* **293,** 695–700 (2016).
4. Sørlie, T., Perou, C. M., Tibshirani, R., Aas, T., Geisler, S., Johnsen, H., Hastie, T., Eisen, M. B., van de Rijn, M., Jeffrey, S. S., Thorsen, T., Quist, H., Matese, J. C., Brown, P. O., Botstein, D., Lønning, P. E. & Børresen-Dale, A. L. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc. Natl. Acad. Sci. U.S.A.* **98,** 10869–10874 (2001).
5. *The Cancer Genome Atlas* `http://cancergenome.nih.gov/`. National Cancer Institute & National Human Genome Research Institute.
6. Weinberg, R. Point: Hypotheses first. *Nature* **464,** 678 (2010).
7. Allison, D. B., Cui, X., Page, G. P. & Sabripour, M. Microarray data analysis: from disarray to consolidation and consensus. *Nat. Rev. Genet.* **7,** 55–65 (2006).
8. Gruvberger, S., Ringnér, M., Chen, Y., Panavally, S., Saal, L. H., Borg, A., Fernö, M., Peterson, C. & Meltzer, P. S. Estrogen receptor status in breast cancer is associated with remarkably distinct gene expression patterns. *Cancer Res* **61** (2001).
9. Van't Veer, L., Dai, H., van de Vijver, M., He, Y., Hart, A., Mao, M., Peterse, H., van der Kooy, K., Marton, M., Witteveen, A., Schreiber, G., Kerkhoven, R., Roberts, C., Linsley, P., Bernards, R. & Friend, S. Gene expression profiling predicts clinical outcome of breast cancer. *Nature* **415** (2002).

10. Budczies, J., Weichert, W., Noske, A., Muller, B. M., Weller, C., Wittenberger, T., Hofmann, H. P., Dietel, M., Denkert, C. & Gekeler, V. Genome-wide gene expression profiling of formalin-fixed paraffin-embedded breast cancer core biopsies using microarrays. *J. Histochem. Cytochem.* **59,** 146–157 (2011).
11. Lai, Y., Wu, B., Chen, L. & Zhao, H. A statistical method for identifying differential gene-gene co-expression patterns. *Bioinformatics* **20,** 3146–3155 (2004).
12. Li, K. C. Genome-wide coexpression dynamics: theory and application. *Proc. Natl. Acad. Sci. U.S.A.* **99,** 16875–16880 (2002).
13. Kostka, D. & Spang, R. Finding disease specific alterations in the co-expression of genes. *Bioinformatics* **20 Suppl 1,** i194–199 (2004).
14. Choi, J. K., Yu, U., Yoo, O. J. & Kim, S. Differential coexpression analysis using microarray data and its application to human cancer. *Bioinformatics* **21,** 4348–4355 (2005).
15. Altay, G., Asim, M., Markowetz, F. & Neal, D. E. Differential C3NET reveals disease networks of direct physical interactions. *BMC Bioinformatics* **12,** 296 (2011).
16. Voy, B. H., Scharff, J. A., Perkins, A. D., Saxton, A. M., Borate, B., Chesler, E. J., Branstetter, L. K. & Langston, M. A. Extracting gene networks for low-dose radiation using graph theoretical algorithms. *PLoS Comput. Biol.* **2,** e89 (2006).
17. Freudenberg, J. M., Sivaganesan, S., Wagner, M. & Medvedovic, M. A semi-parametric Bayesian model for unsupervised differential co-expression analysis. *BMC Bioinformatics* **11,** 234 (2010).
18. Tegge, A. N., Caldwell, C. W. & Xu, D. Pathway correlation profile of gene-gene co-expression for identifying pathway perturbation. *PLoS ONE* **7,** e52127 (2012).
19. Parker, J. S., Mullins, M., Cheang, M. C., Leung, S., Voduc, D., Vickery, T., Davies, S., Fauron, C., He, X., Hu, Z., Quackenbush, J. F., Stijleman, I. J., Palazzo, J., Marron, J. S., Nobel, A. B., Mardis, E., Nielsen, T. O., Ellis, M. J., Perou, C. M. & Bernard, P. S. Supervised risk predictor of breast cancer based on intrinsic subtypes. *J. Clin. Oncol.* **27,** 1160–1167 (2009).
20. Van de Vijver, M. J., He, Y. D., van't Veer, L. J., Dai, H., Hart, A. A., Voskuil, D. W., Schreiber, G. J., Peterse, J. L., Roberts, C., Marton, M. J., Parrish, M., Atsma, D., Witteveen, A., Glas, A., Delahaye, L., van der Velde, T., Bartelink, H., Rodenhuis, S., Rutgers, E. T., Friend, S. H. & Bernards, R. A gene-expression signature as a predictor of survival in breast cancer. *N. Engl. J. Med.* **347,** 1999–2009 (2002).
21. Paik, S., Shak, S., Tang, G., Kim, C., Baker, J., Cronin, M., Baehner, F. L., Walker, M. G., Watson, D., Park, T., Hiller, W., Fisher, E. R., Wickerham, D. L., Bryant, J. & Wolmark, N. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *N. Engl. J. Med.* **351,** 2817–2826 (2004).
22. Kronenwett, R., Bohmann, K., Prinzler, J., Sinn, B. V., Haufe, F., Roth, C., Averdick, M., Ropers, T., Windbergs, C., Brase, J. C., Weber, K. E., Fisch, K., Muller, B. M., Schmidt, M., Filipits, M., Dubsky, P., Petry, C., Dietel, M. & Denkert, C. Decentral gene expression analysis: analytical validation of the Endopredict genomic multianalyte breast cancer prognosis test. *BMC Cancer* **12,** 456 (2012).

