

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

DISSERTATION

Untersuchungen zur immunhistologischen Diagnostik und Prognose des
Magenkarzinoms

zur Erlangung des akademischen Grades
Doctor rerum medicinalium (Dr. rer. medic.)

vorgelegt der Medizinischen Fakultät
Charité – Universitätsmedizin Berlin

von

Hans-Michael Behrens
aus Northeim

Datum der Promotion: 05.06.2016

Diese Arbeit widme ich meinen Eltern.

Inhaltsverzeichnis

Zusammenfassung.....	1
Abstrakt	1
Abstract.....	2
Einführung.....	3
Zielstellung.....	5
Methodik.....	6
Probensammelfehler bei Magenkarzinom-Biopsien.....	6
Reproduzierbarkeit des Her2/neu-Status und der 10%-Grenze	7
Screening weiterer Biomarker	8
Ergebnisse.....	9
Probensammelfehler bei Magenkarzinom-Biopsien.....	9
Reproduzierbarkeit des Her2/neu-Status und der 10%-Grenze	10
Screening weiterer Biomarker	12
Diskussion	15
Probensammelfehler bei Magenkarzinom-Biopsien.....	15
Reproduzierbarkeit des Her2/neu-Status und der 10%-Grenze	16
Screening weiterer Biomarker	18
Literaturverzeichnis	20
Eidesstattliche Versicherung.....	21
Druckexemplare der ausgewählten Publikationen	23
Lebenslauf.....	54
Komplette Publikationsliste	55
Danksagung.....	57

Zusammenfassung

Abstrakt

In drei Arbeiten wurden Aspekte der immunhistologischen Diagnostik am Magenkarzinom im Hinblick auf Therapiesteuerung und Prognose an 482 Fällen untersucht. Der Nachweis der Her2/neu-Positivität an Biopsien oder Resektaten ist Voraussetzung für die Anwendung des Antikörpermedikamentes Trastuzumab. Zwei den Test behindernde Faktoren wurden untersucht: die erste Arbeit untersuchte das durch die Heterogenität der Her2/neu-Expression beim Magenkarzinom entstehende Risiko von Probensammelfehlern bei bioptischer Probenentnahme. Zwei Pathologen bewerteten den Her2/neu-Status an Resektaten und an daraus entnommenen, in Tissue Microarrays (TMAs) überführten Stanzzyllindern in Biopsiengröße. Sie bewerteten 8,1% bzw. 8,4% der Fälle anhand der Resektate als Her2/neu-positiv, aber nur 6,3% anhand der Proben in Biopsiengröße. Es wären 24% der Her2/neu-positiven Patienten bei alleiniger Untersuchung von Biopsien nicht erkannt worden. Die zweite Arbeit untersuchte die Reproduzierbarkeit der immunhistochemischen Färbeintensität und des immungefärbten Tumoranteils, der bei Unterschreitung von 10% den Fall als negativ abwertet. 10 Pathologen bewerteten 12 Resektate mit drei Methoden. Die Übereinstimmung der Färbeintensitäten war moderat und die Konkordanz der gefärbten Tumoranteile 0,55 bis 0,81. Virtuelle Mikroskopie brachte keine Verbesserung. Die Pathologen konnten die Anteile klar umgrenzter Flächen visuell gut abschätzen, aber Gradienten in der Färbeintensität behinderten die Reproduzierbarkeit. Der Grenzwert von 10% führt zu Fehlinterpretationen und falschen Befunden. Die dritte Arbeit untersuchte 8 phänotypische (Mucin 1, Mucin 2, Mucin 5, Mucin 6, CD10, E-Cadherin, β -Catenin, Lysozym) und 5 genotypische Biomarker (BRAF, KRAS, NRAS, PIK3CA, Mikrosatelliteninstabilität) auf ihre prognostische Relevanz und ihre Eignung zur Identifikation von Subtypen oder zur Steuerung der Therapie. Der Laurén-Typ korrelierte mit 10 klinischen Parametern und mit 6 Biomarkern. Der Mucin-Phänotyp und die anderen phänotypischen Marker eigneten sich nicht für die Klassifikation von Subtypen. Genotypisch war nur die Mikrosatelliteninstabilität mit dem Laurén-Typ korreliert (häufiger im intestinalen und unklassifizierten Typ). Das Überleben korrelierte mit einigen klinischen Parametern und Biomarkern. Im proximalen Magenkarzinom war bei KRAS-Mutation das Überleben signifikant kürzer. Multivariate Überlebensanalyse zeigte Mucin 2 als unabhängigen prognostischen Faktor, und PIK3CA beim intestinalen Typ. Obwohl nicht für die histologische Klassifikation geeignet, hingen doch einige Biomarker mit dem Phänotyp oder der Prognose zusammen und könnten zukünftig für die Steuerung der Therapie in Frage kommen, z.B. KRAS, PIK3CA, MSI und der H. pylori-Status.

Abstract

In three studies, aspects of immunohistological diagnostics of gastric cancer were examined with respect to tailoring therapy and patient prognosis on a cohort of 482 cases.

Detection of Her2/neu positivity upon biopsies or resection specimens is required for application of the antibody therapeutic Trastuzumab. Two factors hindering the testing method were examined: the first study assessed the risk of sampling errors when taking biopsies that is raised by the heterogeneous expression of Her2/neu in gastric cancer. Two pathologists determined the Her2/neu status of resection specimens and of corresponding tissue micro array core cylinders taken from the same specimens. They rated 8.1% and 8.4% of the resection specimens as Her2/neu positive, respectively, but only 6.3% of the corresponding biopsy-sized samples. Twenty-four percent of the patients eligible for Trastuzumab would have been missed if solely biopsies had been examined. The second study examined the reproducibility of immunohistochemical staining intensities and the ratio of stained to total tumor tissue. Her2/neu status is downgraded to negative if the stained percentage of tumor tissue is below 10%. Ten pathologists rated 12 resections specimens using three different methods. Agreement on staining intensities was moderate, and concordance of stained tumor ratio was 0.55 to 0.81. Application of virtual microscopy did not improve the agreement. Pathologists are very well able to visually assess the ratio of clearly demarcated areas, but gradients in Her2/neu staining intensity hinder the reproducibility. The 10% cutoff value leads to misinterpretation and therefore to wrong test results. The third study examined 8 phenotypic biomarkers (mucin 1, mucin 2, mucin 5, mucin 6, CD10, E-cadherin, β -catenin, and lysozyme) and 5 genotypic biomarkers (BRAF, KRAS, NRAS, PIK3CA, and microsatellite instability) whether they can be used to identify subtypes of gastric carcinoma, predict prognosis, or possibly tailor therapy. The histological Laurén type correlated with 10 clinical parameters and 6 biomarkers. The mucin phenotype and the other phenotypic markers were not suited to identify subtypes. Within the genotypic markers, only microsatellite instability correlated with Laurén phenotype, being more common in the intestinal and unclassified types. Survival correlated with several clinical parameters and biomarkers. KRAS mutation showed a significantly longer survival in the proximal subtype. In multivariate survival analysis mucin 2 and PIK3CA (in the intestinal type) were independent prognostic factors. Although not suitable for histological classification, some biomarkers correlated with phenotype or prognosis and might be considered to tailor therapy, like KRAS, PIK3CA, MSI and *H. pylori* status.

Einführung

Das Magenkarzinom ist laut dem World Cancer Report 2014 der WHO weltweit die Krebserkrankung mit der fünfthöchsten Inzidenz und der dritthöchsten Mortalität. Als ätiologische Faktoren sind Infektionen mit *Helicobacter pylori* und dem Epstein-Barr-Virus sowie die Ernährungsweise bekannt. Die Klassifikation erfolgt nach dem TNM-System und dem daraus abgeleiteten Tumorstadium, wie auch bei anderen soliden Tumoren, sowie durch den histologischen Typ nach Laurén.

Die therapeutischen Optionen sind begrenzt. Den einzigen kurativen Ansatz bietet bei lokal begrenztem Tumorwachstum die vollständige chirurgische Resektion des Primärtumors in Verbindung mit einer D2-Lymphadenektomie. Das Magenkarzinom wird jedoch meist erst in höheren Stadien diagnostiziert, in denen der Therapieansatz immer palliativ ist und lange nur aus einer Resektion oder Teilresektion des Magens bestand. Einen Fortschritt gab es durch die Einführung einer zusätzlichen perioperativen Chemotherapie, nachdem die MAGIC-Studie¹ dafür einen Überlebensvorteil gezeigt hatte. Dennoch liegt die Fünfjahresüberlebensrate bei nur 27%.

Die Auswahl der Therapie wurde lange nur auf Grund des Tumorstadiums getroffen. Bei vielen Tumorerkrankungen hat man aber beobachtet, dass Patienten im gleichen Tumorstadium verschieden auf die gleiche Chemotherapie ansprechen. Die tumorbiologische Forschung² bietet Erklärungen dafür, und man hat verschiedene genetische Veränderungen und molekulare Phänotypen entdeckt, die Angriffspunkte für zielgerichtete Therapien bieten und das Ansprechen und die Prognose vorhersagen können. Daher werden zunehmend solche prädiktiven oder prognostischen Biomarker zur Steuerung von zielgerichteten Therapien eingesetzt.

Ein solches zielgerichtetes Therapeutikum ist Trastuzumab, ein gegen den Her2/neu-Rezeptor gerichteter monoklonaler Antikörper. Her2/neu ist ein epidermaler Wachstumsfaktor, über den die Zellteilung angeregt und die Apoptose gehemmt wird, und der in manchen Tumoren überexprimiert ist. Trastuzumab blockiert den Rezeptor und soll so das Tumorwachstum reduzieren. Ursprünglich für die Anwendung am Mammakarzinom entwickelt, ist Trastuzumab inzwischen auch für die palliative Anwendung beim fortgeschrittenen Magenkarzinom zugelassen, nachdem die ToGA-Studie³ einen Überlebensvorteil von 2,7 Monaten bei zusätzlicher Gabe von Trastuzumab zur konventionellen Chemotherapie zeigte.

Die Indikation für Trastuzumab ist an den vorherigen Nachweis der Her2/neu-Positivität des Tumors gebunden. Das begleitende Testverfahren ("Companion Diagnostics") für die

Anwendung beim Magenkarzinom sieht eine Kombination aus Immunhistochemie und *in situ*-Hybridisierung vor⁴. Anhand von Schnittpräparaten von Biopsien oder Resektaten beurteilt ein Pathologe die Expression durch die Intensität der membranständigen Immunfärbung eines anti-Her2/neu-Antikörpers mit einem Score von 0 (keine Färbung) bis 3+ (starke zirkumferentielle, laterale oder basolaterale Färbung). Bei einem Score von 0 oder 1+ gilt der Tumor als Her2/neu-negativ, bei 3+ als Her2/neu-positiv. Ein Score von 2+ gilt als unentschieden, worauf nachfolgend eine *in situ*-Hybridisierung durchgeführt wird. Nur wenn dabei eine HER2-Genamplifikation gefunden wird, gilt der Tumor als Her2/neu-positiv. Es kommt allerdings noch ein Grenzwert für den Anteil der positiven Tumorzellen am gesamten Tumorgewebe hinzu, bei dessen Unterschreitung der Tumor dennoch als Her2/neu-negativ gilt. Dieser Grenzwert ist verschieden für Biopsien (mindestens fünf zusammenhängende Zellen müssen positiv sein) und Resektate (mindestens 10% des Tumorgewebes muss positiv sein).

Die Her2/neu-Testung am Magenkarzinom ist aber mit Problemen behaftet: die Expression von Her2/neu ist im Magenkarzinom häufig heterogen verteilt. Dies kann bei der Entnahme von Biopsien zu einem Stichprobenfehler führen, wenn das Her2/neu-positive Tumorgewebe in den Biopsien verfehlt wird. Besonders bei Patienten mit fortgeschrittenem, inoperablem Tumor, für die ja gerade Trastuzumab eine Option wäre, stehen oft nur Biopsien zur Her2/neu-Testung zur Verfügung. Bei einem Anteil des Her2/neu-gefärbten Tumors in der Nähe des Grenzwertes von 10% ist das Testresultat abhängig von der visuellen Einschätzung des Pathologen, was weniger oder mehr als 10% ist. Beide Fehlerquellen wurden in dieser Arbeit systematisch untersucht.

Trastuzumab ist bisher die einzige zielgerichtete Therapie für das Magenkarzinom, bietet aber keinen kurativen Ansatz und nur einen relativ kleinen Überlebensvorteil. Im Vergleich zum kolorektalen Karzinom oder Lungenkarzinom steht die genotypische und phänotypische Charakterisierung des Magenkarzinoms noch in den Anfängen, wodurch vermutlich viele Patienten eine Therapie mit suboptimaler oder sogar fehlender Wirksamkeit bekommen. Es werden also dringend weitere Erkenntnisse benötigt.

Zielstellung

Bei der endoskopischen Entnahme von Biopsien aus der Magenschleimhaut wird nur ein Bruchteil des gesamten Tumors durch punktuelles Sammeln von Proben abgetastet. In der ersten Arbeit wurde dazu die Fragestellung untersucht, wie verlässlich die Testung des Her2/neu-Status an Gewebeproben in der Größe von Biopsien ist im Vergleich zu den Testresultaten, die anhand eines zugehörigen Resektates bestimmt werden. Die Biopsien wurden durch Stanzzyylinder repräsentiert, die aus in Paraffin eingebetteten Resektaten entnommen wurden. Ferner wurde untersucht, ob der Her2/neu-Status bei nur chirurgisch behandelten Patienten einen Einfluss auf das Überleben hat.

Die Fragestellung der zweiten Arbeit war die Reproduzierbarkeit der visuellen Bewertung der Färbeintensität und des Anteils an positivem Tumor an Her2/neu-immungefärbten Schnittpräparaten von Magenkarzinom-Resektaten durch mehrere Pathologen. Es wurden die Auswirkungen des Grenzwertes von 10% auf die Reproduzierbarkeit des Her2/neu-Status untersucht, wenn der Anteil des positiven Tumors in der Nähe dieses Grenzwertes liegt. Es sollte getestet werden, ob gegenüber der Befundung mit einem konventionellen Mikroskop der Einsatz virtueller Mikroskopie oder virtueller Mikroskopie mit einer zusätzlichen Assistenzmethode zum expliziten Einzeichnen von Tumorflächen die Reproduzierbarkeit verbessern kann. Das manuelle Einzeichnen der Tumorareale und des positiv gefärbten Anteils in die virtuellen Präparate sollte gleichzeitig ein Einblick in die visuelle Wahrnehmung der Pathologen geben, um Ursachen für die Diskrepanzen zwischen den Bewertungen zu identifizieren.

Ziel der dritten Arbeit war eine umfassende Analyse von phänotypischen und genotypischen Biomarkern an Magenkarzinomen. Für die Untersuchung wurden Biomarker ausgewählt, die sich bei anderen Tumorerkrankungen als prädiktiv oder prognostisch relevant erwiesen hatten oder für die schon eine zielgerichtete Therapie existiert. Es sollten die Fragestellungen beantwortet werden, ob sich durch die Biomarker Subtypen des Magenkarzinoms identifizieren lassen, ob sie eine prognostische Bedeutung haben, und ob irgendeiner dieser Biomarker in Erwägung gezogen werden könnte, um zukünftig die Therapie des Magenkarzinoms zu lenken.

Methodik

Als Grundlage für die drei Arbeiten wurde aus dem Archiv des Instituts für Pathologie der Universitätsklinik Kiel eine Kohorte von 483 Adenokarzinomen des Magens und des gastroösophagealen Übergangs aus den Jahren 1997 bis 2009 zusammengestellt. Eingeschlossen wurden Fälle mit totaler oder subtotaler Gastrektomie und Vorhandensein von formalinfixierten, paraffineingebetteten Resektaten. Ausgeschlossen wurden Fälle mit perioperativer Chemotherapie. Aus dem elektronischen Archivsystem wurden Geschlecht, Diagnosedatum, Alter bei Diagnose, Lokalisation, histologischer Typ nach Laurén, histologischer Differenzierungsgrad sowie T, N, M, Tumorstadium nach UICC und Anzahlen der resezierten und der positiven Lymphknoten ermittelt. Follow-Up- und Sterbedaten wurden vom Epidemiologischen Landeskrebsregister Schleswig-Holstein, dem klinischen Informationssystem und durch Recherche bei niedergelassenen Ärzten abgefragt. Aus T, N und M wurde auch eine alternative Stadieneinteilung codiert ("Kiel-Stadium"), die in einer separaten Arbeit an dieser Kohorte erarbeitet wurde⁵ und die Überlebenskurven besser trennte als die UICC-Klassifikation.

Von den Resektaten wurden Schnittpräparate für die weiteren Untersuchungen angefertigt. Anhand eines mit Hämatoxylin-Eosin (HE) gefärbten Schnittpräparates wurde jeweils die Position des Tumorgewebes bestimmt, aus dem Tumorgewebe des entsprechenden Präparates Stanzzyylinder mit 1,5 mm Durchmesser entnommen und diese zu Tissue Microarrays (TMAs) kombiniert (je ein Satz TMAs mit 3 und mit 5 Stanzen pro Präparat). In 454 Fällen lag genügend Probenmaterial für die Entnahme der Stanzen und Einbringung in die TMAs vor.

Probensammelfehler bei Magenkarzinom-Biopsien

Mit einem anti-Her2/neu-Antikörper wurden pro Fall ein Schnittpräparat des Resektates sowie Schnittpräparate der TMAs immunhistochemisch gefärbt. Die Stanzzyylinder in den TMAs repräsentierten pro Fall fünf Tumorgewebe enthaltende Gewebeproben in der Größe von Biopsien.

An den Her2/neu-Immunfärbungen der Resektate und der TMAs wurde unabhängig von zwei Pathologen nach dem Auswerteschema für das Magenkarzinom⁴ der immunhistologische Score bewertet. Unter Berücksichtigung der Grenzwerte von mindestens 10% positivem Tumoranteil bei Resektaten bzw. 5 zusammenhängenden gefärbten Tumorzellen bei Biopsien wurde daraus der Her2/neu-Status bestimmt, gegebenenfalls nach der bei einem Score von 2+ zusätzlichen obligaten in situ-Hybridisierung.

Her2/neu-Status und immunhistologischer Score wurden zwischen den beiden Pathologen mit dem Kappa-Test verglichen, sowie auch zwischen den Resektaten und jeweils zugehörigen

Biopsien. Diskrepante Fälle wurden auf die Ursachen der unterschiedlichen Bewertungen hin untersucht, und es wurde zusammen mit einem dritten Pathologen zu jedem Fall ein Konsensus-Her2/neu-Status bestimmt. Dieser wurde mit Fisher's Exact Test explorativ auf Zusammenhänge mit den klinisch-pathologischen Parametern getestet. Schließlich wurde eine Überlebensanalyse nach Kaplan-Meier bezüglich des Her2/neu-Konsensus-Status durchgeführt und die Signifikanz eines Überlebensunterschiedes mit dem log-rank-Test bestimmt.

Reproduzierbarkeit des Her2/neu-Status und der 10%-Grenze

Zur Untersuchung der zweiten Fragestellung, der Reproduzierbarkeit des Her2/neu-Status zwischen mehreren Pathologen im Hinblick auf Auswirkungen der 10%-Grenze, wurden 12 Her2/neu-immungefärbte Schnittpräparate von Magenkarzinom-Resektaten ausgewählt, darunter jene, bei denen es in der ersten Arbeit schwierig zu entscheiden war, ob der Anteil des immungefärbten Tumors über oder unter 10% lag. Diese Präparate wurden von 10 Pathologen regelmäßig bewertet, aber zusätzlich wurde die Schätzung des Her2/neu-gefärbten Anteils des Tumors abgefragt (0 bis 100%). Der immunhistochemische Score (0 bis 3+) sollte auch dann angegeben werden, wenn es durch Unterschreitung des Grenzwertes von 10% zu einer Abwertung, also einem negativen Her2/neu-Status, kam. Durch diese Informationen sowie den mittels in situ-Hybridisierung bestimmten Her2/neu-Amplifikationsstatus konnten separat die Reproduzierbarkeiten des Scores, des gefärbten Tumoranteils und des Her2/neu-Status bewertet und die Ursachen für Diskrepanzen ermittelt werden.

Die Pathologen beurteilten die Präparate zuerst mit ihrem Mikroskop. In einer zweiten Runde wurde die Hypothese getestet, dass sich der Her2/neu-gefärbte Anteil des Tumorgewebes mit virtueller Mikroskopie besser abschätzen lässt als am Mikroskop. Zugrunde lag die Überlegung, dass im Mikroskop immer nur ein kleiner Teil des Präparates auf einmal sichtbar ist, während bei virtueller Mikroskopie ein Gesamtüberblick möglich ist. Die 12 Präparate wurden in 400-facher Vergrößerung mit einem Leica SCN400 "Whole Slide Scanner" digitalisiert und mit einem Betrachtungsprogramm von den 10 Pathologen erneut am Computerbildschirm bewertet. Das Betrachtungsprogramm wurde von mir selbst in der Sprache Java programmiert und erlaubte das Einlesen des Leica-SCN-Dateiformats, in dem die gescannten, virtuellen Präparate vorlagen, das Anzeigen der Bilder in verschiedenen Vergrößerungsstufen und Ausschnitten (inklusive Gesamtansicht) sowie eine effiziente Navigation über Tastatur und Maus.

