



**TECHNISCHE UNIVERSITÄT MÜNCHEN**

Institut für Allgemeine Pathologie und  
Pathologische Anatomie

**Molecular subtypes based on *The Cancer Genome Atlas* (TCGA)  
classification in gastric carcinoma:  
Prognostic and therapeutic implications**

Meike Anna-Lena Kohlruß

Vollständiger Abdruck der von der Fakultät für Medizin der  
Technischen Universität München zur Erlangung des akademischen Grades eines

**Doktors der Naturwissenschaften**

**(Dr. rer. nat.)**

genehmigten Dissertation.

Vorsitzender: Prof. Dr. Markus Gerhard

Prüfende/-r der Dissertation:

1. apl. Prof. Dr. Gisela Keller

2. Prof. Dr. Michael Groll

3. Prof. Dr. Jens H. L. Neumann

Die Dissertation wurde am 03.02.2020 bei der Technischen Universität München eingereicht und durch die Fakultät für Medizin am 14.07.2020 angenommen.



# TABLE OF CONTENTS

<b>LIST OF FIGURES .....</b>	<b>I</b>
<b>LIST OF TABLES .....</b>	<b>III</b>
<b>ABBREVIATIONS .....</b>	<b>V</b>
<b>ABSTRACT .....</b>	<b>1</b>
<b>ZUSAMMENFASSUNG .....</b>	<b>3</b>
<b>1. INTRODUCTION.....</b>	<b>5</b>
1.1. Gastric cancer .....	5
1.1.1. Anatomy and histology of the stomach.....	5
1.1.2. Epidemiology of stomach cancer.....	6
1.1.3. Etiology and risk factors .....	7
1.1.4. Hereditary syndromes of gastric carcinoma.....	8
1.1.5. Staging and histological classification.....	9
1.1.6. Diagnostic and multimodal treatment.....	10
1.2. Molecular classification systems of gastric carcinomas.....	12
1.2.1. TCGA classification and characteristics of molecular subgroups .....	12
1.2.2. ACRG classification .....	13
1.3. Genomic instability in gastric carcinomas.....	15
1.3.1. Microsatellite instability .....	15
1.3.2. Chromosomal instability.....	16
1.4. Mutational landscape and signaling pathways in gastric cancer .....	17
1.5. Objectives and study design .....	20
<b>2. MATERIAL AND METHODS.....</b>	<b>22</b>
2.1. Patients and tumor material .....	22
2.1.1. Patients.....	22
2.1.2. Neoadjuvant chemotherapy and surgery.....	23
2.1.3. Response evaluation.....	23
2.1.4. Tumor cell plasticity .....	24
2.1.5. Overall survival and follow-up time .....	24
2.2. Materials .....	24
2.2.1. Primer and microsatellite marker.....	24
2.2.2. Commercial kits.....	26

2.2.3. Instruments.....	27
2.2.4. Chemicals and reagents.....	28
2.2.5. Buffers and solutions .....	28
2.2.6. Consumption materials .....	29
2.2.7. Software and programs .....	30
2.2.8. Online platforms .....	30
2.3. Methods .....	31
2.3.1. DNA isolation and quantification from formalin-fixed paraffin-embedded tissues ....	31
2.3.2. Methods for determination of molecular classification .....	33
2.3.3. Genome-wide OncoScan analysis and CIN classification .....	41
2.3.4. Assessment of intra-tumor heterogeneity .....	44
2.3.5. Next-generation sequencing.....	44
2.3.6. Statistical analysis.....	52
<b>3. RESULTS.....</b>	<b>54</b>
3.1. Study enrolment and patient characteristics .....	54
3.1.1. Study enrolment.....	54
3.1.2. Patient characteristics.....	55
3.1.3. Response to neoadjuvant CTx .....	57
3.2. Microsatellite analysis and CIN classification .....	57
3.2.1. Comparison of multiplex PCR assays with genome-wide method.....	58
3.2.2. Determination of CIN according to definition of TCGA based on OncoScan and microsatellite based multiplex PCR analysis.....	61
3.2.3. Microsatellite based CIN classification and tumor heterogeneity .....	63
3.2.4. Limit of detection of microsatellite based multiplex PCR assays.....	64
3.3. Analysis of EBV, MSI and AI.....	65
3.3.1. Results of EBV and MSI analysis.....	65
3.3.2. Results of AI analysis .....	68
3.4. Classification of CIN .....	78
3.4.1. CIN classification according to TCGA.....	79
3.4.2. Modified CIN classification according to quartiles of AI ratios.....	83
3.4.3. Restriction of the number of analyzed microsatellite markers and CIN classification	95
3.4.4. CIN classification of paired tumor biopsies and resected tumors.....	97
3.5. Original and modified TCGA classification.....	99
3.5.1. Molecular classification algorithms and frequency of molecular subgroups.....	99
3.5.2. Molecular classification and association with patient characteristics.....	101