23. Bachmayr-Heyda, A., Aust, S., Heinze, G., Polterauer, S., Grimm, C., Braicu, E. I., Sehouli, J., Lambrechts, S., Vergote, I., Mahner, S., Pils, D., Schuster, E., Thalhammer, T., Horvat, R., Denkert, C., Zeillinger, R. & Castillo-Tong, D. C. Prognostic impact of tumor infiltrating CD8+ T cells in association with cell proliferation in ovarian cancer patients–a study of the OVCAD consortium. *BMC Cancer* **13,** 422 (2013).
24. Clarke, B., Tinker, A. V., Lee, C. H., Subramanian, S., van de Rijn, M., Turbin, D., Kalloger, S., Han, G., Ceballos, K., Cadungog, M. G., Huntsman, D. G., Coukos, G. & Gilks, C. B. Intraepithelial T cells and prognosis in ovarian carcinoma: novel associations with stage, tumor type, and BRCA1 loss. *Mod. Pathol.* **22,** 393–402 (2009).
25. Webb, J. R., Milne, K., Watson, P., Deleeuw, R. J. & Nelson, B. H. Tumor-infiltrating lymphocytes expressing the tissue resident memory marker CD103 are associated with increased survival in high-grade serous ovarian cancer. *Clin. Cancer Res.* **20,** 434–444 (2014).
26. Pardoll, D. M. The blockade of immune checkpoints in cancer immunotherapy. *Nat. Rev. Cancer* **12,** 252–264 (2012).
27. R Core Team. *R: A Language and Environment for Statistical Computing* R Foundation for Statistical Computing (Vienna, Austria, 2013). <`http://www.R-project.org/`>.
28. Edgar, R., Domrachev, M. & Lash, A. E. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.* **30,** 207–210 (2002).
29. The Cancer Genome Atlas Research Network. Comprehensive molecular portraits of human breast tumours. *Nature* **490,** 61–70 (2012).
30. The Cancer Genome Atlas Research Network. Integrated genomic analyses of ovarian carcinoma. *Nature* **474,** 609–615 (2011).
31. Bockmayr, M., Klauschen, F., Gyorffy, B., Denkert, C. & Budczies, J. New network topology approaches reveal differential correlation patterns in breast cancer. *BMC Syst Biol* **7,** 78 (2013).
32. Budczies, J., Bockmayr, M., Denkert, C., Klauschen, F., Lennerz, J., Györffy, B., Dietel, M., Loibl, S., Weichert, W. & Stenzinger, A. Classical pathology and mutational load of breast cancer – integration of two worlds. *J Path: Clin Res* **1,** 225–238 (2015).
33. Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N., Schwikowski, B. & Ideker, T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13,** 2498–2504 (2003).
34. Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M. & Sherlock, G. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* **25,** 25–29 (2000).
35. Kanehisa, M. & Goto, S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28,** 27–30 (2000).
36. Milacic, M., Haw, R., Rothfels, K., Wu, G., Croft, D., Hermjakob, H., D'Eustachio, P. & Stein, L. Annotating cancer variants and anti-cancer therapeutics in reactome. *Cancers (Basel)* **4,** 1180–1211 (2012).