In einer dritten Runde sollten die Ursachen der weiterhin bestehenden Diskrepanzen identifiziert werden. Dazu wurden für das Betrachtungsprogramm zusätzliche Funktionen zum manuellen

Einzeichnen von Flächen in beliebiger Vergrößerungsstufe, deren Speicherung in einem eigenen Dateiformat und zur Berechnung von Flächenverhältnissen programmiert. Die Pathologen markierten damit am Bildschirm die Areale, die sie als Tumorgewebe und für angefärbtes Tumorgewebe identifizierten (Abbildung 1A). Abschließend wurde ihnen eine bildschirmfüllende Skizze der markierten Areale gezeigt (Abbildung 1B), anhand der sie den Her2/neu-exprimierenden Tumoranteil visuell abschätzten. Auf diese Weise konnte die Fähigkeit zum Schätzen eines Flächenanteils unter Ausschluss anderer Einflussfaktoren (inhomogene Verteilung oder Intensitätsverläufe) bewertet werden.

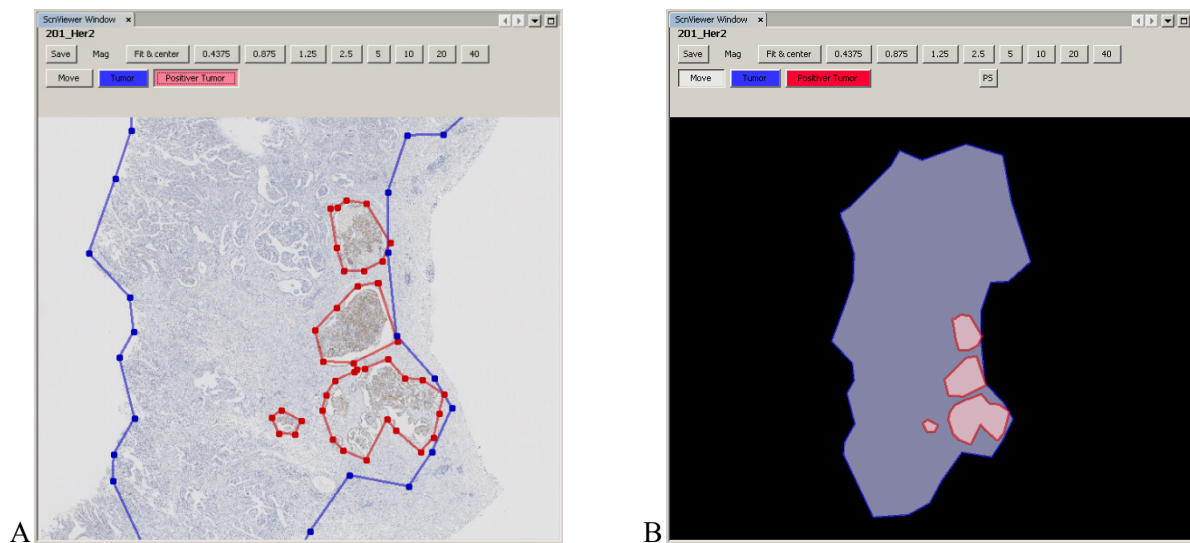


Abbildung 1. A: Funktion zum Markieren der Areale von Tumor (blau) und immunhistochemisch gefärbtem Tumor (rot). B: Finale Ansicht zum visuellen Schätzen des immungefärbten Anteils.

Screening weiterer Biomarker

Für das explorative Screening auf prädiktive oder prognostische Biomarker wurden als phänotypische Kandidaten Mucin 1, Mucin 2, Mucin 5, Mucin 6, CD10, E-Cadherin, β -Catenin und Lysozym ausgewählt. Getestet wurde durch immunhistochemische Färbung der TMAs (3 Stanz pro Fall) mit Antikörpern gegen diese Proteine. Zur Auswertung wurde ein additiver Score von 0 bis 8 aus den zwei Komponenten Färbeintensität (0 bis 3) und Anteil gefärbter Tumorzellen (0 bis 5) gebildet. Die Expression wurde separat für jeden Marker am Median der Scores in negativ und positiv getrennt. Als genotypische Marker wurden die Gene *KRAS*, *NRAS*, *BRAF* und *PIK3CA* (auf Exon 9 und auf Exon 20) ausgewählt. Ihr Mutationsstatus wurde durch PCR und Pyrosequenzierung von Tumorgewebe bestimmt, das durch manuelle Mikrodisektion aus Schnittpräparaten der Resektate gewonnen wurde. Auf Mikrosatelliteninstabilität wurde phänotypisch durch Immunfärbung der Mismatch-Repair-Proteine MLH1, PMS2, MSH2 und MSH6 getestet wie auch genotypisch durch Sequenzierung der fünf Mikrosatelliten BAT-25,

BAT-26, NR-21, NR-22 und NR-27. Wenn mindestens zwei der fünf Mikrosatelliten verschiedene Sequenzlängen zwischen Tumorgewebe und Normalgewebe zeigten, wurde der Fall als mikrosatelliteninstabil (MSI-H) beurteilt. Schließlich wurde auf eine Infektion mit *Helicobacter pylori* histologisch und auch durch Sequenzierung getestet.

Zwischen den klinisch-pathologischen Parametern, den phänotypischen Markern und den genotypischen Markern wurde mit Fisher's Exact Test und Kendall's Tau Test explorativ nach Zusammenhängen gesucht. Für jeden Parameter wurde einzeln eine univariate Überlebensanalyse durchgeführt (Kaplan-Meier-Plot und Log-Rank-Test). Alle Parameter, die der univariaten Überlebensanalyse einen p-Wert unter 0,10 hatten, wurden mit einer multivariaten Überlebensanalyse (Cox-Regression und *Backward-LR-Verfahren*) auf Unabhängigkeit getestet. Alle Analysen wurden für die Untergruppen der intestinalen und diffusen Typen nach Laurén sowie der proximalen und distalen Lokalisationen wiederholt. Wegen der Anzahl der durchgeführten statistischen Tests wurde eine Korrektur des Signifikanzniveaus nach Siemes vorgenommen.

Ergebnisse

Probensammelfehler bei Magenkarzinom-Biopsien

Die beiden Pathologen bewerteten unabhängig 8,1% bzw. 8,4% der Resektate als Her2/neu-positiv, und dies mit sehr guter Übereinstimmung (Kappa 0,985). Dagegen bewerteten sie anhand der zu denselben Fällen gehörenden TMA-Stanzen nur 6,3% der Fälle als positiv, aber auch dies mit sehr guter Übereinstimmung (Kappa 0,962). Diese Diskrepanz zeigt sich auch in deutlichen niedrigeren Kappa-Werten, wenn separat für jeden Pathologen der Her2/neu-Status zwischen Resektat und dazugehörenden TMA verglichen werden (Kappa 0,792 bzw. 0,812). Beim Vergleich der immunhistochemischen Scores (0 bis 3+) zwischen Resektat und TMA-Stanzen zeigen sich noch niedrigere Kappa-Werte (0,473 bzw. 0,448).

Insgesamt wurden 12 Fälle diskrepant zwischen Resektat und TMA bewertet, davon 9 falsch negativ und 3 falsch positiv (jeweils in Bezug auf das Ergebnis des jeweiligen Resektates). Zur Aufklärung der Gründe für die schlechte Übereinstimmung wurden die entsprechenden Resektate und TMAs-Stanzen fallweise mikroskopisch miteinander verglichen. Dabei wurden mehrere Fehlerursachen festgestellt: 6 falsch negative Befunde waren darauf zurückzuführen, dass das Resektat zwar einen Her2/neu-positiven Tumorklon enthielt, der aber nur einen kleinen

Teil der Fläche ausfüllte, meist nur 10-20% (maximal 40%). Der positive Tumorzellklon war jeweils scharf begrenzt, es gab keine graduelle Abnahme der Intensität hin zum ungefärbten Tumorareal, und die Stenzen wurden zufällig im ungefärbten Gebiet entnommen. Die übrigen drei falsch negativen Befunde waren in einer graduellen Verteilung der Färbeintensität begründet. Der Tumor zeigte schwächer und stärker exprimierte Abschnitte, und die Stenzen waren zufällig in einem Abschnitt mit so schwacher Färbung entnommen worden, dass die Befundung negativ ausfiel. Zwei falsch positive Befunde entstanden dadurch, dass die TMA-Stenzen zwar eindeutig positiv zu bewerten waren (mindestens 5 zusammenhängende positive Zellen), aber im Resektat erkennbar war, dass der Tumor weniger als 10% positiven Anteil hatte. Demnach galt der Tumor als negativ, aber die Stenzen waren zufällig im positiven Bereich genommen worden.

Es wurde auch eine Ursache für Abweichungen zwischen den Beobachtern identifiziert: bei Fällen mit einem positiv gefärbten Tumoranteil im Bereich um 10% gab es verschiedene Einschätzungen, ob der Anteil kleiner oder größer als 10% war, also ob demnach eine Abwertung, resultierend in einem negativem Befund, hätte vorgenommen werden müssen.

Der schließlich von drei Pathologen gemeinsam bestimmte Konsensus-Her2/neu-Status wurde mit den klinisch-pathologischen Parametern korreliert. Es ergab sich ein Zusammenhang mit dem Tumortyp ($p=0.002$) und dem Grading ($p=0.003$). Außerdem hatten Tumoren des proximalen Magens eine höhere Positivrate (11,8%) als solche des distalen Magens (6,5%). In der Untergruppe der proximalen Magenkarzinome gab es einen Zusammenhang mit dem Nodalstatus und bei den Distalen mit dem Grading. Die Überlebensanalyse ergab keinen statistisch signifikanten Unterschied der Überlebenskurven von Her2/neu-positiven und –negativen Patienten.

Reproduzierbarkeit des Her2/neu-Status und der 10%-Grenze

Die Bewertungen des immunhistochemisch angefärbten Tumoranteils (0 bis 100%) und des immunhistochemischen Scores (0 bis 3+) durch 10 Pathologen an 12 Präparaten zeigten Schwankungen beider Parameter. Zur Quantifizierung der Übereinstimmung der jeweils 120 Messwerte wurden statistische Tests ausgewählt, die den Vergleich mehrerer Beobachter (Pathologen) an mehreren Subjekten (Präparaten) erlauben. Für den angefärbten Tumoranteil war das der Overall Concordance Correlation Coefficient (OCCC), und für die immunhistochemischen Scores Fleiss' Kappa. Beide Tests liefern einen Zahlenwert zwischen 0 (keine Übereinstimmung) und 1 (vollständige Übereinstimmung).

Bei der Bewertung mit dem normalen Mikroskop ergab sich eine Konkordanz (OCCC) von 0,682 für den angefärbten Tumoranteil und Fleiss' Kappa von 0,44 für die immunhistochemischen Scores. Es gab bei vier Präparaten diskordante Einschätzungen, ob der angefärbte Tumoranteil über oder unter 10% lag. Die Anwendung von virtueller Mikroskopie an Stelle der normalen Mikroskope brachte keine Verbesserung der Übereinstimmung. Die Konkordanz des angefärbten Tumoranteils betrug nun 0,551 und Fleiss' Kappa für die Übereinstimmung der immunhistochemischen Scores 0,45.

Das manuelle Einzeichnen der Umrisse der als Tumor und als gefärbter Tumor eingeschätzten Bereiche im virtuellen Präparat zeigte, dass die Pathologen bei der Festlegung der Grenzen des mit einer bestimmten Intensität (z.B. 2+ oder 3+) gefärbten Tumoreals stärker voneinander abwichen (Konkordanz 0,704) als bei der Festlegung der Grenzen des gesamten Tumorgewebes (Konkordanz 0,807). Die Gegenüberstellung der visuellen Einschätzungen des gefärbten Tumoranteils zwischen den Methoden zeigte eine starke Streuung (Pearsons Korrelationskoeffizienten von 0,634 bis 0,647). Dagegen ergibt sich eine bemerkenswert gute Übereinstimmung (Pearsons Korrelationskoeffizient: 0,979, $p < 0.001$), wenn man die Einschätzung des gefärbten Tumoranteils anhand der Umriss-Zeichnungen mit den aus diesen Zeichnungen berechneten, tatsächlichen Flächenanteilen vergleicht (wobei allerdings der Flächenanteil systematisch um etwa 3,5% überschätzt wurde). Dies gibt einen Hinweis darauf, dass nicht das Vermögen, Flächenverhältnisse abzuschätzen, sondern ein anderer Faktor die Reproduzierbarkeit begrenzt. Bei Betrachtung der Standardabweichungen der Schätzungen an den einzelnen Präparaten fiel auf, dass es Präparate mit niedriger und mit hoher Standardabweichung gibt, also offenbar "leichter" oder "schwerer" zu befundende. Bei näherer Betrachtung der immunhistochemischen Färbemuster fiel auf, dass die Präparate mit niedriger Standardabweichung eher ein Färbemuster mit scharfen Begrenzungen aufwiesen (ein "schwarz-oder-weiß-Muster"). Die Präparate mit hoher Standardabweichung zeigten dagegen eher ein Färbemuster, das durch einen Intensitätsverlauf der Immunfärbung gekennzeichnet war ("Graustufenmuster"), und dessen gefärbte Areale besonders heterogen verteilt waren.

Durch nachträglichen Vergleich mit den Resultaten der *in situ*-Hybridisierungen konnte ermittelt werden, dass in diesem Versuch bis zu 19% der Her2/neu-Befunde falsch negativ waren, weil die Grenze von 10% als unterschritten beurteilt wurde oder der Score unterschätzt wurde.

Eine Subgruppenanalyse ergab, dass Fachärzte den immunhistochemischen Score besser reproduzierbar bestimmen können als Assistenzärzte, aber bei der prozentualen Schätzung des gefärbten Tumoranteils keine Gruppe einen eindeutigen Vorteil hat.

Screening weiterer Biomarker

Die 482 Patienten umfassende Kohorte war charakterisiert durch ein medianes Alter bei Diagnose von 67,9 Jahren und eine Geschlechterverteilung von 61,6% Männern und 38,4% Frauen. Für 466 Patienten lagen Überlebensdaten vor. Das mediane Gesamtüberleben betrug $15,5 \pm 1,2$ Monate (95%-Konfidenzintervall 13,1-17,8 Monate) bei einem verstorbenen Anteil von 71,9% und einer medianen Nachbeobachtungszeit von 30,8 Monaten.

Nach der Laurén-Klassifikation enthielt die Kohorte 247 (51,2%) Fälle mit intestinalem Typ, 152 (31,3%) mit diffusem, 30 (6,2%) mit gemischtem Typ und 53 (10,9%) unklassifizierte Fälle.

Aus dem Expressionsstatus von Mucin 5, CD10 und Mucin 2 wurde der Mucin-Phänotyp codiert.⁶ Gemäß dieser Klassifikation ergaben sich 123 (29,0%) Fälle vom intestinalen Typ, 63 (14,9%) vom gastrischen und 169 (39,9%) vom gemischten Typ sowie 69 (16,3%) vom unklassifizierten Typ.

Die Mutationsanalyse ergab 17 Fälle (3,6%) mit *KRAS*-Mutation sowie für *PIK3CA* 12 Fälle (2,5%) bzw. 9 Fälle (1,9%) mit Mutationen auf Exon 9 bzw. Exon 20. Mutationen von *BRAF* oder *NRAS* konnten in keinem Fall nachgewiesen werden.

Die Auswertung der Immunfärbungen der Mismatch-Repair-Proteine MSH2, MSH6, MLH1 und PMS2 ergaben 158 Fälle, in denen mindestens eines dieser Proteine vermindert oder gar nicht exprimiert wurde. Von diesen Fällen wurde die Mikrosatelliteninstabilität untersucht, die dann 33 (7,3%) hoch instabile Fälle (MSI-H) ergab.

In 411 Großflächenschnitten war neben Tumorgewebe auch Normalgewebe (Schleimhaut) vorhanden, an der in 62 Fällen (15,1%) das Vorhandensein einer Infektion mit *Helicobacter pylori* histologisch oder durch PCR nachgewiesen werden konnte. EBV-Infektionen konnten durch Nachweis von *EBV*-RNA in 15 Fällen (4,0%) nachgewiesen werden.

Der Phänotyp nach Laurén korrelierte signifikant mit Mucin 1, CD10, E-Cadherin, β -Catenin, und Lysozym. Mikrosatelliteninstabilität wurde signifikant gehäuft im intestinalen (9,1% instabiler Anteil) und unklassifizierten Typ (22,4% instabiler Anteil) gefunden. Der Mucin-

Phänotyp korrelierte mit Lysozym und *PIK3CA* (Exon 9) sowie, nach Korrektur für multiples Testen aber nicht mehr signifikant, mit β -Catenin.

Männer zeigten einen höheren positiven Anteil bei Mucin 1, CD10 und mehr Mutationen an *PIK3CA* (Exon 9, nicht signifikant) als Frauen, und umgekehrt Frauen mehr Mucin 6 (auch nicht signifikant). Ältere Patienten zeigten eine höhere Expression von β -Catenin (nicht signifikant nach Korrektur) und einen höheren Anteil an Mikrosatelliteninstabilität (MSI-H) als jüngere.

Der Differenzierungsgrad (Grading) korrelierte negativ mit CD10, E-Cadherin und β -Catenin, d.h. eine stärkere Entdifferenzierung ging mit einer Abnahme der Positivität der Marker einher.

Der MSI-Status zeigte statistisch signifikante Zusammenhänge mit Lysozym und der Mutation von *KRAS*: die hoch instabilen Fälle (MSI-H) wiesen einen kleineren Anteil an Lysozym-Positivität auf (21,9%) als die stabilen (53,0%). Dagegen zeigten die hoch instabilen Fälle eine höhere Rate an *KRAS*-Mutationen (15,2%) als die stabilen (2,9%). Mit steigendem regionalem Lymphknotenbefall nahm der mikrosatelliteninstabile Anteil ab (11,4% MSI-H bei N0, 7,7% bei N1, 6,7% bei N2 und 4,9% bei N3). Dieser Zusammenhang war aber nach Korrektur für multiples Testen nicht mehr signifikant.

Die univariaten Überlebensanalysen zeigten einen statistisch signifikanten Einfluss des Laurén-Typs, T, N, UICC-Stadium, "Kiel-Stadium", Lymphknotenverhältnis, Grading und Resektionsstatus auf das Überleben. Eine Infektion mit *Helicobacter pylori* führte nur beim unklassifizierten Typ nach Laurén zu einem verringerten Überleben.

Für den MSI-Status und Lysozym ergaben sich zwar p-Werte unter 0,05, die aber durch die Siemes-Korrektur für multiples Testen nicht als signifikant gelten konnten (14,1 Monate medianes Überleben bei MSS gegenüber 35,0 Monaten bei MSI-H, $p=0,035$ sowie 16,7 Monate bei negativer Expression von Lysozym gegenüber 12,5 Monaten bei positiver Expression, $p=0,028$).

In der Untergruppe des intestinalen Typs nach Laurén hatten *KRAS*-mutierte Fälle einen Überlebensnachteil, aber ohne statistische Signifikanz (*KRAS*-Wildtyp: 17,9 Monate, *KRAS* mutiert: 9,6 Monate, $p=0,098$).

In das initiale Modell der multivariaten Überlebensanalyse wurden auch die Parameter mit einem p-Wert $>0,05$, aber $<0,10$ aufgenommen. Dies waren (unterschiedlich je nach Subgruppe): Altersgruppe, Mucin-2, *KRAS*, Lysozym und der Mucin-Phänotyp.

Bei den multivariaten Überlebensanalysen zeigten das Lymphknotenverhältnis (Anzahl der positiven Lymphknoten dividiert durch die Anzahl der resezierten Lymphknoten) und der Resektionsstatus immer das höchste Hazard Ratio (Tabelle 1). Von den untersuchten Biomarkern erwiesen sich als unabhängige prognostische Größen Mucin 2 für die gesamte Kohorte und die proximalen Tumore sowie *PIK3CA* auf Exon 20 und der Mucin-Phänotyp beim intestinalen Typ nach Laurén.

Untergruppe	Parameter mit Signifikanz bei univariater Überlebensanalyse	Unabhängige Parameter nach multivariater Überlebensanalyse	p-Wert	Hazard Ratio	95% Konfidenzintervall
Gesamte Kohorte	Altersgruppe, Laurén, T, N, Lymphknotenverhältnis, Grading, Resektionsstatus, MSI-H, Mucin 2, Lysozym	Altersgruppe	0,001	1,525	1,197-1,944
		T	<0,001	1,028	1,013-1,044
		Lymphknotenverhältnis	<0,001	2,887	2,195-3,796
		Resektionsstatus	<0,001	2,274	1,592-3,248
		Mucin 2	0,004	0,689	0,534-0,887
Intestinaler Typ	Altersgruppe, T, N, Lymphknotenverhältnis, Grading, Resektionsstatus, Mucin-Phänotyp, Mucin 2, CD10, KRAS, PIK3CA (Exon 20)	Altersgruppe	0,009	1,632	1,130-2,357
		Lymphknotenverhältnis	<0,001	3,326	2,237-4,946
		Grading	0,002	1,030	1,011-1,049
		Resektionsstatus	0,013	2,135	1,170-3,895
		Mucin-Phänotyp	0,023	1,196	1,025-1,397
		PIK3CA (Exon 20)	0,037	0,117	0,016-0,876
Diffuser Typ	T, N, Lymphknotenverhältnis, Resektionsstatus, PIK3CA (Exon 20)	T	0,037	1,069	1,034-1,106
		N	<0,001	1,034	1,016-1,053
		Resektionsstatus	0,003	2,286	1,315-3,976
Proximale Tumore	Laurén, T, N, Lymphknotenverhältnis, Grading, Resektionsstatus, Mucin-Phänotyp, Mucin 2, KRAS	N	<0,001	1,032	1,016-1,049
		Resektionsstatus	0,007	2,282	1,258-4,140
		Mucin 2	0,026	0,584	0,364-0,937
Distale Tumore	Altersgruppe, Laurén, T, N, Lymphknotenverhältnis, Grading, Resektionsstatus, Lysozym, MSI-H	Altersgruppe	0,010	1,487	1,099-2,012
		T	0,015	1,025	1,005-1,045
		Lymphknotenverhältnis	<0,001	3,450	2,419-4,920
		Resektionsstatus	<0,001	2,550	1,633-3,981

Tabelle 1. Ergebnisse der multivariaten Überlebensanalyse der gesamten Kohorte und der Untergruppen nach Laurén-Typ und Lokalisation. Es sind die initialen Parameter aufgelistet, die bei der univariaten Analyse relevant für das Überleben waren, sowie die nach schrittweiser Elimination übrigen, unabhängigen Parameter.