37. Therneau, T. M. *A Package for Survival Analysis in S* version 2.38 (2015). <`http://CRAN.R-project.org/package=survival`>.

38. Budczies, J., Klauschen, F., Sinn, B. V., Györffy, B., Schmitt, W. D., Darb-Esfahani, S. & Denkert, C. Cutoff Finder: a comprehensive and straightforward Web application enabling rapid biomarker cutoff optimization. *PLoS ONE* **7,** e51862 (2012).

39. Celis, J. E., Gromov, P., Cabezon, T., Moreira, J. M., Friis, E., Jirstrom, K., Llombart-Bosch, A., Timmermans-Wielenga, V., Rank, F. & Gromova, I. 15-prostaglandin dehydrogenase expression alone or in combination with ACSM1 defines a subgroup of the apocrine molecular subtype of breast carcinoma. *Mol. Cell Proteomics* **7,** 1795–1809 (2008).

40. Staaf, J., Jonsson, G., Ringner, M., Vallon-Christersson, J., Grabau, D., Arason, A., Gunnarsson, H., Agnarsson, B. A., Malmstrom, P. O., Johannsson, O. T., Loman, N., Barkardottir, R. B. & Borg, A. High-resolution genomic and expression analyses of copy number alterations in HER2-amplified breast cancer. *Breast Cancer Res.* **12,** R25 (2010).

41. Doane, A. S., Danso, M., Lal, P., Donaton, M., Zhang, L., Hudis, C. & Gerald, W. L. An estrogen receptor-negative breast cancer subset characterized by a hormonally regulated transcriptional program and response to androgen. *Oncogene* **25,** 3994–4008 (2006).

42. Ray, P. S., Wang, J., Qu, Y., Sim, M. S., Shamonki, J., Bagaria, S. P., Ye, X., Liu, B., Elashoff, D., Hoon, D. S., Walter, M. A., Martens, J. W., Richardson, A. L., Giuliano, A. E. & Cui, X. FOXC1 is a potential prognostic biomarker with functional significance in basal-like breast cancer. *Cancer Res.* **70,** 3870–3876 (2010).

43. Treue, D., Bockmayr, M., Heim, D., Lechner, J., Stenzinger, A., Thomas, B. & Klauschen, F. Integrative proteogenomic analysis identifies mechanisms of resistance against targeted EGFR-inhibitory therapy, **2016**; *in preparation.*

44. Goldhirsch, A., Wood, W. C., Coates, A. S., Gelber, R. D., Thurlimann, B. & Senn, H. J. Strategies for subtypes–dealing with the diversity of breast cancer: highlights of the St. Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2011. *Ann. Oncol.* **22,** 1736–1747 (2011).

45. Burrell, R. A., McGranahan, N., Bartek, J. & Swanton, C. The causes and consequences of genetic heterogeneity in cancer evolution. *Nature* **501,** 338–345 (2013).

46. Bockmayr, M. *Applications of spatial statistics in image analysis of cancer tissue* Part III Essay *(University of Cambridge).*

47. Yuan, Y., Failmezger, H., Rueda, O. M., Ali, H. R., Graf, S., Chin, S. F., Schwarz, R. F., Curtis, C., Dunning, M. J., Bardwell, H., Johnson, N., Doyle, S., Turashvili, G., Provenzano, E., Aparicio, S., Caldas, C. & Markowetz, F. Quantitative image analysis of cellular heterogeneity in breast tumors complements genomic profiling. *Sci Transl Med* **4,** 157ra143 (2012).