Abschließend wurde versucht, Zusammenhänge zwischen den genotypischen und phänotypischen Parametern durch hierarchisches Clustering aufzudecken. Dazu wurde die Statistiksoftware R mit dem Bioconductor-Modul benutzt. Das Resultat des Clustering wurde als Heatmap dargestellt. Es ergaben sich allerdings keine Erkenntnisse, die über die bereits durch direkte Korrelation bekannten hinausgingen.

Diskussion

Probensammelfehler bei Magenkarzinom-Biopsien

Die Charakterisierung nach Phänotyp und Genotyp wird zunehmend zur Steuerung der Krebstherapie eingesetzt. Mit der Einführung von Trastuzumab zur palliativen Behandlung des fortgeschrittenen Magenkarzinoms steht erstmals auch eine zielgerichtete Therapie für das Magenkarzinom zur Verfügung. Trastuzumab kann aber nur eingesetzt werden, wenn zuvor die Überexpression des Her2/neu-Rezeptors im Tumorgewebe nachgewiesen wird. Die Zulassung durch die Medikamentenbehörden (*Food and Drug Administration* und *European Medicines Agency*) verlangt dafür den Nachweis durch ein zertifiziertes Testverfahren. Das auch in dieser Arbeit eingesetzte Verfahren⁴ testet an Schnittpräparaten von Resektaten oder Biopsien zuerst die Expression anhand einer immunhistochemischen Färbung mit einem gegen den Her2/neu-Rezeptor gerichteten monoklonalen Antikörper (Klon 4B5, Roche Diagnostics). Der Tumor gilt als Her2/neu-positiv, wenn mindestens 10% des Tumorgewebes von Resektaten oder mindestens 5 zusammenhängende Zellen bei Biopsien eine starke membranständige Färbung aufweisen (Score 3+), oder bei einer moderaten Färbung (Score 2+) in Verbindung mit zusätzlichem Nachweis der Amplifikation des HER2-Gens durch *in situ*-Hybridisierung.

Anhand der Schnittpräparate von Resektaten wurde in der ersten Arbeit eine Her2/neu-Positivität bei 8,1% der Patienten festgestellt. Die Her2/neu-Positivität war häufiger beim intestinalen Typ nach Laurén (12,9%) als beim diffusen Typ (3,9%) sowie häufiger bei proximaler (11,8%) als bei distaler (6,5%) Lokalisation. Diese Angaben reihen sich in veröffentlichte Beobachtungen ein. So berichtet eine Metaanalyse⁷ von einer Prävalenz zwischen 4% und 53% (Median 18%). Die Diskrepanz der Positivrate zwischen Resektaten und TMAs lässt sich also nicht durch methodische Mängel oder die Auswahl der Kohorte erklären.

Die amerikanischen Leitlinien des *National Comprehensive Cancer Network* und die deutsche S3-Leitlinie "Magenkarzinom" empfehlen die Entnahme von 8 bis 10 Biopsien zur Diagnosesicherung eines Karzinoms (damit mindestens in einer Biopsie auch Tumor enthalten ist), aber nicht zur Bestimmung des HER2-Status. In der Praxis ist die Anzahl der entnommenen Biopsien üblicherweise kleiner als acht, und nicht alle enthalten auch Tumorgewebe. Die Untersuchung von 5 Tumorgewebe enthaltenden Stenzen mit der Größe von Biopsien in dieser Arbeit kann also als repräsentativ angesehen werden.

In bisher veröffentlichten Arbeiten über den Vergleich des Her2/neu-Status zwischen Resektaten und Biopsien wurden die Proben nicht aus demselben Tumormaterial entnommen, sondern

teilweise aus verschiedenen Kohorten oder ohne Berücksichtigung der Auswirkungen einer zwischen Entnahme der Biopsie und Resektion gegebenen Chemotherapie. Dagegen wurden in dieser Arbeit Schnittpräparate von Resektaten und von Gewebeproben in der Größe von Biopsien, die aus demselben Tumormaterial gewonnen wurden, im Hinblick auf Probensammelfehler untersucht. Dabei ergaben sich deutliche Hinweise, dass die Probenentnahme durch Stanzen ein signifikantes Risiko für einen falsch negativen Her2/neu-Status birgt. Fünfundzwanzig Prozent der Magenkarzinompatienten mit Her2/neu-Überexpression wären bei einer bioptischen Untersuchung mit Entnahme von 5 Tumorgewebe enthaltenden Stanzen übersehen worden. Dieser Probensammelfehler hängt mit der bekannten Heterogenität der Her2/neu-Expression im Magenkarzinom zusammen. Allerdings birgt die bioptische Probenentnahme auch ein geringes Risiko für falsch positive Ergebnisse. Sechs Fälle zeigten ein "schwarz-oder-weiß"-Färbemuster. In insgesamt Her2/neu-negativen Magenkarzinomen gab es kleine Her2/neu überexprimierende Tumorzellklone (<10% der gesamten Tumormenge). Die Probenentnahme aus einem solchen Gebiet kann nach den Bewertungsregeln für das Magenkarzinom zu einem falsch positiven Testresultat führen. Der prognostische Einfluss solcher Tumorzellklone ist unbekannt. Zumindest konnte an dieser Kohorte keine prognostische Signifikanz der Her2/neu-Überexpression nachgewiesen werden.

Reproduzierbarkeit des Her2/neu-Status und der 10%-Grenze

In der zweiten Arbeit wurden die Auswirkungen der Grenze von 10% auf die Reproduzierbarkeit des Her2/neu-Status untersucht, indem 10 Pathologen 12 Schnittpräparate mit dem normalen Mikroskop, mit einfacher virtueller Mikroskopie und mit virtueller Mikroskopie mit Assistenzfunktion zum Einzeichnen von Tumorflächen befundeten.

Die Resultate zeigen, dass Pathologen gut in der Lage sind, Flächenverhältnisse zu schätzen, unabhängig von ihrer Erfahrung, aber mit einer leichten systematischen Überbewertung. Es ist der Gradient in der immunhistologischen Färbintensität (Graustufen-Muster), der zu Fehlern führt, weil er es erschwert, die Grenzen zwischen Färbintensitäten (z.B. 2+ und 3+) verlässlich festzulegen. Fälle mit einem scharf begrenzten "schwarz-oder-weiß"-Färbemuster können einfacher und verlässlicher beurteilt werden. Die Reproduzierbarkeit und Fehlerrate hängt hochgradig vom individuellen Färbemuster jedes Falles ab. Die Situation wird verschärft durch die häufig heterogene Verteilung der Her2/neu-positiven Tumorzellklone beim Magenkarzinom, die eine zusätzliche Komplexität zum Färbemusters hinzufügt. Dies kann zu falsch negativen Bewertungen führen, die eine darauffolgende in situ-Hybridisierung verhindern, die den Patienten potentiell einer Therapie mit Trastuzumab zuführen könnte. In dieser Untersuchung

war die Rate der falsch negativen Ergebnisse bis zu 19%, wobei allerdings für diese Studie hauptsächlich Fälle ausgewählt wurden, die im für die Therapieentscheidung kritischen Bereich um 10% lagen.

Die Kenntnis dieser Fehlerquellen ist wichtig, weil sie zum Vorenthalten einer Therapie mit Trastuzumab führen können, oder andersherum zur Anwendung einer Therapie ohne Evidenz für einen Vorteil. Dies wird sogar noch wichtiger durch das Aufkommen neuer zielgerichteter Therapien, die auf an Antikörper gekoppelten Zytostatika beruhen (z.B. Trastuzumab-Emansine), für die präzise "Companion Diagnostics" mit niedrigen falsch negativ- und falsch positiv-Raten benötigen werden.

Der Versuch, die Auswirkungen dieser Effekte durch den Einsatz von virtueller Mikroskopie und das explizite Einzeichnen der Areale von Tumor und angefärbtem Tumor zu reduzieren, führte nicht zu einer Verbesserung. Mangels Verfügbarkeit wohldefinierter "*on slide*"-Referenzen würde auch eine automatische Bildanalyse vermutlich nicht helfen. Die in der Subgruppenanalyse gefundene bessere Reproduzierbarkeit des immunhistochemischen Scores durch Fachärzte als durch Assistenzärzte bestätigt, dass erfahrene Pathologen den Her2/neu-Status bestimmen sollten. Die schlechte Reproduzierbarkeit des Her2/neu-immungefärbten Tumoranteils ist dagegen intrinsisch und in diesem Versuch nicht durch Erfahrung verbesserbar. In Fällen, in denen der gefärbte Tumoranteil um 10% herum liegt, sollte im pathologischen Begutachtungsbericht darauf hingewiesen werden, dass die Möglichkeit einer Fehlinterpretation besteht, und im Zweifel sollte eine *in situ*-Hybridisierung zur Klärung vorgenommen werden. Beim Mammakarzinom wurde kürzlich die Grenze von 10% sogar ganz verworfen⁸, um die Rate an falsch negativen Befunden zu verringern. Dies sollte auch beim Magenkarzinom in Betracht gezogen werden.

Zusammengefasst führt die heterogene Expression und Verteilung prädiktiver Biomarker zu zwei Hauptproblemen: zunächst führt nicht-repräsentative Probenentnahme von Biopsien zu falsch positiven und falsch negativen Testresultaten, was aber durch die Auswertung einer größeren Anzahl von Biopsien oder eines Resektates überwunden werden kann. Das zweite Problem ist die subjektive Bewertung der Immunfärbung. Der Anteil an positiv gefärbtem Tumorgewebe hängt stark davon ab, was der jeweilige Pathologe als das am stärksten gefärbte Tumorgewebe einschätzt, in Verbindung mit einer systematischen Überschätzung dieses Anteils. Bei der Etablierung von zukünftigen auf Immunhistochemie basierenden "Companion Diagnostics" mag besondere Aufmerksamkeit auf heterogene Verteilungen vom Biomarker-

Expressionen und Schwierigkeiten bei der Klassifikation im Bereich um einen Grenzwert gelegt werden.

Screening weiterer Biomarker

Die dritte Arbeit untersuchte die Fragestellungen, ob für das Magenkarzinom weitere phänotypische oder genotypische Biomarker identifizierbar sind, die prädiktive oder prognostische Bedeutung haben, oder mit denen Subtypen klassifiziert werden können. Während für andere Tumorerkrankungen bereits verschiedene zielgerichtete Therapieoptionen bestehen, ist beim Magenkarzinom die bisher einzige solche Möglichkeit die palliative Anwendung von Trastuzumab.

In Übereinstimmung mit vorherigen Studien erwiesen sich an dieser Kohorte der Laurén-Phänotyp, die Ausdehnung des Primärtumors (T), die Invasion regionärer Lymphknoten (N), das Tumorstadium, das Grading und der Resektionsstatus als prognostisch relevante klinisch-pathologische Parameter, wodurch die Kohorte gleichzeitig validiert wird. Zusätzlich wurde gezeigt, dass mikrosatelliteninstabile Magenkarzinome eine bessere Prognose haben als mikrosatellitenstabile, und dass *PIK3CA* (Exon 20) ein unabhängiger prognostischer Faktor für die Vorhersage des Überlebens beim Magenkarzinom des intestinalen Typs sein könnte.

Bei verschiedenen Tumoren werden immunhistochemische Färbungen angewendet, um die Unterklassifikation zu unterstützen. Obwohl E-Cadherin, β -Catenin und Lysozym beim Magenkarzinom zwischen dem intestinalen und diffusen Subtyp signifikant verschieden exprimiert werden, kann keines dazu dienen, verlässlich zwischen den Subtypen zu unterscheiden. Dies ist in Übereinstimmung mit einer Studie⁹, in der Magenkarzinomzelllinien auf translationaler Ebene untersucht wurden. Dort erwies sich allerdings das Gen für Lysozym als signifikant differentiell exprimiert, und deshalb wurde Lysozym auch für diese Studie als weiterer Marker ausgewählt. Lysozym war in dieser Studie in den Tumorzellen von 262 Magenkarzinomen exprimiert und wurde dabei signifikant häufiger im diffusen (62,2%) als im intestinalen (46,8%) Subtyp gefunden. Wie auch für E-Cadherin und β -Catenin war die differentielle Expression von Lysozym zwischen dem intestinalen und diffusen Subtyp hoch signifikant und zeigte sogar prognostische Signifikanz. Dennoch zeigten die Expressionen erhebliche Überlappungen. Auch der Mucin-Phänotyp wurde für die Klassifikation des Magenkarzinoms vorgeschlagen⁶. Obwohl der Laurén-Phänotyp und der Mucin-Phänotyp miteinander korrelieren, kann Immunhistochemie nicht zur Unterstützung der histologischen Unterklassifikation des Magenkarzinoms empfohlen

werden. Die Charakteristika der Immunfärbungen zeigen zu große Überlappungen zwischen den verschiedenen histologischen Subtypen.

In verschiedenen Studien¹⁰⁻¹² wurden genetische Veränderungen an Magenkarzinomen und Magenkarzinomzelllinien untersucht. Dabei wurde eine Vielzahl an Mutationen, Deletionen und epigenetischen Veränderungen gefunden, z.B. an Genen für die Zelladhäsion, den Chromatin-Umbau, Tumorsuppressoren und Tyrosinkinasen, darunter auch *PIK3CA* und *KRAS*. In dieser Arbeit wurde der Mutationstatus von *BRAF*, *KRAS*, *NRAS* und *PIK3CA* durch konventionelle Sequenzierung und der MSI-Status bestimmt und mit den phänotypischen Parametern korreliert. Obwohl der Mutationsstatus von *KRAS* und *PIK3CA* nicht zufällig über die histologischen Subtypen verteilt war, war doch die Prävalenz insgesamt zu niedrig, um die histologische Klassifikation zu unterstützen. Die *KRAS*-mutierten Magenkarzinome zeigten ein medianes Überleben von $3,5 \pm 3,1$ Monaten im Vergleich zu $12,7 \pm 0,7$ Monaten beim Wildtyp. Das Testen des *KRAS*-Mutationsstatus könnte also beim proximalen Magenkarzinom empfohlen werden, um Patienten mit sehr schlechter Prognose zu identifizieren.

In dieser Arbeit wurden 33 (7,0%) Patienten mit hoch mikrosatelliteninstabilen (MSI-H) Magenkarzinomen gefunden. Diese Patienten waren älter und hatten eine längere mediane Überlebenszeit als die Patienten mit mikrosatellitenstabilen Magenkarzinomen. Auf Grund dieser Beobachtungen kann man in Erwägung ziehen, die Gabe von Chemotherapie in dieser kleinen Untergruppe von älteren Patienten zu vermeiden. Das Testen des MSI-Status am Magenkarzinom ist also sinnvoll.

Die umfassende Untersuchung von 13 Biomarkern zeigte, dass diese nicht geeignet sind, um die histologische Klassifikation des Magenkarzinoms zu unterstützen. Allerdings korrelierten einige von ihnen mit dem Phänotyp und/oder der Prognose und könnten in Betracht gezogen werden, um zukünftig die Therapie maßzuschneidern, z.B. der *KRAS*-, *PIK3CA*-, MSI- und H. pylori-Status.

Literaturverzeichnis

1. Cunningham D, Allum WH, Stenning SP, *et al.* Perioperative chemotherapy versus surgery alone for resectable gastroesophageal cancer. *N. Engl. J. Med.* 2006;**355**(1):11-20.
2. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell* 2011;**144**(5):646-674.
3. Bang YJ, Van CE, Feyereislova A, *et al.* Trastuzumab in combination with chemotherapy versus chemotherapy alone for treatment of HER2-positive advanced gastric or gastro-oesophageal junction cancer (ToGA): a phase 3, open-label, randomised controlled trial. *Lancet* 2010;**376**(9742):687-697.
4. Rüschoff J, Dietel M, Baretton G, *et al.* HER2 diagnostics in gastric cancer-guideline validation and development of standardized immunohistochemical testing. *Virchows Arch.* 2010;**457**(3):299-307.
5. Warneke VS, Behrens HM, Hartmann JT, *et al.* Cohort study based on the seventh edition of the TNM classification for gastric cancer: proposal of a new staging system. *J. Clin. Oncol.* 2011;**29**(17):2364-2371.
6. Namikawa T, Hanazaki K. Mucin phenotype of gastric cancer and clinicopathology of gastric-type differentiated adenocarcinoma. *World J. Gastroenterol.* 2010;**16**(37):4634-4639.
7. Chua TC, Merrett ND. Clinicopathologic factors associated with HER2-positive gastric cancer and its impact on survival outcomes--a systematic review. *Int. J. Cancer* 2012;**130**(12):2845-2856.
8. Wolff AC, Hammond ME, Hicks DG, *et al.* Recommendations for human epidermal growth factor receptor 2 testing in breast cancer: American Society of Clinical Oncology/College of American Pathologists clinical practice guideline update. *J. Clin. Oncol.* 2013;**31**(31):3997-4013.
9. Tan IB, Ivanova T, Lim KH, *et al.* Intrinsic subtypes of gastric cancer, based on gene expression pattern, predict survival and respond differently to chemotherapy. *Gastroenterology* 2011;**141**(2):476-85, 485.
10. Deng N, Goh LK, Wang H, *et al.* A comprehensive survey of genomic alterations in gastric cancer reveals systematic patterns of molecular exclusivity and co-occurrence among distinct therapeutic targets. *Gut* 2012;**61**(5):673-684.
11. Lee J, van HP, Go C, *et al.* High-throughput mutation profiling identifies frequent somatic mutations in advanced gastric adenocarcinoma. *PLoS. One.* 2012;**7**(6):e38892.
12. Zang ZJ, Cutcutache I, Poon SL, *et al.* Exome sequencing of gastric adenocarcinoma identifies recurrent somatic mutations in cell adhesion and chromatin remodeling genes. *Nat. Genet.* 2012;**44**(5):570-574.

Eidesstattliche Versicherung

„Ich, Hans-Michael Behrens, versichere an Eides statt durch meine eigenhändige Unterschrift, dass ich die vorgelegte Dissertation mit dem Thema: "Untersuchungen zur immunhistologischen Diagnostik und Prognose des Magenkarzinoms" 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/der Betreuer/in, 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.“

Datum

Unterschrift

Anteilerklärung an den erfolgten Publikationen

Hans-Michael Behrens hatte folgenden Anteil an den folgenden Publikationen:

Publikation 1: Warneke VS, Behrens HM, Böger C, Becker T, Lordick F, Ebert MP, Röcken C. Her2/neu testing in gastric cancer: evaluating the risk of sampling errors. Ann Oncol 2013;24(3):725-33.

Beitrag im Einzelnen: Datenmanagement, Auswertung und Darstellung der Daten, Auswahl, Durchführung und Interpretation der statistischen Tests, Erstellung von Tabellen sowie Abbildungen, Mitarbeit am Manuskript (u.a. Verfassung des Abschnitts über Statistik), Korrektur des Manuskriptes

Publikation 2: Behrens HM, Warneke VS, Böger C, Garbrecht N, Jüttner E, Klapper W, Mathiak M, Oschlies I, Rudolph U, Stuhlmann-Laeisz C, Trick D, Röcken C, Hufnagl P. Reproducibility of Her2/neu scoring in gastric cancer and assessment of the 10% cut-off rule. Cancer Med 2014. doi:10.1002/cam4.365

Beitrag im Einzelnen: Konzeption und Betreuung der Versuche, Programmierung der Software zum Einzeichnen und Ausmessen der Areale in virtuelle mikroskopische Präparate, Anleitung der Versuchsteilnehmer, Datenmanagement, Erfassung, Speicherung, Auswertung und Darstellung der Ergebnisse, Auswahl, Durchführung und Interpretation der statistischen Tests, Erstellung aller Tabellen und Abbildungen, Interpretation der Ergebnisse, selbständige Verfassung und Korrektur des Manuskriptes

Publikation 3: Warneke VS, Behrens HM, Haag J, Balschun K, Böger C, Becker T, Ebert MP, Lordick F, Röcken C. Prognostic and putative predictive biomarkers of gastric cancer for

personalized medicine. *Diagn Mol Pathol* 2013;22(3):127-37.

Beitrag im Einzelnen: Datenmanagement, Auswertung, Darstellung und Interpretation der Daten, Auswahl und Durchführung der statistischen Tests, Erstellung der Tabellen sowie Abbildungen, Mitarbeit am Manuskript (Verfassung des Abschnitts über Statistik), Korrektur des Manuskriptes

Unterschrift, Datum und Stempel des betreuenden Hochschullehrers/der betreuenden Hochschullehrerin

Unterschrift des Doktoranden/der Doktorandin

Druckexemplare der ausgewählten Publikationen

1. Warneke, V. S., Behrens, H. M., Böger, C., Becker, T., Lordick, F., Ebert, M. P. A., Röcken, C. Her2/neu testing in gastric cancer: evaluating the risk of sampling errors. *Annals of oncology* 2013;24(3):725-733. doi:10.1093/annonc/mds528
Impact Factor: 7.384 (Rang 18 von 197 im Bereich Onkologie)
Eigenfactor: 0.07205
2. Behrens HM, Warneke VS, Böger C, Garbrecht N, Jüttner E, Klapper W, Mathiak M, Oshlies I, Rudolph U, Stuhlmann-Laeisz C, Trick D, Röcken C, Hufnagl P. Reproducibility of Her2/neu scoring in gastric cancer and assessment of the 10% cut-off rule. *Cancer Med* 2014. doi:10.1002/cam4.365
Cancer Medicine besteht erst seit 2012 und hat daher noch keinen Impact Factor.
3. Warneke, V. S., Behrens, H. M., Haag, J., Balschun, K., Böger, C., Becker, T., Ebert M. P. A., Lordick F., & Röcken, C. (2013). Prognostic and putative predictive biomarkers of gastric cancer for personalized medicine. *Diagnostic Molecular Pathology* 2013;22(3):127-137. doi:10.1097/PDM.0b013e318284188e
Impact Factor: 1.861 (Rang 46 von 77 im Bereich Pathologie)
Eigenfactor: 0.00176

30. Velenik V, Ocvirk J, Oblak I et al. A phase II study of cetuximab, capecitabine and radiotherapy in neoadjuvant treatment of patients with locally advanced resectable rectal cancer. *Eur J Surg Oncol* 2010; 36: 244–250.
31. Kim SY, Hong YS, Kim DY et al. Preoperative chemoradiation with cetuximab, irinotecan, and capecitabine in patients with locally advanced resectable rectal cancer: a multicenter phase II study. *Int J Radiat Oncol Biol Phys* 2010; 81(3): 677–683.
32. Pinto C, Di Fabio F, Maiello E et al. Phase II study of panitumumab, oxaliplatin, 5-fluorouracil, and concurrent radiotherapy as preoperative treatment in high-risk locally advanced rectal cancer patients (StarPan/STAR-02 Study). *Ann Oncol* 2011; 22(11): 2424–2430.
33. National Comprehensive Cancer Network. NCCN Clinical Practice Guidelines in Oncology. Colon Cancer (version 3. 2011). www.nccn.org. Last accessed 29 March 2011.
34. National Comprehensive Cancer Network. NCCN Clinical Practice Guidelines in Oncology. Rectal Cancer (version 4. 2011). www.nccn.org. Last accessed 29 March 2011.
35. Bengala C, Bettelli S, Bertolini F et al. Epidermal growth factor receptor gene copy number, K-ras mutation and pathological response to preoperative cetuximab, 5-FU and radiation therapy in locally advanced rectal cancer. *Ann Oncol* 2009; 20: 469–474.
36. Debucquoy A, Haustermans K, Daemen A et al. Molecular response to cetuximab and efficacy of preoperative cetuximab-based chemoradiation in rectal cancer. *J Clin Oncol* 2009; 27: 2751–2757.
37. Gaedcke J, Grade M, Jung K et al. KRAS and BRAF mutations in patients with rectal cancer treated with preoperative chemoradiotherapy. *Radiother Oncol* 2010; 94: 76–81.
38. Dewdney A, Cunningham D, Tabernero J et al. Multicenter randomized phase II clinical trial comparing neoadjuvant oxaliplatin, capecitabine, and preoperative radiotherapy with or without cetuximab followed by total mesorectal excision in patients with high-risk rectal cancer (EXPERT-C). *J Clin Oncol* 2012; 30(14): 1620–1627.
39. Beddy D, Hyland JM, Winter DC et al. A simplified tumor regression grade correlates with survival in locally advanced rectal carcinoma treated with neoadjuvant chemoradiotherapy. *Ann Surg Oncol* 2008; 15: 3471–3477.
40. Guillem JG, Chessin DB, Cohen AM et al. Long-term oncologic outcome following preoperative combined modality therapy and total mesorectal excision of locally advanced rectal cancer. *Ann Surg* 2005; 241: 829–836; discussion 836–828.
41. Lacouture ME, Mitchell EP, Piperdi B et al. Skin toxicity evaluation protocol with panitumumab (STEPP), a phase II, open-label, randomized trial evaluating the impact of a pre-emptive skin treatment regimen on skin toxicities and quality of life in patients with metastatic colorectal cancer. *J Clin Oncol* 2010; 28: 1351–1357.
42. Law WJ, Chu KW, Ho JW et al. Risk factors for anastomotic leakage after low anterior resection with total mesorectal excision. *Am J Surg* 2000; 179: 92–96.
43. Nesbakken A, Nygaard K, Lunde OC. Outcome and late functional results after anastomotic leakage following mesorectal excision for rectal cancer. *Br J Surg* 2001; 88: 400–404.
44. Rullier E, Laurent C, Garrelon JL et al. Risk factors for anastomotic leakage after resection of rectal cancer. *Br J Surg* 1998; 85: 355–358.
45. Kalady MF, DeJulius KL, Sanchez JA et al. BRAF mutations in colorectal cancer are associated with distinct clinical characteristics and worse prognosis. *Dis Colon Rectum* 2012; 55: 128–133.

Annals of Oncology 24: 725–733, 2013
doi:10.1093/annonc/mds528
Published online 8 November 2012

Her2/neu testing in gastric cancer: evaluating the risk of sampling errors

V. S. Warneke^{1,†}, H.-M. Behrens^{1,2,†}, C. Böger¹, T. Becker³, F. Lordick⁴, M. P. A. Ebert⁵ & C. Röcken^{1*}

¹Department of Pathology, Christian-Albrechts University, Kiel; ²Department of Pathology, Charité University Hospital, Berlin; ³Department of General Surgery and Thoracic Surgery, Christian-Albrechts University, Kiel; ⁴University Cancer Centre Leipzig, University of Leipzig, Leipzig; ⁵Department of Medicine II, Faculty of Clinical Medicine Mannheim, University of Heidelberg, Mannheim, Germany

Received 11 April 2012; revised 2 September 2012; accepted 4 September 2012

Background: We evaluated the risk of sampling errors in specimens of biopsy size, which may be caused by heterogeneous overexpression of Her2/neu in gastric cancer (GC).

Patients and methods: The study cohort comprised 454 gastrectomy patients with adenocarcinoma of the stomach or esophago-gastric junction. Tissue micro-arrays (TMAs) served as 'biopsy procedure' and were generated from formalin-fixed and paraffin-embedded tissue: five tissue cylinders were collected randomly from each tumor, rendering 2230 core cylinders. These were compared with 454 whole tissue sections obtained from the same paraffin blocks. Her2/neu expression and gene amplification were analyzed by immunohistochemistry and *in situ* hybridization. The Her2/neu status was determined according to GC scoring system by two independent observers.

*Correspondence to: Dr C. Röcken, Department of Pathology, Christian-Albrechts University, Arnold-Heller-Str. 3, Haus 14, D-24105 Kiel, Germany.
Tel: +49-431-597-3401; Fax: +49-431-597-3462; E-mail: christoph.roecken@uk-sh.de

[†]VSW and H-MB contributed equally to this work.

Results: In whole tissue sections, 37 (8.1%; observer 1) and 38 (8.4%; observer 2) of the GCs, and in the corresponding TMAs, 28 (6.3%; observer 1) and 28 (6.3%; observer 2) of the GCs were classified as Her2/neu-positive (kappa value 98.5% and 96.2%; $P < 0.001$). Comparison of whole tissue sections with corresponding TMAs showed a false-negative rate of 24% and a false-positive rate of 3% for TMAs.

Conclusion: Assessment of the Her2/neu status in tissue biopsies carries a significant risk of sampling errors, thereby rendering patients unsuitable for treatment with trastuzumab.

Key words: biopsy, gastric cancer, Her2/neu, sampling error

introduction

For many years, only the anatomical location of the primary tumor, its histological phenotype and the tumor stage tailored chemotherapy. However, in clinical practice, many patients with a seemingly identical tumor responded differently to the same therapy. Research on cancer biology provided ample explanations [1]. Various genetic alterations and distinct molecular phenotypes were unraveled, which influence patient prognosis and response to chemotherapy. Meanwhile, the choice of treatment increasingly depends on the identification of these particular molecular phenotypes. As a consequence, predictive biomarkers were sought, which prognosticate treatment responses before therapy. This approach bears hope to make treatment more efficient in responders and to avoid unnecessary side-effects in non-responders. Despite these advancements, new problems emerged. The value of a predictive biomarker test depends on the robustness, sensitivity and specificity of the assay. These factors can be negatively influenced by tumor heterogeneity and tissue sampling errors.

Recently, Her2/neu was introduced as a predictive biomarker for the treatment of gastric cancer (GC) with trastuzumab. Trastuzumab is an antibody targeting Her2/neu and is applied in combination with chemotherapy for the treatment of Her2/neu-positive advanced GC [2]. The Her2/neu status is assessed in tumor biopsies or resection specimens by a combination of immunohistochemistry and *in situ* hybridization. Based on the breast cancer scoring system, a modified scoring system was developed for GC [3]. An almost overwhelming number of studies recently demonstrated the robustness of the Her2/neu testing (for a review, see also [4]). However, evidence is accumulating that Her2/neu overexpression is heterogeneous in GC, carrying the risk of a sampling error [5–14]. Particularly in patients with advanced, unresectable GC eligible for trastuzumab, only biopsies may be available for Her2/neu testing. In this study, we assessed the reliability of Her2/neu testing in tissue samples of biopsy size by comparing the Her2/neu status of whole tissue sections with corresponding tissue micro-arrays (TMAs).

materials and methods

ethics statement

This project was approved by the local ethics committee of the University Hospital in Kiel, Germany (reference number D 453/10). All patient data were pseudonymized before study inclusion.

study population

We retrieved from the archive of the Institute of Pathology, University Hospital Kiel, all Caucasian patients who had undergone either total or partial gastrectomy for adenocarcinoma of the stomach or esophago-gastric

junction between 1997 and 2009. The following patient characteristics were retrieved: type of surgery, age at diagnosis, gender, tumor localization, tumor type, tumor grade, depth of invasion, number of lymph nodes resected and number of lymph nodes with metastases. Each resection specimen underwent gross sectioning and histological examination by trained histopathologists. Date of patient death was obtained from the Epidemiological Cancer Registry of the state of Schleswig-Holstein, Germany. Follow-up data of patients still alive were retrieved from hospital records and general practitioners.

study inclusion and exclusion criteria

Inclusion and exclusion criteria were defined as follows: patients were included when histology confirmed an adenocarcinoma of the stomach or esophago-gastric junction, and the date of death or survival data were available. Patients were excluded when histology identified a tumor type other than adenocarcinoma, histopathological data were incomplete, patients had previously undergone a partial stomach resection (Billroth-II) and now had locally recurrent GC and date of patient death or survival data had not been recorded. Patients who received perioperative chemotherapy were also excluded.

histology and TNM classification

Tissue specimens were fixed in formalin and embedded in paraffin. Deparaffinized sections were stained with hematoxylin and eosin. Tumors were classified according to the Laurén classification [15]. pTNM stage of all study patients was determined according to seventh edition of the UICC guidelines [16] and our recent proposal ('Kiel-stage') [17] and was based solely on surgical pathological examination including classification of distant metastases (pM-category). In the seventh edition, all tumors of the esophago-gastric junction as well as tumors of the proximal 5 cm of the stomach with extension into the esophagus are classified as esophageal tumors [16]. Patients were re-categorized accordingly.

TMA construction

Formalin-fixed and paraffin-embedded tissue samples were used to generate TMAs as described previously [18]. Briefly, five morphologically representative regions of the paraffin 'donor' blocks were chosen. Tissue cylinders of 1.5 mm diameter were punched from these areas and precisely arrayed into a new 'recipient' paraffin block using a custom-built instrument (Beecher Instruments, Silver Spring, MD). After completing block construction, 4 µm sections of the resulting tumor TMA block were cut for further analysis.

Her2/neu status

The Her2/neu status was assessed using the monoclonal anti-Her2/neu antibody (clone 4B5), the HER2-SISH double-labeling *in situ*-hybridization system and the Ventana BenchMark XT automated slide staining system (all Roche Diagnostics GmbH, Mannheim, Germany). The Her2/neu status was assessed by two independent observers (VSW and CB) according to the GC scoring system separately for resection specimens (whole tissue

sections) and for biopsies (TMAs) [3]. The intensity of Her2/neu immunostaining varied from negative (0) to strong (3+). All GCs with an Her2/neu immunostaining of 2+ were forwarded to silver-enhanced *in situ* hybridization. Both observers were blinded with regard to the clinicopathological patient characteristics. After independent evaluation of the whole tissue sections and the TMAs by both observers, a final consensus evaluation was carried out with three observers (VSW, CB and CR).

statistics

Statistical analyses were carried out using IBM SPSS Version 20 (IBM Corp., Armonk, NY). Median overall survival was determined using the Kaplan–Meier method, and the log-rank test was used to determine significance. For comparison purposes, the median survival time, its standard deviation and 95% confidence interval were calculated. Prognostic relevance was investigated by multivariate Cox regression analysis, using the backward LR method with inclusion/exclusion significances of 0.100, respectively. The significance of correlation between clinicopathological parameters and the Her2/neu status was tested using Fisher's exact test. For parameters of ordinal scale (T, N, stage), we applied Kendall's tau test instead. For measuring the interrater agreement of Her2/neu scores between two observers, Cohen's kappa coefficient was calculated. A kappa value of 0.20 was considered to be poor, of 0.21–0.40 to be fair, of 0.41–0.60 to be moderate, of 0.61–0.80 to be good, and of 0.81–1.00 to be very good. A *P*-value ≤ 0.05 was considered statistically significant. No adjustments were made.

results

In this study, 454 patients fulfilled all study criteria (Table 1). According to Laurén, an intestinal type GC was found in 232 (51.1%), a diffuse type in 140 (30.8%), an unclassified type in 53 (11.7%) and a mixed type in 29 (6.4%) patients.

Her2/neu status in whole tissue sections

The Her2/neu status was assessed according to the GC scoring system for resection specimens, using whole tissue sections [3]: GCs were classified as Her2/neu-positive when $\geq 10\%$ of the tumor area showed strong immunostaining (3+) for Her2/neu, or moderate immunostaining (2+) in conjunction with *HER2* gene amplification (ratio ≥ 2.0). Observer 1 classified 37 (8.1%) and observer 2 classified 38 (8.4%) of the GCs as Her2/neu-positive. The interobserver agreement was very good (kappa 98.5%, *P* < 0.001; Table 2).

Her2/neu status in TMAs

Next we assessed the Her2/neu status in TMAs obtained from the same tumor-bearing paraffin blocks used for the assessment of whole tissue sections. Five core cylinders were obtained from each paraffin 'donor' block, finally rendering 2230 tumor-bearing core cylinders (in eight cases, the amount of material was too low for the generation of a TMA). The Her2/neu status was assessed according to the GC scoring system for biopsy specimens [3]: GCs were classified as Her2/neu-positive when five or more tumor cells showed strong immunostaining (3+) for Her2/neu, or moderate immunostaining (2+) in conjunction with *HER2* gene amplification (ratio ≥ 2.0) in at least one of the five core cylinders obtained from a given tumor. Observer 1 classified 28 (6.3%) and observer 2 classified 28 (6.3%) of the GCs as Her2/

neu-positive. Again, the interobserver agreement was very good (kappa 96.2%, *P* < 0.001; Table 2).

comparison of the Her2/neu status between whole tissue sections and TMAs

In the next step, we compared the Her2/neu status between whole tissue sections and TMAs. Observer 1 classified nine (2.0%) GCs as Her2/neu-positive in whole tissue sections and negative in the corresponding TMAs. To the contrary, three (0.7%) cases were classified as Her2/neu-positive in TMAs and negative in whole tissue sections. Observer 2 classified nine (2.0%) GCs as Her2/neu-positive in whole tissue sections and negative in the corresponding TMAs. To the contrary, two (0.4%) cases were classified as Her2/neu-positive in the TMAs and negative in the whole tissue sections. The interassay agreement was good to very good for both observers (kappa 79.2% and 81.2%, respectively; *P* < 0.001; Table 2).

comparison of Her2/neu immunostaining between whole tissue sections and TMAs

According to the GC scoring system, Her2/neu immunostaining is the first diagnostic procedure for the assessment of the Her2/neu status. An Her2/neu immunoreactivity score of $\leq 1+$ and of 3+ does not necessitate *in situ* hybridization. Thus, a heterogeneous Her2/neu immunostaining has a major impact on the Her2/neu status classified in a biopsy specimen. Therefore, we next compared the results of immunostaining between whole tissue sections and TMAs. As shown in Table 2, the interassay agreement further declined for both observers, when only the Her2/neu immunostaining results were compared between whole tissue sections and TMAs (kappa 47.3% and 44.8%, respectively). However, Her2/neu immunostaining results correlated still highly significantly between whole tissue sections and TMAs for both observers (*P* < 0.001).

case-by-case analysis of discrepant Her2/neu status

Following the comparison of the immunostaining results, we compared case-by-case the whole tissue sections with the TMAs. Interestingly, the discrepancies between whole tissue sections and TMAs were mainly related to sampling errors (Table 3). In six cases, the TMA was false negative since the GC enclosed an Her2/neu-positive tumor cell clone (IHC 3+ or IHC2+/ISH-amplified) comprising only 10%–20% (four cases) or 40% (one case) of the entire tumor area. The remaining tumor area was completely devoid of Her2/neu expression ('black-and-white' expression pattern; Figure 1). In three cases, the intensity of immunostaining for Her2/neu was uneven. The GCs enclosed areas with weak (1+), moderate (2+) and strong (3+) immunostaining ('gray-scale' expression pattern). The intensity of Her2/neu-immunostained tumor areas enclosed in the TMAs was not representative for the entire tumor.

Two cases were false positive in the TMA for similar reasons ('black-and-white' expression pattern): five or more cells showed strong (3+) or moderate immunostaining (2+) in

Table 1. Clinicopathological patient characteristics

Patient characteristics	
Patients (<i>n</i>)	454
Age (years)	
Mean ± SD	67.3 ± 11.1
Median	68
Gender, <i>n</i> (%)	
Men	283 (62.3)
Women	171 (37.7)
Follow-up data, <i>n</i> (%)	
Alive	120 (27.1)
Dead	322 (72.9)
Laurén phenotype, <i>n</i> (%)	
Intestinal	232 (51.1)
Diffuse	140 (30.8)
Unclassified	53 (11.7)
Mixed	29 (6.4)
Localization, <i>n</i> (%)	
Proximal	144 (31.7)
Distal	310 (68.3)
pT-category, <i>n</i> (%)	
pT1a	10 (2.2)
pT1b	37 (8.1)
pT2	55 (12.1)
pT3	185 (40.7)
pT4a	128 (28.2)
pT4b	39 (8.6)
pN-category, <i>n</i> (%)	
pN0	124 (27.5)
pN1	65 (14.4)
pN2	81 (18.0)
pN3/a/b	181 (40.1)
Resection margin, <i>n</i> (%)	
pR0	380 (88.0)
pR1	49 (11.3)
pR2	3 (0.7)
Lymphatic invasion, <i>n</i> (%)	
pL0	212 (48.3)
pL1	227 (51.7)
Venous invasion, <i>n</i> (%)	
pV0	386 (88.3)
pV1	51 (11.7)
UICC stage (7th edition), <i>n</i> (%)	
IA	36 (8.1)
IB	30 (6.7)
IIA	57 (12.8)
IIB	46 (10.3)
IIIA	54 (12.1)
IIIB	80 (18.0)
IIIC	64 (14.4)
IV	78 (17.5)
Stage according to Kiel proposal, <i>n</i> (%)	
I	36 (8.0)
II	83 (18.4)
IIIA	47 (10.4)
IIIB	148 (32.7)
IV	138 (30.5)

Continued

Table 1.. Continued

Patient characteristics	
Resected lymph nodes	
Mean ± SD	19.1 ± 8.2
Median, <i>n</i>	18
Positive lymph nodes	
Mean ± SD	6.5 ± 7.4
Median, <i>n</i>	3.5
Lymph node ratio	
Median, <i>n</i>	0.22
Tumor grade, <i>n</i> (%)	
G1/G2	99 (22.4)
G3/G4	343 (77.6)

conjunction with *HER2* gene amplification (ratio ≥ 2.0) in the TMAs samples, while the whole tissue section showed that these clones comprised $<10\%$ of the entire tumor area (Figure 1). In these cases, we also noticed that the interobserver variability was due to observer-dependent differences in the estimation of the cut-off value, i.e. $<10\%$ versus $\geq 10\%$ of the tumor area (case no. 10; Table 3).

Finally, one case (no. 12) was categorized as false positive due to a staining artifact. The monoclonal antibody used in this study is known to occasionally generate a strong cytoplasmic and nuclear immunoreaction [19]. The signet ring cell carcinoma of patient no. 12 showed strong cytoplasmic staining, which was misinterpreted by both observers as membranous staining and became evident for both observers in the whole tissue section (Figure 1).

Statistical analyses

Finally, a consensus score was reached by three observers (VSW, CB, CR) and a total of 37 (8.1%) GCs were classified as Her2/neu-positive (supplementary Table S1, available at *Annals of Oncology* online). The Her2/neu status correlated highly significantly with tumor type ($P = 0.002$) and tumor grade ($P = 0.003$). Proximal tumors were more commonly Her2/neu-positive (11.8%), compared with distal tumors (6.5%). However, this difference did not reach statistical significance ($P = 0.065$). We next separated proximal tumors from distal tumors. This showed that the Her2/neu status correlated significantly with nodal spread ($P = 0.034$; proximal tumors) and tumor grade ($P = 0.009$; distal tumors). Patient survival did not correlate with the Her2/neu status (Figure 2). However, patient prognosis highly significantly depended on patient age, Laurén phenotype, tumor grade, T-category, N-category, lymph node ratio, R-status, as well as UICC stage and 'Kiel stage' (data not shown).

discussion

The phenotypic and genotypic characterization of cancer is increasingly used to tailor patient management. The introduction of trastuzumab for distant metastatic Her2/neu-positive GC in the first-line treatment added GC to a growing list of cancers to which targeted therapy can be applied [2].