48. Denkert, C., Loibl, S., Noske, A., Roller, M., Muller, B. M., Komor, M., Budczies, J., Darb-Esfahani, S., Kronenwett, R., Hanusch, C., von Torne, C., Weichert, W., Engels, K., Solbach, C., Schrader, I., Dietel, M. & von Minckwitz, G. Tumor-associated lymphocytes as an independent predictor of response to neoadjuvant chemotherapy in breast cancer. *J. Clin. Oncol.* **28,** 105–113 (2010).

49. Zhang, Y., Kang, S., Shen, J., He, J., Jiang, L., Wang, W., Guo, Z., Peng, G., Chen, G., He, J. & Liang, W. Prognostic significance of programmed cell death 1 (PD-1) or PD-1 ligand 1 (PD-L1) Expression in epithelial-originated cancer: a meta-analysis. *Medicine (Baltimore)* **94,** e515 (2015).

50. Stenzinger, A., Weichert, W., Lennerz, J. K. & Klauschen, F. Basket Trials: Just the End of the First Quarter. *J. Clin. Oncol.* **33,** 2823–2824 (2015).

51. Klauschen, F., Heim, D. & Stenzinger, A. Histological tumor typing in the age of molecular profiling. *Pathol. Res. Pract.* **211,** 897–900 (2015).

Affirmation in lieu of an oath/
Eidesstattliche Erklärung

„Ich, Michael Bockmayr, versichere an Eides statt durch meine eigenhändige Unterschrift, dass ich die vorgelegte Dissertation mit dem Thema: „Integrative analysis of "omics" data and histopathological features in breast and ovarian cancer" selbstständig und ohne nicht offengelegte Hilfe Dritter verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel genutzt habe. Alle Stellen, die wörtlich oder dem Sinne nach auf Publikationen oder Vorträgen anderer Autoren beruhen, sind als solche in korrekter Zitierung (siehe „Uniform Requirements for Manuscripts (URM)" des ICMJE -www.icmje.org) kenntlich gemacht. Die Abschnitte zu Methodik (insbesondere praktische Arbeiten, Laborbestimmungen, statistische Aufarbeitung) und Resultaten (insbesondere Abbildungen, Graphiken und Tabellen) entsprechen den URM (s.o) und werden von mir verantwortet. Meine Anteile an den ausgewählten Publikationen entsprechen denen, die in der untenstehenden gemeinsamen Erklärung mit dem Betreuer, angegeben sind. Sämtliche Publikationen, die aus dieser Dissertation hervorgegangen sind und bei denen ich Autor bin, entsprechen den URM (s.o) und werden von mir verantwortet. Die Bedeutung dieser eidesstattlichen Versicherung und die strafrechtlichen Folgen einer unwahren eidesstattlichen Versicherung (§156, 161 des Strafgesetzbuches) sind mir bekannt und bewusst."

____

M. Bockmayr

Declaration of own contributions

Michael Bockmayr contributed to the publications as follows:

Publication 1

M. Bockmayr, F. Klauschen, B. Györffy, C. Denkert and J. Budczies: **New network topology approaches reveal differential correlation patterns in breast cancer** *BMC Syst Biol.* 2013 Aug 15;7(1):78; IF 2.435.
Contribution: 60 %
Details: contributions to the design of the study, development of the algorithms, implementation of the algorithms in R, complete analysis of data, preparation of all the figures, writing of the methods and the results section, contributions to the other sections.

Publication 2

J. Budczies*, M. Bockmayr*, C. Denkert, F. Klauschen, J.K. Lennerz, B. Györffy, M. Dietel, S. Loibl, W. Weichert and A. Stenzinger: **Classical pathology and mutational load of breast cancer – integration of two worlds** *J Path: Clin Res* 2015 Oct;1(4):225–238; [1] (* J. Budczies and M. Bockmayr contributed equally to this work.)
Contribution: 30 %
Details: contributions to the design of the study, contributions to data analysis, preparation of all the figures, contributions to the manuscript.