Table 2. Interrater agreement on the assessment of the Her2/neu status (kappa)

Subject	Comparing	Contingency table	p-value of Fisher's exact test	Kappa	p-value of Kappa																																								
Her2/neu-status of whole section	Observer 1 versus Observer 2	<table border="1"> <thead> <tr> <th rowspan="2">Whole Section Her2 status</th> <th colspan="2">Observer 2</th> <th rowspan="2">Total</th> </tr> <tr> <th>negative</th> <th>positive</th> </tr> </thead> <tbody> <tr> <td>Observer 1 negative</td> <td>416</td> <td>1</td> <td>417</td> </tr> <tr> <td>Observer 1 positive</td> <td>0</td> <td>37</td> <td>37</td> </tr> <tr> <td>Total</td> <td>416</td> <td>38</td> <td>454</td> </tr> </tbody> </table>	Whole Section Her2 status	Observer 2		Total	negative	positive	Observer 1 negative	416	1	417	Observer 1 positive	0	37	37	Total	416	38	454	<0.001	98.5%	<0.001																						
Whole Section Her2 status	Observer 2			Total																																									
	negative	positive																																											
Observer 1 negative	416	1	417																																										
Observer 1 positive	0	37	37																																										
Total	416	38	454																																										
Her2/neu-status of TMA	Observer 1 versus Observer 2	<table border="1"> <thead> <tr> <th rowspan="2">TMA Her2 status</th> <th colspan="2">Observer 2</th> <th rowspan="2">Total</th> </tr> <tr> <th>negative</th> <th>positive</th> </tr> </thead> <tbody> <tr> <td>Observer 1 negative</td> <td>417</td> <td>1</td> <td>418</td> </tr> <tr> <td>Observer 1 positive</td> <td>1</td> <td>27</td> <td>28</td> </tr> <tr> <td>Total</td> <td>418</td> <td>28</td> <td>446</td> </tr> </tbody> </table>	TMA Her2 status	Observer 2		Total	negative	positive	Observer 1 negative	417	1	418	Observer 1 positive	1	27	28	Total	418	28	446	<0.001	96.2%	<0.001																						
TMA Her2 status	Observer 2			Total																																									
	negative	positive																																											
Observer 1 negative	417	1	418																																										
Observer 1 positive	1	27	28																																										
Total	418	28	446																																										
Her2/neu-status of Observer 1	Whole section versus TMA	<table border="1"> <thead> <tr> <th rowspan="2">Observer 1 Whole Section Her2 status</th> <th colspan="2">TMA Her2 status</th> <th rowspan="2">Total</th> </tr> <tr> <th>negative</th> <th>positive</th> </tr> </thead> <tbody> <tr> <td>negative</td> <td>409</td> <td>3</td> <td>412</td> </tr> <tr> <td>positive</td> <td>9</td> <td>25</td> <td>34</td> </tr> <tr> <td>Total</td> <td>418</td> <td>28</td> <td>446</td> </tr> </tbody> </table>	Observer 1 Whole Section Her2 status	TMA Her2 status		Total	negative	positive	negative	409	3	412	positive	9	25	34	Total	418	28	446	<0.001	79.2%	<0.001																						
Observer 1 Whole Section Her2 status	TMA Her2 status			Total																																									
	negative	positive																																											
negative	409	3	412																																										
positive	9	25	34																																										
Total	418	28	446																																										
Her2/neu-status of Observer 2	Whole section versus TMA	<table border="1"> <thead> <tr> <th rowspan="2">Observer 2 Whole Section Her2 status</th> <th colspan="2">TMA Her2 status</th> <th rowspan="2">Total</th> </tr> <tr> <th>negative</th> <th>positive</th> </tr> </thead> <tbody> <tr> <td>negative</td> <td>409</td> <td>2</td> <td>411</td> </tr> <tr> <td>positive</td> <td>9</td> <td>26</td> <td>35</td> </tr> <tr> <td>Total</td> <td>418</td> <td>28</td> <td>446</td> </tr> </tbody> </table>	Observer 2 Whole Section Her2 status	TMA Her2 status		Total	negative	positive	negative	409	2	411	positive	9	26	35	Total	418	28	446	<0.001	81.2%	<0.001																						
Observer 2 Whole Section Her2 status	TMA Her2 status			Total																																									
	negative	positive																																											
negative	409	2	411																																										
positive	9	26	35																																										
Total	418	28	446																																										
Her2/neu-IHC score of Observer 1	Whole section versus TMA	<table border="1"> <thead> <tr> <th rowspan="2">Observer 1 Whole section IHC score</th> <th colspan="4">TMA IHC score</th> <th rowspan="2">Total</th> </tr> <tr> <th>0</th> <th>1+</th> <th>2+</th> <th>3+</th> </tr> </thead> <tbody> <tr> <td>0</td> <td>363</td> <td>15</td> <td>4</td> <td>0</td> <td>382</td> </tr> <tr> <td>1+</td> <td>16</td> <td>5</td> <td>0</td> <td>0</td> <td>21</td> </tr> <tr> <td>2+</td> <td>6</td> <td>5</td> <td>3</td> <td>0</td> <td>14</td> </tr> <tr> <td>3+</td> <td>4</td> <td>1</td> <td>7</td> <td>17</td> <td>29</td> </tr> <tr> <td>Total</td> <td>389</td> <td>26</td> <td>14</td> <td>17</td> <td>446</td> </tr> </tbody> </table>	Observer 1 Whole section IHC score	TMA IHC score				Total	0	1+	2+	3+	0	363	15	4	0	382	1+	16	5	0	0	21	2+	6	5	3	0	14	3+	4	1	7	17	29	Total	389	26	14	17	446	X ² : <0.001	47.3 %	<0.001
Observer 1 Whole section IHC score	TMA IHC score				Total																																								
	0	1+	2+	3+																																									
0	363	15	4	0	382																																								
1+	16	5	0	0	21																																								
2+	6	5	3	0	14																																								
3+	4	1	7	17	29																																								
Total	389	26	14	17	446																																								
Her2/neu-IHC score of Observer 2	Whole section versus TMA	<table border="1"> <thead> <tr> <th rowspan="2">Observer 2 Whole section IHC score</th> <th colspan="4">TMA IHC score</th> <th rowspan="2">Total</th> </tr> <tr> <th>0</th> <th>1+</th> <th>2+</th> <th>3+</th> </tr> </thead> <tbody> <tr> <td>0</td> <td>352</td> <td>17</td> <td>1</td> <td>2</td> <td>372</td> </tr> <tr> <td>1+</td> <td>23</td> <td>6</td> <td>0</td> <td>0</td> <td>29</td> </tr> <tr> <td>2+</td> <td>6</td> <td>6</td> <td>5</td> <td>2</td> <td>19</td> </tr> <tr> <td>3+</td> <td>4</td> <td>1</td> <td>5</td> <td>16</td> <td>26</td> </tr> <tr> <td>Total</td> <td>385</td> <td>30</td> <td>11</td> <td>20</td> <td>446</td> </tr> </tbody> </table>	Observer 2 Whole section IHC score	TMA IHC score				Total	0	1+	2+	3+	0	352	17	1	2	372	1+	23	6	0	0	29	2+	6	6	5	2	19	3+	4	1	5	16	26	Total	385	30	11	20	446	X ² : <0.001	44.8 %	<0.001
Observer 2 Whole section IHC score	TMA IHC score				Total																																								
	0	1+	2+	3+																																									
0	352	17	1	2	372																																								
1+	23	6	0	0	29																																								
2+	6	6	5	2	19																																								
3+	4	1	5	16	26																																								
Total	385	30	11	20	446																																								

TMA denotes tissue micro array and IHC immunohistochemistry

Trastuzumab can only be used when the cancer has been shown to overexpress Her2/neu, i.e. when $\geq 10\%$ of the tumor area of a GC resection specimen or when a tumor cell cluster (five or more tumor cells) of a biopsy specimen shows strong immunostaining (3+) for Her2/neu, or moderate immunostaining (2+) in conjunction with Her2/neu gene amplification (ratio ≥ 2.0) (<http://www.ema.europa.eu>). Various protocols have been tested and found suitable to test the Her2/neu status in GC. This includes different primary antibodies and *in situ*-hybridization protocols [3, 20]. We used the monoclonal anti-Her2/neu antibody (clone 4B5), the HER2-SISH double-labeling *in situ*-hybridization system and the Ventana BenchMark XT automated slide staining system, which is a highly standardized protocol and was validated independently in the past [3].

Using whole tissue sections, we found an Her2/neu overexpression in 8.1% of our patients. It was more common in

intestinal-type GCs compared with diffuse-type GCs (12.9% versus 3.6%), and was more common in proximal GCs compared with distal GCs (11.8% versus 6.5%). These findings are in line with previous observations. A recent systematic review reported a prevalence of Her2/neu overexpression ranging from 4%–53% (median 18%) [4, 21]. A correlation with the anatomical localization was also found by Yu et al. [22]. Thus, neither methodological issues (fixation and staining protocols) nor the selection of the study cohort explain the discrepancy between whole tissue sections and TMAs.

The NCCN Guidelines recommend that multiple (8–10) biopsies should be carried out to provide an adequate-sized material for histological interpretation [23]. Similarly, the German S3-guideline for the diagnosis and treatment of esophago-gastric cancer recommends that at least eight biopsies should be taken from tumor suspicious lesions [24]. Ten biopsies should be obtained from large lesions.

Table 3. Case-by-case analysis of discrepant Her2/neu status

	No.	Observer 1						Observer 2						Comment
		Tissue micro-array			Whole tissue section			Tissue micro-array			Whole tissue section			
		IHC	ISH	Her2	IHC	ISH	Her2	IHC	ISH	Her2	IHC	ISH	Her2	
False negative	1	2+	NA	-	3+	A	+	2+	NA	-	3+	A	+	Tumor cell clone with IHC 3+ comprising 40% of the tumor area; clone reaching the mucosal surface; available for endoscopic biopsy procedure
	2	0		-	3+	A	+	0		-	3+	A	+	Mucinous adenocarcinoma; tumor cell clone with IHC 3+ comprising 15% of the tumor area; clone localized at the invasion front unavailable for endoscopic biopsy procedure
	3	0		-	3+	A	+	0		-	3+	A	+	Tumor cell clone IHC 3+ comprising 10%–20% of the tumor area; clone reaching the mucosal surface; available for endoscopic biopsy procedure
	4	0		-	3+	A	+	0		-	3+	A	+	Tumor cell clone IHC 3+ comprising 10%–20% of the tumor area; clone reaching from mucosal surface to the level of the subserosa; available for endoscopic biopsy procedure
	5	0		-	3+		+	0		-	3+		+	Tumor cell clone IHC 3+ comprising 10%–20% of the tumor area; clone reaching from mucosal surface down to the level of the muscularis propria; available for endoscopic biopsy procedure
	6	0		-	2+	A	+	0		-	2+	A	+	Tumor cell clone IHC 2+ comprising 20% of the tumor area; clone reaching mucosal surface; available for endoscopic biopsy procedure
	7	1+		-	2+	A	+	1+		-	2+	A	+	Variable Her2/neu expression (IHC 0, 1+, 2+, 3+); tumor areas with IHC 3+ localized at the mucosal surface; available for endoscopic biopsy procedure
	8	1+		-	3+	A	+	1+		-	3+	A	+	Variable Her2/neu expression (IHC 0, 1+, 2+, 3+); tumor areas with IHC 3+ localized at the mucosal surface; available for endoscopic biopsy procedure
	9	1+		-	2+	A	+	1+		-	2+	A	+	Variable Her2/neu expression (IHC 0, 1+, 2+, 3+); gene amplification was even across entire tumor area; possibly fixation artifact.
False positive	10	2+	A	+	0		-	2+	A	+	2+	A	+	Tumor cell clone IHC 3+ comprising <10% of the tumor area; in the TMA, more than five cells with IHC 3+ were present in one of five core cylinders; in the remaining four core cylinders tumor cells were completely Her2/neu negative
	11	2+	A	+	0		-	2+	NA	-	0		-	Tumor cell clone IHC 2+ comprising <10% of the tumor area in whole tissue section; clone localized at the mucosal surface; available for endoscopic biopsy procedure
	12	2+	NA	-	0		-	3+		+	0		-	Staining artifact: signet ring cells showed strong cytoplasmic immunostaining, which was misinterpreted as membranous staining in TMA

A, amplified; NA, not amplified; IHC, immunohistochemistry; Her2, Her2/neu status; ‘-’ denotes negative; ‘+’ denotes positive.

Particularly, diffuse-type GC can easily be missed by biopsies and these numbers were chosen to reach a safe diagnosis of GC by avoiding false-negative results. They were not based on the assumption that GC can only be diagnosed when a minimum number of eight biopsies enclose cancer. In our experience, the number of cancer-bearing biopsies is usually less than eight, and five core cylinders obtained from the formalin-fixed and paraffin-embedded tumor may be

considered to be representative for the evaluation of endoscopic biopsy sampling of GC.

We believe that we are the first to systematically study the effect of tissue sampling on the Her2/neu status in a single patient cohort of 454 Caucasian patients, using the GC scoring system. Yano et al. [25] compared whole tissue sections and biopsies in an Asian study population. In their cohort, the biopsy was positive in 76.7% of the cases with overexpression

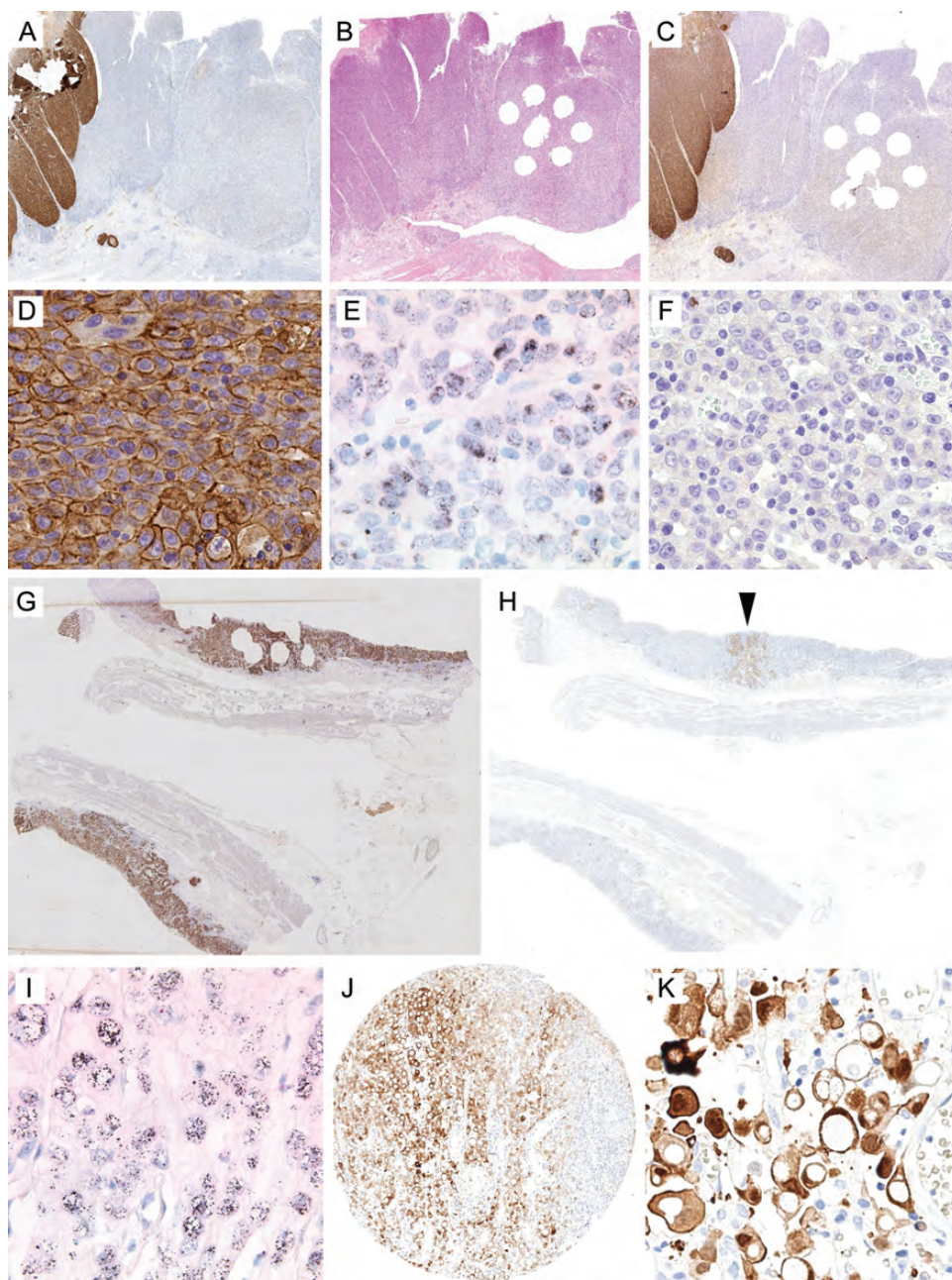


Figure 1. Her2/neu expression of gastric cancer assessed in whole tissue sections and tissue micro-arrays (TMAs). A case-by-case analysis illustrated the causes of discrepant Her2/neu test results between whole tissue sections and TMAs: (A–F) Case no. 5 (see Table 3) encloses a tumor cell clone comprising 10%–20% of the entire tumor area, which showed strong (IHC 3+) expression of Her2/neu (A, C and D) and *HER2* gene amplification (E). The remaining tumor area was completely negative for Her2/neu (IHC 0; F). A tissue section cut after the TMA was generated shows that, unintentionally, only the negative tumor area was ‘biopsied’. (G and H) Case no. 10 (see Table 3) illustrates the reverse situation of a false-positive test result. An Her2/neu-overexpressing tumor cell (H, arrowhead) clone comprised <10% of the entire tumor area, as illustrated by cytotokeratin immunolabeling (G). This clone also showed *HER2* gene amplification (I). (J and K) Case no. 12 illustrates a unique but well-recognized staining artifact of the monoclonal antibody we used (clone 4B5). Strong cytoplasmic immunostaining of a signet ring cell carcinoma was erroneously interpreted as membrane staining in TMAs. However, both observers also classified the same case as negative in the whole tissue sections, where the staining artifact was more apparent. Her2/neu immunostaining (A, B, D, F, H, J and K); *HER2*-SISH double-labeling *in situ* hybridization (E and I); hematoxylin and eosin (B); pancyokeratin (the brown stain is virtually identical with the tumor area (G)). No magnification (A–C, G, H, J); original magnifications $\times 400$ (D, F and K) and $\times 600$ (E and I). The digitalized original slides of (C) and (H) can be viewed at <http://www.uni-kiel.de/path/vm/her2-manuscript>.

in subsequent resection specimens. However, Yano et al. [25] applied the breast cancer scoring system to both biopsy and resection specimens. Yan et al. [14] compared 15 resection specimens with 12 matched biopsy specimens obtained from

the same patients before gastrectomy, using the GC scoring system. They did not find a significant correlation of the Her2/neu status between biopsy and surgical resection specimens. Neither Yano et al. [25] nor Yan et al. [14] specified the

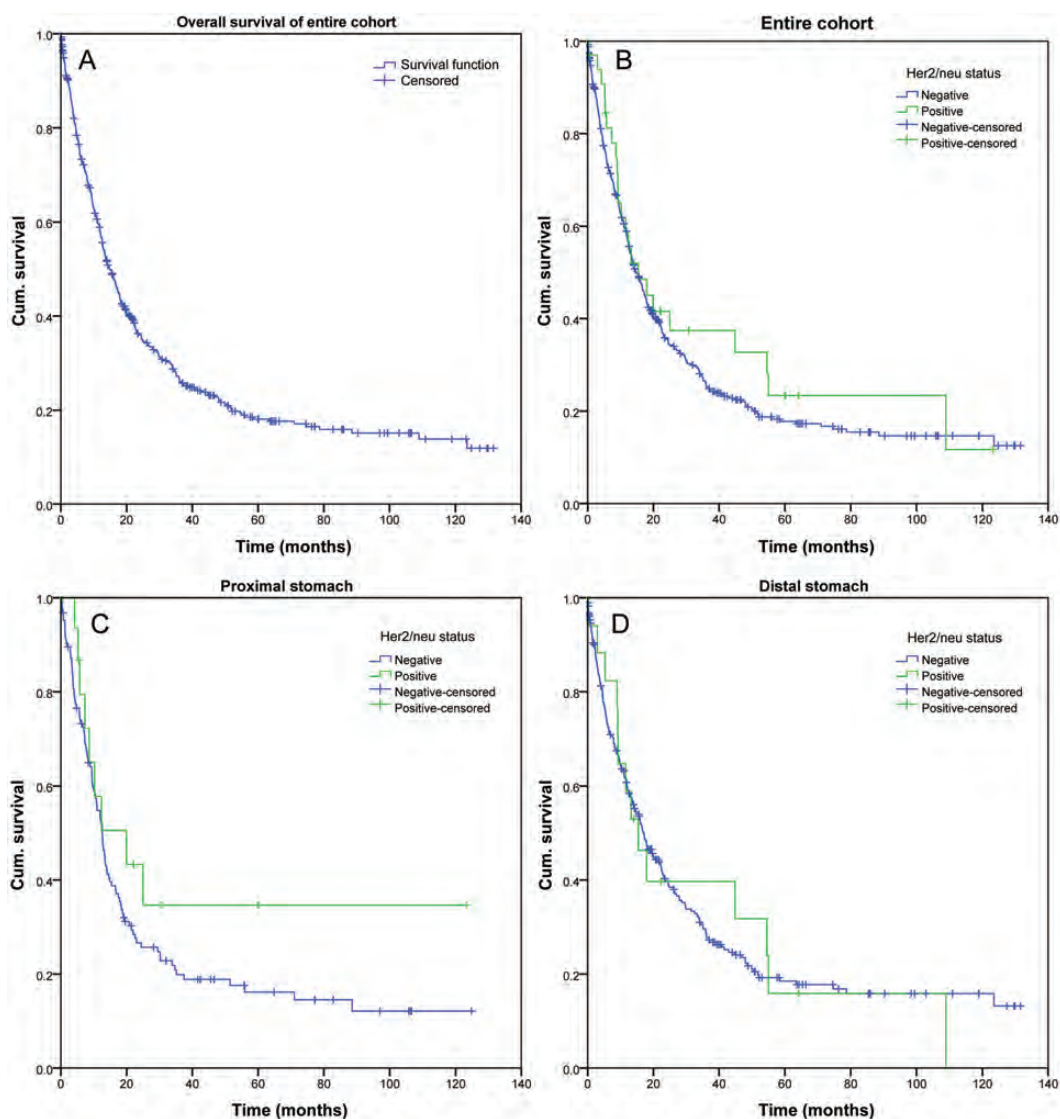


Figure 2. Patient survival according to the Her2/neu status. Kaplan–Meier curves depicting patient survival of the entire patient cohort (A), according to the Her2/neu status in all cases (B; $P = 0.452$), the Her2/neu status in proximal (C; $P = 0.203$) and distal cancers (D; $P = 0.984$).

number of tumor-bearing biopsies studied. Kim et al. [26] recently studied the Her2/neu status, by immunohistochemistry only, in two large separate Asian cohorts. In the first cohort of 1414 patients, whole tissue sections were analyzed. In the second, independent cohort TMAs were generated from 595 resection specimens (one core cylinder per tumor and patient). The frequency of Her2/neu-positive cases was 12.3% in whole tissue sections and 17% in TMAs [26]. However, based on the study design, no conclusions can be drawn from this study regarding the positive or negative predictive value of biopsy sampling. Finally, Lee et al. [11] compared biopsies (mean number of biopsy fragments was 4.9; range 1–11) and resection specimens obtained from 54 patients. The concordance for the Her2/neu status between biopsy and gastrectomy was 74.1% [11]. However, the time interval between biopsy procedure and gastrectomy was not reported, and confounding factors, such as perioperative chemotherapy, were not specifically mentioned and cannot be completely ruled out [11].

To the contrary, our study investigated whole tissue sections and TMAs obtained from the same paraffin blocks, addressing specifically the issue of a sampling error. We provide strong evidence that sampling procedures carry a significant risk to generate a false-negative Her2/neu status. Twenty-five percent of our GC patients with Her2/neu overexpression would have been missed in a biopsy procedure harvesting five tumor-bearing biopsies. This sampling error is related to the well-recognized heterogeneity of Her2/neu expression in GC. This has been reported independently now by many authors [5–8, 10, 12–14]. However, biopsy sampling carries also a minor risk of a false-positive rate. Six cases showed a ‘black-and-white’ expression pattern (Figure 1 and Table 3). Small Her2/neu-overexpressing tumor cell clones (<10% of the entire tumor volume) were present in an overall Her2/neu-negative GC. Sampling of these tumor cell clones carries the risk of a false-positive test result, according to the GC scoring system. The prognostic impact of these cell clones is unknown. At least

we were unable to show any prognostic significance of Her2/neu overexpression in our cohort.

Patient-individualized treatment aims to avoid unnecessary medication in patients who are unlikely to respond to therapy. To the contrary, targeted therapy should reach every patient eligible for the treatment. In this respect, unresectable GC patients may be withheld a life-prolonging medication due to a sampling error. Since the overall number of GC patients eligible for trastuzumab is already small (<20%), we recommend that, whenever possible, the Her2/neu status should be assessed on whole tissue sections of gastrectomy specimens. Furthermore, in case of metastatic disease, it should be considered to re-evaluate the Her2/neu status from metastatic sites, when Her2/neu results are negative from primary tumor biopsies, in order to increase the probability of finding Her2/neu-positivity, which allows for treatment with trastuzumab in stage-IV GC. Additionally, more in-depth investigations on the correlation of the percentage of Her2/neu-immunopositive GCs with clinical outcomes may help to improve predictive diagnostics. Changing the cut-off value to above or below 10% positivity influences the risk of a sampling error, i.e. it may also alter the discrepancies between whole tissue sections obtained from a GC resection specimen and tumor biopsies. Alternatively, a predictive biomarker should be sought, which is less sensitive to sampling procedures.

acknowledgements

We wish to thank Sandra Krüger for her excellent technical assistance.

funding

This work was supported by grant no. F344605 of Roche Diagnostics Deutschland GmbH and Roche Pharma AG. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

disclosure

The authors have declared no conflicts of interest.

references

- Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell* 2011; 144: 646–674.
- Bang YJ, Van CE, Feyereislova A et al.. Trastuzumab in combination with chemotherapy versus chemotherapy alone for treatment of HER2-positive advanced gastric or gastro-oesophageal junction cancer (ToGA): a phase 3, open-label, randomised controlled trial. *Lancet* 2010; 376: 687–697.
- Rüschoff J, Dietel M, Baretton G et al.. HER2 diagnostics in gastric cancer—guideline validation and development of standardized immunohistochemical testing. *Virchows Arch* 2010; 457: 299–307.
- Chua TC, Merrett ND. Clinicopathologic factors associated with HER2-Positive gastric cancer and its impact on survival outcomes – a systematic review. *Int J Cancer* 2012; 130(12): 2845–2856.
- Bilous M, Osamura RY, Ruschoff J et al.. HER-2 amplification is highly homogenous in gastric cancer. *Hum Pathol* 2010; 41: 304–305.
- Bozzetti C, Negri FV, Lagrasta CA et al.. Comparison of HER2 status in primary and paired metastatic sites of gastric carcinoma. *Br J Cancer* 2011; 104: 1372–1376.
- Kimura M, Tsuda H, Morita D et al.. Usefulness and limitation of multiple endoscopic biopsy sampling for epidermal growth factor receptor and c-erbB-2 testing in patients with gastric adenocarcinoma. *Jpn J Clin Oncol* 2005; 35: 324–331.
- Langer R, Rauser S, Feith M et al.. Assessment of ErbB2 (Her2) in oesophageal adenocarcinomas: summary of a revised immunohistochemical evaluation system, bright field double in situ hybridisation and fluorescence in situ hybridisation. *Mod Pathol* 2011; 24: 908–916.
- Abraham SC, Park SJ, Lee JH et al.. Genetic alterations in gastric adenomas of intestinal and foveolar phenotypes. *Mod Pathol* 2003; 16: 786–795.
- Kim MA, Lee HJ, Yang HK et al.. Heterogeneous amplification of ERBB2 in primary lesions is responsible for the discordant ERBB2 status of primary and metastatic lesions in gastric carcinoma. *Histopathology* 2011; 59: 822–831.
- Lee S, de Boer WB, Fermoy S et al.. Human epidermal growth factor receptor 2 testing in gastric carcinoma: issues related to heterogeneity in biopsies and resections. *Histopathology* 2011; 59: 832–840.
- Tafe LJ, Janjigian YY, Zaidinski M et al.. Human epidermal growth factor receptor 2 testing in gastroesophageal cancer: correlation between immunohistochemistry and fluorescence in situ hybridization. *Arch Pathol Lab Med* 2011; 135: 1460–1465.
- Yan B, Yau EX, Bte Omar SS et al.. A study of HER2 gene amplification and protein expression in gastric cancer. *J Clin Pathol* 2010; 63: 839–842.
- Yan B, Yau EX, Choo SN et al.. Dual-colour HER2/chromosome 17 chromogenic in situ hybridisation assay enables accurate assessment of HER2 genomic status in gastric cancer and has potential utility in HER2 testing of biopsy samples. *J Clin Pathol* 2011; 64: 880–883.
- Lauren T. The two histologic main types of gastric carcinoma: diffuse and so-called intestinal-type carcinoma. *Acta Pathol Microbiol Scand* 1965; 64: 31–49.
- Sobin LH, Gospodarowicz M, Wittekind C. *TNM Classification of Malignant Tumours*, 7th edition. Wiley-Blackwell: Oxford, 2009.
- Warneke VS, Behrens HM, Hartmann JT et al.. Cohort study based on the seventh edition of the TNM classification for gastric cancer: proposal of a new staging system. *J Clin Oncol* 2011; 29: 2364–2371.
- Weichert W, Röske A, Gekeler V et al.. Association of patterns of class I histone deacetylase expression with patient prognosis in gastric cancer: a retrospective analysis. *Lancet Oncol* 2008; 9: 139–148.
- Rüschoff J, Hanna W, Bilous M et al.. HER2 testing in gastric cancer: a practical approach. *Mod Pathol* 2012; 25: 637–650.
- Fox SB, Kumarasinghe MP, Armes JE et al.. Gastric HER2 Testing Study (GaTHER): an evaluation of gastric/gastroesophageal junction cancer testing accuracy in Australia. *Am J Surg Pathol* 2012; 36: 577–582.
- Park YS, Hwang HS, Park HJ et al.. Comprehensive analysis of HER2 expression and gene amplification in gastric cancers using immunohistochemistry and in situ hybridization: which scoring system should we use. *Hum Pathol* 2012; 43: 413–422.
- Yu GZ, Chen Y, Wang JJ. Overexpression of Grb2/HER2 signaling in Chinese gastric cancer: their relationship with clinicopathological parameters and prognostic significance. *J Cancer Res Clin Oncol* 2009; 135: 1331–1339.
- Ajani JA, Barthel JS, Bekaii-Saab T et al.. Gastric cancer. *J Natl Compr Canc Netw* 2010; 8: 378–409.
- Moehler M, Al-Batran SE, Andus T et al.. German S3-guideline 'Diagnosis and treatment of esophagogastric cancer'. *Z Gastroenterol* 2011; 49: 461–531.
- Yano T, Doi T, Ohtsu A et al.. Comparison of HER2 gene amplification assessed by fluorescence in situ hybridization and HER2 protein expression assessed by immunohistochemistry in gastric cancer. *Oncol Rep* 2006; 15: 65–71.
- Kim KC, Koh YW, Chang HM et al.. Evaluation of HER2 protein expression in gastric carcinomas: comparative analysis of 1,414 cases of whole-tissue sections and 595 cases of tissue microarrays. *Ann Surg Oncol* 2011; 18: 2833–2840.

ORIGINAL RESEARCH

Reproducibility of Her2/neu scoring in gastric cancer and assessment of the 10% cut-off rule

Hans-Michael Behrens^{1,2}, Viktoria S. Warneke¹, Christine Böger¹, Nele Garbrecht¹, Eva Jüttner¹, Wolfram Klapper¹, Micaela Mathiak¹, Ilske Oschlies¹, Ursula Rudolph¹, Christiane Stuhlmann-Laeisz¹, David Trick¹, Christoph Röcken^{1*} & Peter Hufnagl^{2*}

¹Department of Pathology, Christian-Albrechts-University, Kiel, Germany

²Department of Pathology, Charité University Hospital, Berlin, Germany

Keywords

Cut-off value, gastric cancer, Her2/neu, virtual microscopy, visual perception

Correspondence

Christoph Röcken, Department of Pathology, Christian-Albrechts-University, Arnold-Heller-Str. 3, Haus 14, D-24105 Kiel, Germany.
Tel: +49(0)431-597-3401;
Fax: +49(0)431-597-3462;
E-mail: christoph.roecken@uk-sh.de

Funding Information

CR is supported by grants of the German Research Council (Ro 1173/12)

Received: 17 July 2014; Revised: 19 September 2014; Accepted: 24 September 2014

doi: 10.1002/cam4.365

*Shared senior authorship.

Abstract

The application of Trastuzumab on gastric cancer patients is based on Her2/neu immunostaining. The testing method relies on visual estimation of both membranous staining intensity, and positive tumor ratio with respect to a 10% cutoff. We evaluated the effect of inter- and intraobserver variations of both factors on therapeutic decision, especially if the positive tumor ratio hovers around the 10% cutoff. Ten pathologists scored 12 Her2/neu immunohistologically stained whole sections of gastric cancer. Applying the common rules for Her2/neu testing for gastric cancer, they separately noted the strongest identifiable staining intensity and the corresponding positive tumor ratio. Scoring was done repeatedly using the microscope, plain virtual microscopy, and virtual microscopy with a manual outline drawing function. Agreements on the strongest identified staining intensities were moderate. Overall concordance correlation coefficients of positive tumor ratios ranged from 0.55 to 0.81. Reproducibility was not improved by virtual microscopy. Pathologists have a good ability to estimate ratios of clearly demarcated areas, but gradients in staining intensities hinder reproducible visual demarcation of positive tumor areas. When hovering around the 10% positive tumor ratio cutoff there is a risk of misinterpretation of the staining results. This could lead to a denial of Trastuzumab therapy. Assessment of Her2/neu expression should be carried out by experienced pathologists because they can more reproducibly rate membranous staining intensities. The low reproducibility of positive tumor ratio is inherent in the testing method and cannot be improved by virtual microscopy. Therefore, we propose to reconsider the 10% cut-off limit.

Introduction

Gastric cancer (GC) is the second most common cause of cancer-related deaths in the world. Approximately 70% of the patients have already lymph node metastases at the time of the diagnosis. Complete resection of the primary tumor with D2-lymphadenectomy offers the only chance of cure in the early stage of the disease. Survival of more locally advanced GCs was significantly improved by the introduction of perioperative, adjuvant, and palliative chemotherapy.

Recently, Her2/neu was introduced as a predictive biomarker for the treatment of GC with trastuzumab. Trast-

uzumab is an antibody targeting Her2/neu and is applied in combination with chemotherapy for the treatment of Her2/neu positive advanced GC [1]. The Her2/neu status is assessed by surgical pathologists using tumor tissue obtained by biopsy or by resection and immunohistochemistry in combination with in situ hybridization. A GC is Her2/neu positive, when $\geq 10\%$ of the tumor cells show strong circumferential, lateral, or baso-lateral immunostaining, or when $\geq 10\%$ of the tumor cells show weak to moderate circumferential, lateral, or baso-lateral immunostaining in combination with HER2/neu gene amplification.

An almost overwhelming number of studies demonstrated the robustness of the Her2/neu testing (for a

review see also [2]). However, the assessment of Her2/neu status is hampered by (1) its heterogeneous expression in GC, carrying the risk of a sampling error [3–14], and (2) by the surgical pathologist's visual perception of what is below and above 10%. In a previous study [14], we evaluated the risk of sampling errors in specimens of biopsy size, which may be caused by heterogeneous overexpression of Her2/neu in GC. Tissue microarrays served as "biopsy procedure" and were compared with 454 whole tissue sections obtained from the same paraffin blocks used for the generation of tissue microarrays. The Her2/neu status was determined according to GC scoring system [15] by two independent observers using immunohistochemistry and in situ hybridization. In that study, we identified the particular problem of visual assessment of positive ($\geq 10\%$ positive tumor cells) or negative ($< 10\%$) when the amount of positive tumor cells is near the cut-off value of 10%. This motivated us to design an experiment to further validate the problem of the cut-off value and assess the agreement of Her2/neu scoring between multiple observers and trying to find a method leading to more reproducible results.

Our experiment now assesses the agreement of the strongest identifiable staining intensity as well as the positive tumor ratio between pathologists and methods, using (1) the standard microscopic method, (2) virtual microscopy, and (3) virtual microscopy with additional assistance for outlining tissue areas.

Materials and Methods

Participants

Ten pathologists were recruited as participants. Six had been practicing as board-certified pathologists for 4–10 years (median 5.5 years), and four as residents for 3–6 years (median 4.5 years).

Samples

Twelve Her2/neu immunohistochemically stained large sections (monoclonal antibody 4B5; Roche Diagnostics GmbH, Mannheim, Germany) were selected from a previous study of 454 cases of GC [14]. Since we wanted to assess reproducibility of positive tumor with respect to the 10% cut-off threshold, we selected cases from the previous study that had been problematic in whether the positive tumor ratio was above or below 10%, and added some cases with a positive tumor ratio in higher ranges. Additionally, for all 12 cases Her2/neu gene amplification was evaluated using the HER2-SISH double-labeling in situ hybridization system and the Ventana BenchMark XT automated slide staining system (all Roche Diagnostics

GmbH). Identities or further data of the 12 cases were not known to the pathologists. Characteristics of the cases are given in Table S1.

Virtual microscopy

The samples were scanned using a Leica SCN400 microscopic whole-slide scanner (Leica Biosystems, Nussloch, Germany) at its maximum, nominally 40 times magnification. In the scanned images, pixel-to-pixel distance represents $0.26 \mu\text{m}$. Images were exported from the scanner system into files of Leica SCN format. For performing the computer-assisted parts of the experiment, a viewer program was written to display images of Leica SCN file format. This gave us the flexibility to create the screen layout, user interaction, assistance tool, and calculation routines we needed for our experiment. A view of the program is depicted in Figure S1.

Procedure

The 10 pathologists rated each of the 12 slides repeatedly using three methods (summing up to 36 ratings per pathologist and a total of 360 ratings in the entire study). The pathologists were asked for two values: First, the raw value of the strongest immunostaining they could identify within the specimen on the scale 0–3 analogous to the rules described by Rüschoff et al. [15] (0, no reactivity; 1, barely visible; 2, weak to moderate; 3, strong), but still without applying the cut-off rule of 10% positive tumor ratio; Second, an estimation of the ratio of only this strongest stained tumor tissue to total tumor tissue (0–100%). The intent was to spot the cases where the 10% cut-off rule would have had to be applied. In order to minimize memorization at least 2 months were allowed to pass between rounds.

Method 1: Microscope

First the pathologists performed a Her2/neu scoring of the samples using their own, familiar microscope as in daily work. They sequentially received the set of 12 slides along with a questionnaire asking for the strongest staining they could identify, applying also Rüschoff's magnification rule [15], as well as the positive tumor ratio.

Method 2: Virtual microscopy

Rating of staining intensity and positive tumor ratio was repeated on the scanned slides, utilizing our viewer program. The program was configured to offer only viewing functions (zoom and move), but no additional aid. Magnification buttons ($40\times$, $20\times$, $10\times$, $5\times$) allowed to switch

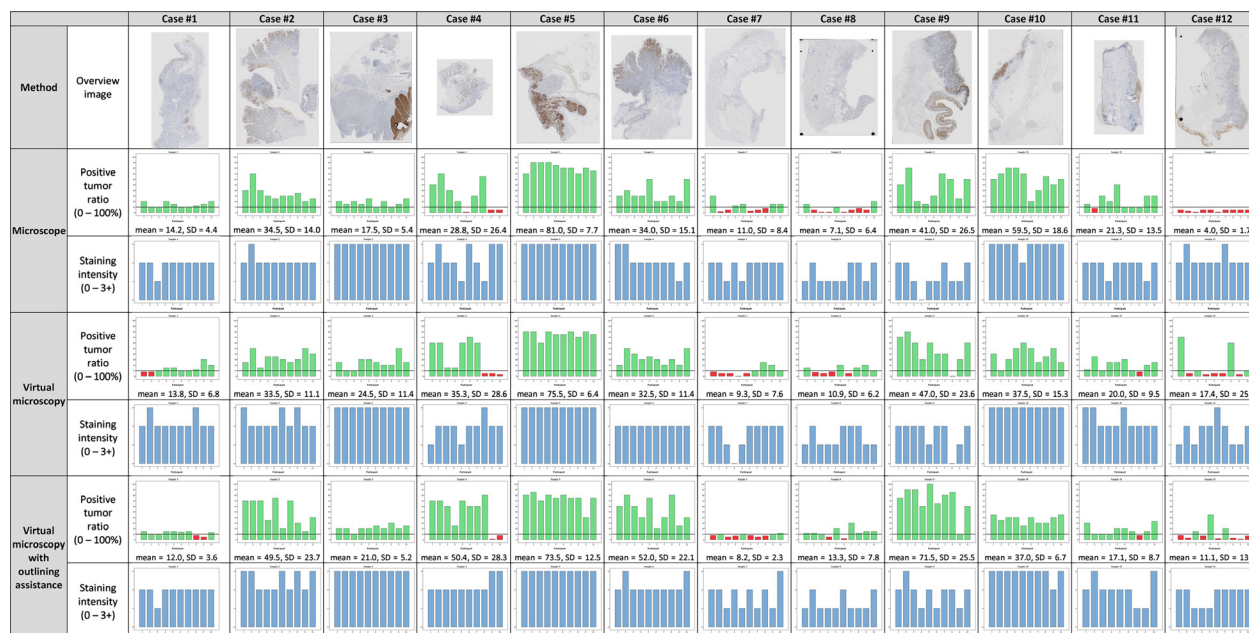


Figure 1. Overview of all ratings, grouped by the rating method (microscopy, virtual microscopy, and assisted virtual microscopy.) Each diagram shows 10 bars, corresponding to the ratings of one specimen by 10 pathologists, using the indicated method. Ratings are shown separately for positive tumor ratio (green bars $\geq 10\%$, and red bars $< 10\%$), and staining intensity (blue bars). Below the diagrams, mean and standard deviation of positive tumor ratio of the 10 ratings are indicated. The top row shows overview images of the 12 scanned slides.

magnification in order to imitate the handling of a conventional microscope.

Method 3: Virtual microscopy with area outlining assistance

In the third round of rating we first repeated the estimation of staining intensities with the viewer program. For measuring positive tumor ratio we tested an alternative, computer-assisted method. We extended the viewer program by a polygon line drawing function. This was used by the pathologists to separately trace the outlines of total tumor tissue and positive tumor tissue (Fig. S1A). Finally they were presented a homogeneously color-filled sketch of the outlines they had drawn (Fig. S1B), and were asked to visually estimate the positive tumor ratio from this sketch. The drawing was saved in a file, so we could calculate the exact ratio afterward for comparison.