Publication 3

S. Darb-Esfahani*, C.A. Kunze*, H. Kulbe, J. Sehouli, S. Wienert, J. Lindner, J. Budczies, M. Bockmayr, M. Dietel, C. Denkert, I. Braicu and K. Jöhrens: **Prognostic impact of programmed cell death-1 (PD-1) and PD-ligand 1 (PD-L1) expression in cancer cells and tumor-infiltrating lymphocytes in ovarian high-grade serous carcinoma** *Oncotarget* 2015 Nov;7(2):1486-1499; IF 6.359 (* S. Darb-Esfahani and C.A. Kunze contributed equally to this work.)
Contribution: 10 %
Details: analysis and visualization of TCGA data, contributions to the manuscript.

M. Bockmayr PD Dr. J. Budczies

[1] *The Journal of Pathology: Clinical Research* is a sister journal to the *The Journal of Pathology*, IF 7.43, launched in 2014 with a more clinical focus, it has not yet been assigned an impact factor.

M. Bockmayr, F. Klauschen, B. Györffy, C. Denkert and J. Budczies

New network topology approaches reveal differential correlation patterns in breast cancer

BMC Syst Biol. 2013 Aug 15;7(1):78.

http://dx.doi.org/10.1186/1752-0509-7-78

Pages 19 - 32

J. Budczies*, M. Bockmayr*, C. Denkert, F. Klauschen, J.K. Lennerz, B. Györffy, M. Dietel, S. Loibl, W. Weichert and A. Stenzinger

Classical pathology and mutational load of breast cancer – integration of two worlds

J Path: Clin Res 2015 Oct;1(4):225–238

(* J. Budczies and M. Bockmayr contributed equally to this work.)

http://dx.doi.org/10.1002/cjp2.25

Pages 33 - 46

S. Darb-Esfahani*, C.A. Kunze*, H. Kulbe, J. Sehouli, S. Wienert, J. Lindner, J. Budczies, M. Bockmayr, M. Dietel, C. Denkert, I. Braicu and K. Jöhrens

Prognostic impact of programmed cell death-1 (PD-1) and PD-ligand 1 (PD-L1) expression in cancer cells and tumor-infiltrating lymphocytes in ovarian high-grade serous carcinoma

Oncotarget 2015 Nov;7(2):1486-1499

(* S. Darb-Esfahani and C.A. Kunze contributed equally to this work.)

http://dx.doi.org/10.18632/oncotarget.6429

Pages 47 - 60

Curriculum vitae

For privacy reasons, the curriculum vitae is not published in the online version.

Complete publication list

Peer reviewed articles

1) M. Bockmayr, F. Klauschen, B. Györffy, C. Denkert und J. Budczies: **New network topology approaches reveal differential correlation patterns in breast cancer** *BMC Syst Biol.* 2013 Aug 15;7(1):78. `http://dx.doi.org/10.1186/1752-0509-7-78` IF: 2.435

2) J. Budczies*, M. Winterfeld*, F. Klauschen, M. Bockmayr, J.K. Lennerz, C. Denkert, T. Wolf, A. Warth, M. Dietel, I. Anagnostopoulos, W. Weichert, D. Wittschieber und A. Stenzinger: **The landscape of metastatic progression patterns across major human cancers** *Oncotarget* 2015 Jan;6(1):570:583. `http://dx.doi.org/10.18632/oncotarget.2677` IF:6.359

2) J. Budczies, M. Bockmayr, D. Treue, F. Klauschen und C. Denkert: **Semiconductor sequencing: how many flows do you need?** *Bioinformatics* 2015 Apr;31(8):1199:1203. `http://dx.doi.org/10.1093/bioinformatics/btu805` IF:4.981

4) J. Budczies*, M. Bockmayr*, C. Denkert, F. Klauschen, J.K. Lennerz, B. Györffy, M. Dietel, S. Loibl, W. Weichert und A. Stenzinger: **Classical pathology and mutational load of breast cancer – integration of two worlds** *J Path: Clin Res* 2015 Oct;1(4):225–238. `http://dx.doi.org/10.1002/cjp2.25`