Statistics

Statistical analyses and tests were conducted using SPSS version 20 (IBM Corporation, Armonk, NY) and R version 3.0.1 (R Foundation for Statistical Computing, Vienna, Austria). Intrarater agreements of categorical variables (immunoreactive scores [IRs]) between two rounds, respectively, were determined using the kappa

test. A kappa value of 0.01–0.20 was considered to be slight agreement, of 0.21–0.40 to be fair, of 0.41–0.60 to be moderate, of 0.61–0.80 to be substantial agreement, and of 0.81–1.00 to be almost perfect agreement [16]. Interrater agreements of categorical variables (IRs within one round) were calculated using Fleiss' kappa test [17], which is appropriate for multiple observers rating multiple subjects, using the irr package for R [18]. Agreements between continuous variables (positive tumor ratios) were calculated using the overall concordance correlation coefficient (OCCC) [19] which is implemented in the epiR package for R [20].

External quality assurance

Both HER2/neu testing methods, immunohistochemistry as well as in situ hybridization, were certified successfully in 2013 by the quality assurance program of the German Society of Pathology and the *Bundesverband Deutscher Pathologen e.V.*

Results

The pathologists' individual ratings of staining intensity (0, 1+, 2+, or 3+) and positive tumor ratio (i.e., percentage of stained tumor tissue, 0–100%) were plotted as bar diagrams to illustrate the interobserver variations (Fig. 1).

Table 1. Interobserver agreements between pathologists using one method of rating.

Method	Agreement of positive tumor ratio			Agreement of staining intensity	
	OCCC	Mean of standard deviations [%] (range)	Number of specimens with 10% cut-off disagreements	Fleiss' kappa	
1. Microscope	0.682	12.1 (1.7–26.5)	4	0.44 ($P < 0.001$)	
2. Virtual microscopy	0.551	13.7 (6.2–25.3)	7	0.45 ($P < 0.001$)	
3. Virtual microscopy with area outlining assistance	Visual estimation by the pathologists	0.672	13.1 (2.3–28.3)	6	0.40 ($P < 0.001$)
	By computation of outlined areas	0.694	12.3 (1.9–25.2)	7	

Positive tumor ratios were compared using the mean of standard deviations, overall concordance correlation coefficient (OCCC), and number of discordant ratings with respect to the 10% cutoff. Staining intensities were compared using Fleiss' kappa test.

We quantified interobserver agreements separately for each method using statistical tests appropriate for the situation of multiple observers (pathologists) rating multiple subjects (slides). For staining intensities (ordinal scale) we used Fleiss' Kappa test [17], which takes a value between 0 (no agreement) and 1 (perfect agreement). For positive tumor ratios (continuous scale) we used the OCCC [19], which also takes a value between 0 and 1. Additionally, we calculated the standard deviations of positive tumor ratios per slide (Fig. 1) and averaged over all slides. Table 1 summarizes the interobserver agreements.

We took the consensus Her2/neu status that was rated by a panel of three pathologists from the previous study [14] as reference with which we compared each of the ratings in this study. This allowed identification of false-negative ratings due to the following causes: (1) Staining intensity was rated as strong (3+), but underestimation of stained tumor ratio (<10%) led to negative Her2/neu status. (2) Staining intensity was rated as moderate (2+) and stained tumor ratio was underestimated (<10%), thus no SISH would have been carried out. (3) Staining intensity was underestimated as 1+ or 0, but previously

was 3+ or 2+ in combination with a positive SISH result.

Results from Method 1 (Microscope)

When using the microscope, we found a 12.1% mean standard deviation of positive tumor ratio, ranging from 1.7% to 26.5% over the 12 cases, and an OCCC of 0.682. Four cases were discordantly rated with respect to the 10% cutoff. Interobserver agreement of the rating of staining intensity showed a moderate agreement (Fleiss' kappa was 0.44). A comparison of the IHC findings with the reference results illustrates the risk of underestimation of Her2/neu immunopositive tumor area (<10%): six ratings underestimated the tumor area (<10%) and no SISH would have been carried out leading to denial of medication, although HER2/neu gene amplification was found by SISH. Thus, Her2/neu 2+ cases are extremely sensitive to misrating at the 10% cut-off point (Table 2). Interestingly, underestimation of staining intensity even led to 10 false-negative ratings in five cases. Combination of both causes showed 17 false-negative ratings in five cases. In

Table 2. Numbers of false-negative ratings and cases due to underestimation of positive tumor area and/or staining intensity.

	Number of false-negative ratings/number of cases affected		
	Microscope	Virtual microscopy	Assisted virtual microscopy
False-negative because positive tumor ratio was underestimated (<10%), while staining intensity was 2+ or 3+ and SISH was positive	6 (7%)/2	6 (7%)/3	6 (7%)/2
False-negative because staining intensity was underestimated (0 or 1+), while positive tumor ratio \geq 10% and SISH positive	10 (11%)/5	2 (2%)/1	7 (8%)/4
Total false-negative rate because positive tumor ratio or staining intensity or both were underestimated, while SISH was positive	17 (19%)/5	12 (13%)/4	16 (18%)/4

Numbers in parentheses denote the false-negative percentage related to the 90 ratings per method that were made for the nine cases with positive reference Her2/neu status (10 pathologists times nine positive cases). Interestingly, all these cases showed HER2/neu gene amplification, illustrating the risk of underestimating the percentage of Her2/neu-immunostained tumor area or intensity of immunostaining.

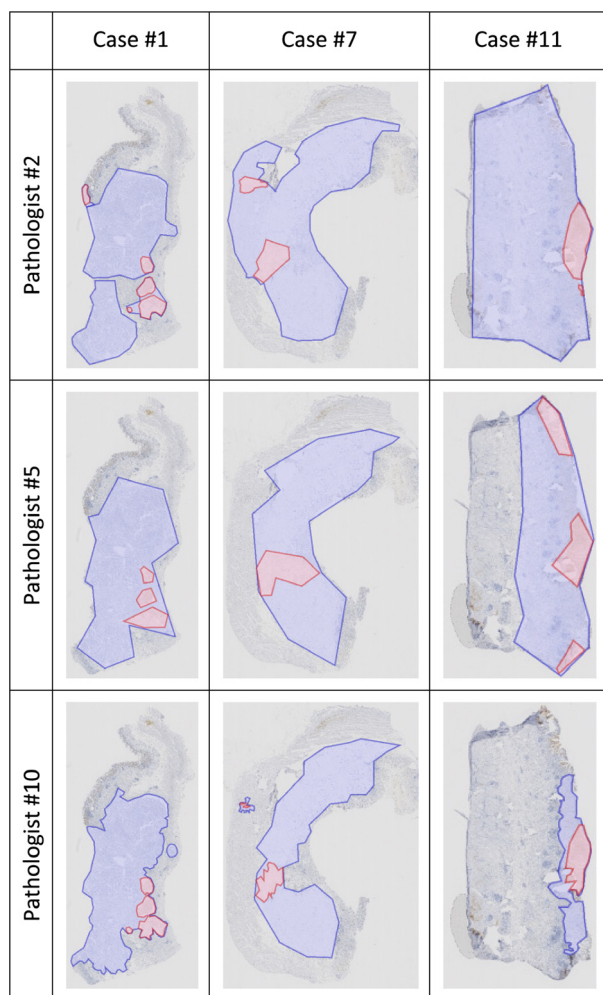


Figure 2. Examples of the outlines of tumor areas (pale blue) and positive tumor areas (pale red) that were manually drawn by three pathologists. They illustrate the variations in assessment of positive tumor areas, and also of total tumor area. Full data are shown in Table S2B.

relation to the 90 ratings that were performed on the nine SISH-positive cases, this corresponds to a false-negative rate of 19%.

Results from Method 2 (Virtual microscopy)

Applying plain virtual microscopy did not improve the reproducibility of the positive tumor ratio. We found 13.7% mean standard deviation of positive tumor ratio (range 6.2–25.3%), and an OCCC of 0.551. Seven cases were discordantly rated concerning the 10% cutoff. Staining intensities showed a moderate agreement (Fleiss' kappa was 0.45). Comparison with the reference consensus Her2/neu status revealed six false-negative ratings in three cases due to underestimation of posi-

tive tumor ratio, two false-negative ratings in one case due to underestimation of staining intensity, and combined 12 false-negative ratings in four cases (Table 2).

We noticed that cases having low standard deviations of positive tumor ratio tend to express a black-and-white staining pattern with sharply demarcated borders (e.g., cases 1, 3, and 5). On the other hand, cases with a more gray scale or heterogeneous expression pattern show a high standard deviation (e.g., cases 2, 4, 6, and 9).

Results from Method 3 (Virtual microscopy with area outlining assistance)

The addition of manual area outlining prior to visual estimation of positive tumor ratio rendered several findings: First, there were differences in what the pathologists considered positively stained tumor, but also what they considered tumor tissue (Figs. 2 and S2). Second, interobserver reproducibility (OCCC 0.672) was about the same as in microscopy, and slightly better than with plain virtual microscopy. Third, we found a nearly perfect correlation between the positive tumor ratio that was visually estimated from the outline sketches (Figs. 2 and S2) and the ratios that were afterward calculated from the sketches (Fig. 3C, Pearson's correlation coefficient 0.979, $P < 0.001$). This demonstrates that the pathologists are very well capable of estimating area ratios. However, we found 19 positive tumor ratios visually estimated on the outline sketches were above 10%, but turned out to be below 10% by calculation. Linear regression showed that the positive tumor ratio was visually overestimated compared with the calculated values. The overestimation is about 3.5% when the positive tumor ratio hovers around 10%. Thus, all 10 surgical pathologists (almost perfectly but systematically) overestimated the percentage of the positive tumor area.

An overview of the comparison of Her2/neu status of this study with the reference consensus Her2/neu status by the panel of three pathologists from the previous study [14] is given in Table 3. Percentages of Her2/neu status ratings not concordant with the reference status ranged from 10% to 22.5% between the methods. We found discordances in six of 12 cases when rating was carried out using the microscope (Method 1). Application of virtual microscopy (Methods 2 and 3) showed a harmonizing effect by reducing the number of discordantly rated cases to 4. Interestingly, in case nos. 1 and 7 the computation of outlined areas from method 3 showed a positive tumor ratio below 10% in 19 of 20 outline drawings, indicating that probably the positive reference result was overestimated.

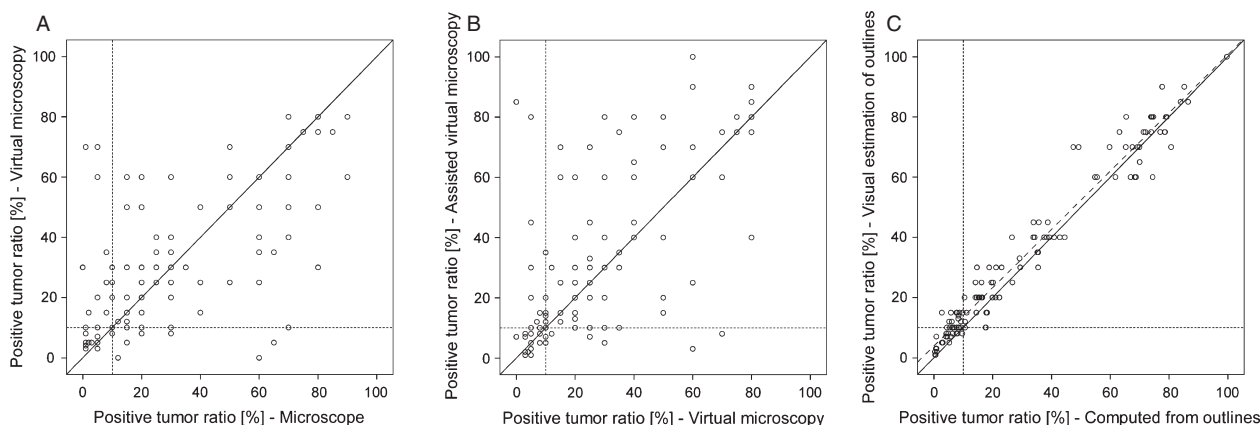


Figure 3. Comparison of positive tumor ratios between methods. Each data point represents two positive tumor ratios of the same case, rated by the same pathologist, using two methods. Solid diagonal lines mark complete agreement. Dotted lines denote 10% cutoff and isolate the discordant ratings in the upper left and lower right quadrant. (A) compares the microscopic method with virtual microscopy, and (B) compares virtual microscopy with assisted virtual microscopy. (C) How well the pathologists were able to estimate the positive tumor ratios of the outline sketches they had drawn, compared with their calculated ratios (Pearson’s correlation coefficient was 0.974, $P < 0.001$.) Linear regression (dashed line) shows that positive tumor ratios were systematically overestimated.

Agreements between methods

Agreements between methods were calculated using the Kappa test for staining intensities (Tables S2 and 4), and OCCC for positive tumor ratios (Table 4.) Additionally, we displayed the corresponding positive tumor ratio estimations from two methods, respectively, as

scatter plots (Fig. 3) and calculated Pearson’s correlation coefficient.

Comparing Method 1 and Method 2 (microscope vs. virtual microscopy), intraobserver variation of positive tumor ratio showed a broad scattering (Fig. 3A), accompanied by an OCCC of 0.647 and a Pearson’s correlation coefficient of 0.651 ($P < 0.001$). Between Method 2 and

Table 3. Comparison of Her2/neu status results (10 individual pathologists) with consensus results from the previous study [14] (panel of three pathologists).

Case no.	Consensus results from the previous study [14] (panel of three pathologists)			Number of discordant ratings (10 individual pathologists)			
				Method 1: microscope		Method 3: virtual microscopy with manual outlining	
	Her2/neu IRS	Her2/neu SISH	Her2/neu Status	Method 1: microscope	Method 2: virtual microscopy	Visual estimation by pathologists	Computation of outlined areas
1	3+	Positive	Positive	1	2	3	10
2	3+	Positive	Positive	0	0	0	0
3	3+	Positive	Positive	0	0	0	0
4	2+	Negative	Negative	1	0	0	0
5	3+	Positive	Positive	0	0	0	0
6	3+	Positive	Positive	1	0	0	0
7	3+	Positive	Positive	6	6	8	10
8	0	Negative	Negative	0	0	0	0
9	3+	Positive	Positive	6	3	3	3
10	3+	Positive	Positive	0	0	0	0
11	3+	Positive	Positive	3	1	2	4
12	0	Negative	Negative	0	0	0	0
Total				18 ¹ (15%)	12 (10%)	16 (13.3%)	27 (22.5%)

Except for one false-positive rating (case no. 4, Method 1), all discordant ratings were false-negative with respect to the consensus Her2/neu status. Total percentages of disagreements refer to the 120 ratings per method that were carried out by the 10 pathologists on the 12 slides.

¹17 false-negative and 1 false-positive rating.

Table 4. Intraobserver agreements, comparing two rating methods.

Methods compared	Agreement of positive tumor ratio			Agreement of staining intensity
	OCCC	Pearson's correlation coefficient	Number of disagreements	Kappa
Microscopy versus virtual microscopy	0.647	0.651 ($P < 0.001$)	20	0.505 ± 0.068 ($P < 0.001$)
Plain virtual microscopy versus virtual microscopy with area outlining assistance	0.634	0.655 ($P < 0.001$)	18	0.509 ± 0.067 ($P < 0.001$)
Microscopy versus virtual microscopy with area outlining	0.665	0.680 ($P < 0.001$)	18	0.551 ± 0.066 ($P < 0.001$)

Agreements of positive tumor ratio were compared using OCCC, Pearson's correlation coefficient, and the number of discordant ratings with respect to the 10% cutoff. Agreements of staining intensities were calculated as kappa values.

Method 3 (plain virtual microscopy versus virtual microscopy with drawing of outlines) OCCC was 0.634 and Pearson's correlations coefficient was 0.655 ($P < 0.001$).

We counted the intraobserver disagreements whether positive tumor ratio was above or below 10% between methods (Fig. 3). Between microscope and virtual microscopy we found 20 discordant estimations, 18 between plain and assisted virtual microscopy and 18 between microscopy and assisted virtual microscopy. Irrespective of the methodology, the intraobserver discordance always ranged between 15% and 17% and could not be improved by drawing outlines.

An additional hint that positive tumor areas are difficult to circumscribe could be found by comparing the interobserver agreements of the manually marked tumor areas and positive tumor areas: tumor areas could be more reliably delineated (OCCC was 0.807) than positive tumor areas (OCCC was 0.704).

Intraobserver comparisons of staining intensities are shown in Table S2. Between microscopy and virtual microscopy we found 84 matches (70%) and 36 mismatches (30%), as much as between plain and assisted microscopy. Between microscopy and assisted virtual microscopy there were 87 matches (72.5%) and 33 mismatches (27.5%). In all comparisons, the kappa value was between 0.505 and 0.551 ($P < 0.001$), indicating a moderate agreement.

These data show that intraobserver agreement declines when intensity of immunostaining is added to the scoring system. A thorough analysis of the individual drawings (see for instance Fig. S2, cases 6 and 12) illustrates that each pathologist draws unique outlines of what he/she considers to be the strongest identifiable immunostaining intensity, which then affects the overall IRS.

Subgroup analysis

We divided the pathologists by their formal degree of experience into board-certified pathologists ($n = 6$) and residents ($n = 4$) to test whether experience has an influ-

ence on reproducibility. All calculations were repeated for both subgroups. Resulting data are shown in Tables S3 and S4. P -values of the tests remain significant. The measures of agreement of positive tumor ratio (i.e., mean of standard deviations, OCCC, and Pearson's correlation coefficient) vary between the subgroups, but we cannot identify a trend indicating an advantage for either subgroup. In contrast, inter- and intraobserver agreements of staining intensity show that board-certified pathologists can better reproduce staining intensities. Calculation of the interobserver agreements of the absolute tumor areas that were manually outlined in round 3 yielded an OCCC of 0.811 for the board-certified pathologists subgroup and 0.756 for the residents subgroup, respectively, showing a slight advantage for the board-certified pathologists. Agreement between visual estimation of area ratio of the manually drawn outlines and the corresponding computed values is nearly perfect (OCCC 0.966 and 0.983, respectively) for both subgroups. The number of specimens discordantly rated concerning the 10% cutoff could not be reduced using virtual microscopy in either group. The rate of false-negative ratings shows no clear advantage for either subgroup (Table S5).

Discussion

For many years, only the anatomical location of the primary tumor, its histological phenotype and the tumor stage tailored chemotherapy. However, in clinical practice many patients with a seemingly identical tumor responded differently to the same therapy. Research on cancer biology provided ample explanations [21]. Various genetic alterations and distinct molecular phenotypes were unraveled, which influence patient prognosis and response to chemotherapy. With the advent of targeted therapy, companion diagnostics is increasingly used to tailor patient treatment. Tumor-bearing tissue obtained prior to treatment is used to identify tumor-specific alterations (gene mutations, gene amplifications, protein expression-patterns), which predict therapeutic response. Testing the

Her2/neu status has long been used in breast cancer. More recently, the To A-study provided strong evidence that Her2/neu overexpressing GCs also respond to treatment with trastuzumab. However, the assessment of Her2/neu overexpression is far more complicated in GC compared with breast cancer. First, it necessitated the development of a novel scoring system, which is different from the breast cancer scoring [22]. GC cells more commonly harbor basolateral expression and rarely circumferential Her2/neu-staining [15, 22]. Second, expression of Her2/neu is heterogeneous bearing the risk of false-positive- and false-negative results due to sampling errors [14]. In this study we examined the third pitfall, that is, the correct assessment of the cut-off value (i.e., 10%).

Our results demonstrate that pathologists are well capable estimating area ratios, independent of their experience, but with a slight systematic overestimation. It is the gradient in immunostaining intensity (“gray scale pattern”) that introduces errors by making it difficult to reliably delineate the borders between IRS values (e.g., 2+ vs. 3+). This difficulty is illustrated in Figure S4. Cases showing a sharply delineated “black-and-white” staining pattern can be rated easier and more reliably. Reproducibility and error rates highly depend on the individual staining pattern of each case. The situation is aggravated by the often heterogeneous distribution of positive tumor clones in GC, which introduces additional complexity to the staining pattern. This results in false-negative ratings, preventing an ISH analysis that could possibly render the patient eligible for Trastuzumab therapy. In our analysis, this false-negative rate was up to 19% (Table 2). However, we had mainly selected cases known to be hovering around 10% for this study.

Knowledge of these sources of error is important because they may lead to the denial of Trastuzumab or, vice versa, to the prescription of medication without evidence of a treatment benefit. This will become even more important with the emergence of new targeted therapies using antibody-coupled drugs (e.g., Trastuzumab-Emtansine), which will need precise companion diagnostics with low false-negative and false-positive error rates. Future developments of immunohistochemistry-based companion diagnostics may wish to spend particular attention to heterogeneous biomarker expressions and the obstacles of a correct classification of percentage areas near a cut-off value.

How could the consequences of our observation be minimized? First, our trial to apply a computer-assisted method, namely manual outlining of the positive and total tumor areas within the virtual slide, did not help. A fully automated image analysis would presumably render higher precision and reproducibility, but not a more exact result as long as there are no well-defined on-slide refer-

ences for calibration of Her2/neu staining intensities. Second, while we found that the reproducibility of positive tumor ratio estimation does not profit from experience, the rating of staining intensities does. Aside the existing recommendation, this gives further evidence that Her2/neu status should be evaluated by board-certified pathologists who are trained in the method. Third, cut-off limits should be established very carefully henceforth. There should be evidence for its necessity, its value and the viability of its measurement. Fourth, in cases having a positive tumor ratio hovering around 10% there should be a statement in the pathologic report that there is a chance of misinterpretation of the staining results. Anyway, in any case of doubt we would recommend to apply an ISH test.