5) S. Darb-Esfahani*, C.A. Kunze*, H. Kulbe, J. Sehouli, S. Wienert, J. Lindner, J. Budczies, M. Bockmayr, M. Dietel, C. Denkert, I. Braicu, K. Jöhrens: **Prognostic impact of programmed cell death-1 (PD-1) and PD-ligand 1 (PD-L1) expression in cancer cells and tumor-infiltrating lymphocytes in ovarian high-grade serous carcinoma.** *Oncotarget* 2015 Nov;7(2):1486-1499. `http://dx.doi.org/10.18632/oncotarget.6429` IF:6.359

6) N. Pfarr, HP Sinn, F. Klauschen, C. Flechtenmacher, M. Bockmayr, K. Ridinger, M. von Winterfeld, A. Warth, K. Lorenz, J. Budczies, R. Penzel, J.K. Lennerz, V. Endris, W. Weichert, A. Stenzinger: **Mutations in genes encoding PI3K-AKT and MAPK signaling define anogenital papillary hidradenoma.** *Genes Chromosomes Cancer* 2016 Feb;55(2):113-9. `http://dx.doi.org/10.1002/gcc.22315` IF:4.041

7) J. Budczies, M. Bockmayr, C. Denkert, F. Klauschen, S. Gröschel, S. Darb-Esfahani, N. Pfarr, J. Leichsenring, M. L. Onozato, J. K. Lennerz, M. Dietel, S. Fröhling, P. Schirmacher, A. J. Iafrate, W. Weichert and A. Stenzinger **Pan-cancer analysis of copy number changes in programmed death-ligand 1 (PD-L1, CD274) - associations with gene expression, mutational load and survival.** *Genes Chromosomes Cancer* 2016 *accepted* `http://dx.doi.org/10.1002/gcc.22365` IF:4.041

(* equal contribution)

Theses

8) M. Bockmayr: **Analysis of the bifurcation structure in a physiologically realistic but reduced mathematical model of cortical spreading depression** Bachelor thesis Mathematics, Freie Universität Berlin, 2012

9) M. Bockmayr: **Applications of spatial statistics in image analysis of cancer tissue** Part III Essay Mathematics, University of Cambridge, 2013

Abstracts

10) M. Bockmayr, F. Klauschen, C. Denkert, J. Budczies: **Identification of differential co-expression patterns in breast cancer.** *Annual meeting of the German Society of Pathology, Der Pathologe [Suppl 1]* 2012; 33:177–180 (Talk)

11) N. Hübel, M. Bockmayr, E. Schoell and M. A. Dahlem: **Phase Space Analysis of a Ionic Model for Cortical Spreading Depression.** *Front. Comput. Neurosci. Conference Abstract: Bernstein Conference* 2012 (Poster)

12) J. Budczies, M. Bockmayr, F. Klauschen, C. Denkert: **New network topology approaches reveal differential correlation patterns in breast cancer.** *German Conference on Bioinformatics* 2012 (Poster)

13) J. Budczies, M. Bockmayr, F. Klauschen, W. Weichert, C. Denkert, A. Stenzinger: **Pan-cancer analysis of programmed death-ligand 1 (PD-L1) amplifications.** *AACR Annual Meeting* 2016 (Talk)

Acknowledgment

I was most fortunate to have PD Dr. Jan Budczies and PD Dr. Frederick Klauschen as my advisors. They provided me with the opportunity to complete this thesis in the interdisciplinary field of mathematics and medicine and introduced me to many exciting topics in pathology and bioinformatics. I am very thankful for the fruitful and inspiring discussions, as well as all their help and encouragement over the last years. I would also like to thank Dr. Albrecht Stenzinger for the very pleasant scientific cooperation. The second part of this thesis is based on joint work with him. Many thanks go to Sarah for reading the manuscript and her helpful comments. Finally, I am very grateful to my family and my girlfriend Bernadette for their constant support.