It should be discussed whether to alter the rules for Her2/neu assessment in GC as it has recently been done for breast cancer [23], where cases with a strong staining and less than 10% positive tumor ratio are newly assigned to ISH testing to reduce the false-negative rate. To our knowledge, the 10% cutoff is not validated for GC, but rather adopted from the Her2/neu scoring rules for breast cancer.

Having available an H&E stained slide as a reference to help identify tumor areas might be helpful. In this study, the participants did not use H&E slides. Nevertheless, the overall agreement on what was tumor tissue (OCCC 0.807) was better than agreement on positively stained tumor areas (OCCC 0.704), so the larger source of errors certainly comes from the detection of stained areas and not from identifying tumor areas.

Another topic deserves attention: The Her2/neu testing recommendations require an additional ISH testing only in cases of equivocal, that is, IHC 2+, staining intensity, to distinguish positive from negative cases. IHC 1+ always counts as negative and IHC 3+ always as positive, without confirmation by ISH testing. A recent prospective study [24] has systematically investigated the relation between HER2 gene amplification and response to therapy in 66 cases of advanced GC. Using ROC analysis, an optimal HER2/CEP17 ratio threshold of 4.7 could be determined to predict response. It could also be predicted whether 12-month and 16-month survival was reached, respectively. This indicates the advantages of ISH testing, which is probably underestimated in the current recommendations, and might be considered for every case bearing any IHC staining greater than 0.

In conclusion, heterogeneous expression and distribution of predictive biomarkers in tumor tissue poses two major problems. First, nonrepresentative biopsy sampling leads to false-positive or -negative test results and may be overcome by analyzing greater numbers of biopsies or by combining the Her2/neu analysis of biopsy and surgical

resection specimens [25]. Second, the subjective assessment of an immunoreactivity IRS. Positive tumor ratios strongly depend on what each individual pathologist considers as maximum immunostaining intensity coupled with a systematical overestimation of tumor areas. Future developments of immunohistochemistry-based companion diagnostics may wish to spend particular attention to heterogeneous biomarker expressions and the obstacles of a correct classification of percentage areas near a cut-off value.

Limitations

Our study could not assess the consequences for therapeutic outcome, because the patients of our cohort were only surgically resected and received no chemotherapy, neither adjuvant nor neoadjuvant.

Conflict of Interest

None declared.

References

- Bang, Y. J., C. E. Van, A. Feyereislova, H. C. Chung, L. Shen, A. Sawaki, et al. 2010. Trastuzumab in combination with chemotherapy versus chemotherapy alone for treatment of HER2-positive advanced gastric or gastro-oesophageal junction cancer (ToGA): a phase 3, open-label, randomised controlled trial. *Lancet* 376:687–697.
- Chua, T. C., and N. D. Merrett. 2012. Clinicopathologic factors associated with HER2-positive gastric cancer and its impact on survival outcomes—a systematic review. *Int. J. Cancer* 130:2845–2856.
- Bilous, M., R. Y. Osamura, J. Ruschoff, M. van de Viver, W. HannaF. Penault-Llorca, et al. 2010. HER-2 amplification is highly homogenous in gastric cancer. *Hum. Pathol.* 41:304–305.
- Bozzetti, C., F. V. Negri, C. A. Lagrasta, P. Crafa, C. Bassano, I. Tamagnini, et al. 2011. Comparison of HER2 status in primary and paired metastatic sites of gastric carcinoma. *Br. J. Cancer* 104:1372–1376.
- Fusco, N., E. G. Rocco, C. C. Del, C. Pellegrini, G. Bulfamante, N. F. Di, et al. 2013. HER2 in gastric cancer: a digital image analysis in pre-neoplastic, primary and metastatic lesions. *Mod. Pathol.* 26:816–824.
- Kimura, M., H. Tsuda, D. Morita, E. Shinto, T. Tanimoto, T. Ichikura, et al. 2005. Usefulness and limitation of multiple endoscopic biopsy sampling for epidermal growth factor receptor and c-erbB-2 testing in patients with gastric adenocarcinoma. *Jpn. J. Clin. Oncol.* 35:324–331.
- Langer, R., S. Rauser, M. Feith, J. M. Nahrig, A. Feuchtinger, H. Friess, et al. 2011. Assessment of ErbB2 (Her2) in oesophageal adenocarcinomas: summary of a revised immunohistochemical evaluation system, bright field double in situ hybridisation and fluorescence in situ hybridisation. *Mod. Pathol.* 24:908–916.
- Abraham, S. C., S. J. Park, J. H. Lee, L. Mugartegui, and T. T. Wu. 2003. Genetic alterations in gastric adenomas of intestinal and foveolar phenotypes. *Mod. Pathol.* 16:786–795.
- Kim, M. A., H. J. Lee, H. K. Yang, Y. J. Bang, and W. H. Kim. 2011. Heterogeneous amplification of ERBB2 in primary lesions is responsible for the discordant ERBB2 status of primary and metastatic lesions in gastric carcinoma. *Histopathology* 59:822–831.
- Lee, S., W. B. de Boer, S. Fermoye, M. Platten, and M. P. Kumarasinghe. 2011. Human epidermal growth factor receptor 2 testing in gastric carcinoma: issues related to heterogeneity in biopsies and resections. *Histopathology* 59:832–840.
- Tafe, L. J., Y. Y. Janjigian, M. Zaidinski, C. V. Hedvat, M. R. Hameed, L. H. Tang, et al. 2011. Human epidermal growth factor receptor 2 testing in gastroesophageal cancer: correlation between immunohistochemistry and fluorescence in situ hybridization. *Arch. Pathol. Lab. Med.* 135:1460–1465.
- Yan, B., E. X. Yau, S. S. Bte Omar, C. W. Ong, B. Pang, K. G. Yeoh, et al. 2010. A study of HER2 gene amplification and protein expression in gastric cancer. *J. Clin. Pathol.* 63:839–842.
- Yan, B., E. X. Yau, S. N. Choo, C. W. Ong, K. J. Yong, B. Pang, et al. 2011. Dual-colour HER2/chromosome 17 chromogenic in situ hybridisation assay enables accurate assessment of HER2 genomic status in gastric cancer and has potential utility in HER2 testing of biopsy samples. *J. Clin. Pathol.* 64:880–883.
- Warneke, V. S., H. M. Behrens, C. Boger, T. Becker, F. Lordick, M. P. Ebert, et al. 2013. Her2/neu testing in gastric cancer: evaluating the risk of sampling errors. *Ann. Oncol.* 24:725–733.
- Rüschhoff, J., M. Dietel, G. Baretton, S. Arbogast, A. Walch, G. Monges, et al. 2010. HER2 diagnostics in gastric cancer—guideline validation and development of standardized immunohistochemical testing. *Virchows Arch.* 457:299–307.
- Landis, J. R., and G. G. Koch. 1977. The measurement of observer agreement for categorical data. *Biometrics* 33:159–174.
- Fleiss, J. L., B. Levin, and M. C. Paik. 2003. *Statistical methods for rates and proportions*. 3rd ed. John Wiley & Sons, New York, NY.
- Gamer, M., J. Lemon, I. Fellows, and P. Singhirr. 2012. *irr: various coefficients of interrater reliability and agreement*. R package, version 0.84. [computer program].
- Barnhart, H. X., M. Haber, and J. Song. 2002. Overall concordance correlation coefficient for evaluating

- agreement among multiple observers. *Biometrics* 58:1020–1027.
20. Stevenson, M., T. Nunes, C. Heuer, J. Marshall, J. Sanchez, R. Thornton, et al. 2013. *epiR*: An R package for the analysis of epidemiological data. R package, version 0.9-48 [computer program].
 21. Hanahan, D., and R. A. Weinberg. 2011. Hallmarks of cancer: the next generation. *Cell* 144:646–674.
 22. Hofmann, M., O. Stoss, D. Shi, R. Buttner, V. van de Vijver, W. Kim, et al. 2008. Assessment of a HER2 scoring system for gastric cancer: results from a validation study. *Histopathology* 52:797–805.
 23. Wolff, A. C., M. E. Hammond, D. G. Hicks, M. Dowsett, L. M. McShane, K. H. Allison, et al. 2013. Recommendations for human epidermal growth factor receptor 2 testing in breast cancer: American Society of Clinical Oncology/College of American Pathologists clinical practice guideline update. *J. Clin. Oncol.* 31:3997–4013.
 24. Gomez-Martin, C., J. C. Plaza, R. Pazo-Cid, A. Salud, F. Pons, P. Fonseca, et al. 2013. Level of HER2 gene amplification predicts response and overall survival in HER2-positive advanced gastric cancer treated with trastuzumab. *J. Clin. Oncol.* 31:4445–4452.
 25. Watson, S., P. Validire, P. Cervera, N. Zorkani, A. Scriva, F. Lemay, et al. 2013. Combined HER2 analysis of biopsies and surgical specimens to optimize detection of trastuzumab-eligible patients in eso-gastric adenocarcinoma: a GERCOR study. *Ann. Oncol.* 24:3035–3039.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Figure S1. (A) Screenshot of our virtual microscopy program in outline-tracing mode. The pathologists manually marked tumor tissue (blue) and positive tumor tissue (red). (B) To test the ability of visually estimating area ratios, the pathologists were shown color-filled sketches of the regions they had outlined. These were only slightly tinted to avoid effects of color contrast or color perception. In a routine application one would not visually estimate the positive tumor ratios from drawn outlines because they could as well be computed; this only served to get an insight into the process of scoring.

Figure S2. Outlines manually drawn by the pathologists around areas of all tumor tissue (blue) and positive tumor tissue (red). Numbers below images denote the staining intensities from rating by microscope, by virtual

microscopy, and by assisted virtual microscopy, respectively. Below are given the corresponding positive tumor ratios, plus the ratio that was calculated from the drawings. Continued on second page.

Figure S3. Areas of tumor tissue (blue) and positive tumor tissue (red) were calculated from the outline drawings, both given in square millimeters in the bar diagrams. The third row shows bar diagrams of the resulting, calculated positive tumor ratio. For direct comparison, in the bottom row the visually estimated positive tumor ratio from Figure 1 is repeated.

Figure S4. Illustration of 15% variability of intensity on a continuous gray scale at different levels. Each box spans 15% width of the gray scale. Viewing one box alone, the variation within is hardly noticeable.

Table S1. Clinical characteristics of the 12 cases of gastric cancer. UICC data (T, N, M, Stage) were recorded to seventh edition if only older data were available.

Table S2. Comparison of staining intensity ratings between methods. Numbers in the cross tables indicate the count of staining intensity combinations from two methods. Numbers on the diagonal (green cells) indicate agreement, and off-diagonal numbers indicate disagreement. (A) Agreement of staining intensity ratings between microscope and virtual microscopy. (B) Agreement of staining intensity ratings between virtual microscopy and assisted virtual microscopy. (C) Agreement of staining intensity ratings between microscope and assisted virtual microscopy.

Table S3. Subgroup analysis of interobserver agreements between pathologists using one method of rating, respectively. Positive tumor ratios were compared using the mean of standard deviations, OCCC*), and number of discordant ratings with respect to the 10% cutoff. Staining intensities were compared using Fleiss' kappa.

Table S4. Subgroup analysis of intraobserver agreements, comparing two rating methods, respectively. Agreements of positive tumor ratio were compared using OCCC, Pearson's correlation coefficient, and the number of discordant ratings with respect to the 10% cutoff. Agreements of staining intensities were calculated as kappa values.

Table S5. Subgroup analysis of the numbers of false-negative ratings and cases, respectively, due to underestimation of positive tumor area and/or staining intensity. Numbers in parentheses denote the false-negative percentage related to the total number of ratings that were done in each subgroup on the nine SISH-positive cases (72 ratings by six board-certified pathologists, and 48 ratings by four residents.)

Warneke, V. S., Behrens, H. M., Haag, J., Balschun, K., Böger, C., Becker, T., Ebert M. P. A., Lordick F., & Röcken, C. (2013). Prognostic and putative predictive biomarkers of gastric cancer for personalized medicine. *Diagnostic Molecular Pathology* 2013;22(3):127-137.
[doi:10.1097/PDM.0b013e318284188e](https://doi.org/10.1097/PDM.0b013e318284188e)

Lebenslauf

Mein Lebenslauf wird aus datenschutzrechtlichen Gründen in der elektronischen Version meiner Arbeit nicht veröffentlicht.

Komplette Publikationsliste

Originalpublikationen

1. Behrens HM, Warneke VS, Böger C, Garbrecht N, Jüttner E, Klapper W, Mathiak M, Oschlies I, Rudolph U, Stuhlmann-Laeisz C, Trick D, Röcken C, Hufnagl P. Reproducibility of Her2/neu scoring in gastric cancer and assessment of the 10% cut-off rule. *Cancer Med* 2014. doi:10.1002/cam4.365
2. Böger C, Warneke VS, Behrens HM, Kalthoff H, Goodman SL, Becker T, Röcken C. Integrins $\alpha\beta3$ and $\alpha\beta5$ as prognostic, diagnostic, and therapeutic targets in gastric cancer. *Gastric Cancer* 2014. doi:10.1007/s10120-014-0435-2
3. Böger C, Kalthoff H, Goodman SL, Behrens HM, Röcken C. Integrins and their ligands are expressed in non-small cell lung cancer but not correlated with parameters of disease progression. *Virchows Arch* 2014;464(1):69-78. doi:10.1007/s00428-013-1506-1
4. Egberts F, Krüger S, Behrens HM, Bergner I, Papaspyrou G, Werner JA, Alkatout I, Haag J, Hauschild A, Röcken C. Melanomas of unknown primary frequently harbor TERT-promoter mutations. *Melanoma Res* 2014;24(2):131-6. doi:10.1097/CMR.0000000000000048
5. Egberts F, Bergner I, Krüger S, Haag J, Behrens HM, Hauschild A, Röcken C. Metastatic melanoma of unknown primary resembles the genotype of cutaneous melanomas. *Ann Oncol* 2014 ;25(1):246-50. doi:10.1093/annonc/mdt411
6. Hess K, Böger C, Behrens HM, Röcken C. Correlation between the expression of integrins in prostate cancer and clinical outcome in 1284 patients. *Ann Diagn Pathol* 2014;18(6):343-50. doi:10.1016/j.anndiagpath.2014.09.001
7. Ingold HB, Behrens HM, Balschun K, Haag J, Krüger S, Becker T, Röcken C. HER2/neu testing in primary colorectal carcinoma. *Br J Cancer* 2014;111(10):1977-84. doi:10.1038/bjc.2014.483
8. Warneke VS, Behrens HM, Haag J, Krüger S, Simon E, Mathiak M, Ebert MP, Röcken C. Members of the EpCAM signalling pathway are expressed in gastric cancer tissue and are correlated with patient prognosis. *Br J Cancer* 2013;109(8):2217-27. doi:10.1038/bjc.2013.536
9. Warneke VS, Behrens HM, Haag J, Balschun K, Böger C, Becker T, Ebert MP, Lordick F, Röcken C. Prognostic and putative predictive biomarkers of gastric cancer for personalized medicine. *Diagn Mol Pathol* 2013;22(3):127-37. doi:10.1097/PDM.0b013e318284188e
10. Warneke VS, Behrens HM, Böger C, Becker T, Lordick F, Ebert MP, Röcken C. Her2/neu testing in gastric cancer: evaluating the risk of sampling errors. *Ann Oncol* 2013;24(3):725-33. doi:10.1093/annonc/mds528
11. Simon E, Petke D, Böger C, Behrens HM, Warneke V, Ebert M, Röcken C. The spatial distribution of LGR5+ cells correlates with gastric cancer progression. *PLoS One* 2012;7(4):e35486. doi: 10.1371/journal.pone.0035486
12. Warneke VS, Behrens HM, Hartmann JT, Held H, Becker T, Schwarz NT, Röcken C. Cohort study based on the seventh edition of the TNM classification for gastric cancer: proposal of a new staging system. *J Clin Oncol* 2011;29(17):2364-71. doi:10.1200/JCO.2010.34.4358

13. Schmuck R, Warneke V, Behrens HM, Simon E, Weichert W, Röcken C. Genotypic and phenotypic characterization of side population of gastric cancer cell lines. *Am J Pathol* 2011;178(4):1792-804. doi:10.1016/j.ajpath.2010.12.043
14. Markert A, Thierry V, Kleber M, Behrens M, Engelhardt M. Chemotherapy safety and severe adverse events in cancer patients: strategies to efficiently avoid chemotherapy errors in in- and outpatient treatment. *Int J Cancer* 2009;124(3):722-8. doi:10.1002/ijc.23991
15. Behrens M, Fröchtenicht R, Hartmann M, Siebers JG, Buck U, Hagemeister FC. Vibrational spectroscopy of methanol and acetonitrile clusters in cold helium droplets. *The Journal of Chemical Physics* 1999;111(6):2436-43.
16. Behrens M, Buck U, Fröchtenicht R, Hartmann M, Huisken F, Rohmund F. Rotationally resolved IR spectroscopy of ammonia trapped in cold helium clusters. *The Journal of Chemical Physics* 1998;109(14):5914-20.
17. Behrens M, Buck U, Fröchtenicht R, Hartmann M, Havenith M. The ammonia dimer spectrum in cold helium clusters. *The Journal of Chemical Physics* 1997;107(18):7179-86.

Posterbeiträge

1. Behrens HM, Warneke V, Böger C, Garbrecht N, Jüttner E, Klapper W, Mathiak M, Oshlies I, Rudolph U, Stuhlmann-Laeisz C, Trick D, Hufnagl P, Röcken C. Comparison of Her2/neu scoring between microscopic and computer assisted methods assessing the 10 percent cut-off rule in gastric cancer. *Der Pathologe* 2014;35(Supplement 1):93. doi:10.1007/s00292-014-1944-9
2. Böger C, Warneke VS, Behrens HM, Röcken C. Putative tumor biological and prognostic value of integrins in gastric cancer. *Der Pathologe* 2014;35(Supplement 1):27. doi:10.1007/s00292-014-1944-9
3. Mathiak M, Warneke V, Behrens HM, Haag J, Krüger S, Röcken C. Microsatellite instability is uncommon in gastric cancer and correlates with patient survival. *Der Pathologe* 2014;35(Supplement 1):92. doi:10.1007/s00292-014-1944-9
4. Warneke V, Behrens HM, Haag J, Balschun K, Böger C, Becker T, Hartmann J, Ebert M, Röcken C. Prognostic biomarkers of gastric cancer. *Der Pathologe* 2012;33(Supplement 1):54. doi:10.1007/s00292-012-1584-x
5. Haag J, Ingold-Heppner B, Behrens HM, Aigner T, Röcken C. FoxO6 expression in gastric cancer. *Der Pathologe* 2012;33(Supplement 1):99. doi:10.1007/s00292-012-1584-x
6. Böger C, Warneke V, Behrens HM, Röcken C. Her2/neu in gastric cancer: comparison of tissue micro arrays with corresponding whole tissue sections. *Der Pathologe* 2012;33(Supplement 1):104. doi:10.1007/s00292-012-1584-x
7. Simon E, Petke D, Böger C, Warneke V, Behrens HM, Röcken C. LGR5 is differentially expressed in hepato-gastrointestinal tumors. *Der Pathologe* 2012;33(Supplement 1):118. doi:10.1007/s00292-012-1584-x

Danksagung

Ganz wichtig für das Entstehen dieser Dissertation waren meine Betreuer und Koautoren, für deren Unterstützung, Rat, Mitarbeit und Geduld ich meinen Dank ausdrücken möchte:

Zuerst danke ich Herrn Professor Dr. med. Manfred Dietel dafür, dass er mir die Möglichkeit zur Promotion am Institut für Pathologie der Charité gegeben hat.

Herrn Professor Dr. rer. nat. Peter Hufnagl danke ich für die Übernahme der direkten Betreuung und die freundliche Aufnahme in seine Arbeitsgruppe. Er unterstützte mich durch Rat, Anregungen und Ermutigung bei der Versuchsplanung und den Auswertungen.

Herrn Professor Dr. med. Christoph Röcken danke ich dafür, dass er den Anstoß zu diesem Promotionsvorhaben gab, es durch die Bereitstellung der Ressourcen überhaupt erst ermöglicht hat, und mich an die Fragestellungen zum Magenkarzinom herangeführt hat. In unseren Diskussionen habe ich viel von ihm über die pathologisch-klinische Forschung gelernt. Bei Fragen konnte ich ihn jederzeit um Rat bitten und habe immer klare und hilfreiche Antworten bekommen. In schwierigen Situationen lotste er das Projekt wieder in Richtung Ziel. Dafür ganz herzlichen Dank!

Den Koautoren der Publikationen danke ich für die angenehme Zusammenarbeit und insbesondere für ihre Ausdauer bei der Mitarbeit am Her2/neu-Auswerterversuch. An Frau Dr. Viktoria Warneke geht mein Dank für die gute Zusammenarbeit bei der geteilten Erstautorenschaft der ersten Publikation.

Außerdem gilt mein Dank den Kolleginnen und Kollegen in der Arbeitsgruppe und im Institut für ihre Unterstützung und die angenehme Arbeitsumgebung. Beispielhaft seien genannt: Frau Sandra Krüger u.a. für die zusätzlichen HER2-CISH-Färbungen, Frau Dr. Christine Böger für die Auswertung weiterer Immunfärbungen und Frau Dr. Franziska Wilhelm für die kritische Durchsicht des Manuskriptes zum Her2-Auswerterversuch.

Meiner Lebensgefährtin Dr. Eva Jüttner danke ich für ihre Geduld und Unterstützung.