3.5.3. Molecular classification and association with response to CTx .....	104
3.5.4. Molecular classification and patients survival .....	105
3.6. Mutation profiling by next-generation sequencing .....	113
3.6.1. Frequency of tumors with alterations in GC related genes .....	113
3.6.2. <i>TP53</i> mutations .....	114
3.6.3. <i>CDHI</i> and <i>RHOA</i> mutations .....	115
3.6.4. Alterations in GC related genes and association with the MSI-L subtype.....	116
3.6.5. Alterations in GC related genes and association with the CIN-status .....	117
3.6.6. Prediction of copy number variants .....	119
3.7. Main characteristics of the molecular subgroups .....	120
<b>4. DISCUSSION .....</b>	<b>121</b>
4.1. Identification of novel molecular subgroups in gastric cancer.....	121
4.2. Determination of AI and CIN .....	125
4.3. Mutational landscape of gastric cancer .....	129
4.4. Limitations of the study and perspectives .....	131
<b>5. CONCLUSION.....</b>	<b>132</b>
<b>6. APPENDIX .....</b>	<b>133</b>
<b>7. BIBLIOGRAPHY .....</b>	<b>153</b>
<b>8. PUBLICATIONS AND CONGRESS CONTRIBUTIONS.....</b>	<b>167</b>
8.1. Publications .....	167
8.2. Congress contributions .....	167
<b>9. ACKNOWLEDGEMENT .....</b>	<b>168</b>

# LIST OF FIGURES

Figure 1: Anatomy of the stomach and layers of the stomach wall .....	5
Figure 2: Incidence and mortality of the most common cancers worldwide for both sexes and all ages.....	6
Figure 3: Pathogenesis of intestinal gastric carcinoma including environmental and genetic factors .....	8
Figure 4: Characteristics of the four gastric cancer subtypes according to the TCGA classification system.....	13
Figure 5: Molecular classification algorithms according to TCGA and ACRG .....	14
Figure 6: Examples of various karyotypes as results of numerical and structural CIN.....	17
Figure 7: Study design and objectives of the study.....	21
Figure 8: Workflow for the detection of AI and CIN using microsatellite based multiplex PCR assays.....	35
Figure 9: Calculation of allele ratios of a heterozygous marker.....	38
Figure 10: Dilution experiments to assess the limitation of AI detection by the multiplex PCR assays.....	39
Figure 11: Molecular inversion probe technology of the Affymetrix OncoScan® FFPE Assay .....	41
Figure 12: Analysis of copy number variants and AI using Nexus Express software.....	42
Figure 13: Determination of B-allele frequency .....	43
Figure 14: Next-generation sequencing workflow using the Ion Torrent technology.....	47
Figure 15: Preparation of DNA libraries using the multiplex PCR based Ion AmpliSeq™ targeted sequencing technology.....	49
Figure 16: Flow chart diagram of patient’s inclusion for the analysis and evaluation of the molecular classification .....	55
Figure 17: Response to neoadjuvant CTx and association with overall survival in the pre-therapeutic tumor biopsy cohort .....	57
Figure 18: Examples of tumors which are negative or positive for AI in the OncoSan and multiplex PCR assays .....	59
Figure 19: Overview of the ratios of alterations in the 30 tumors detected with the OncoScan assay compared to the multiplex PCR assays.....	60
Figure 20: Correlation of the ratios of chromosomal alterations detected with the OncoScan assay compared to the AI ratios determined with the multiplex PCR assays.....	61
Figure 21: Genome-wide OncoScan analysis of 30 tumors and OncoScan based CIN classification .....	61
Figure 22: Microsatellite based CIN classification and tumor heterogeneity .....	63
Figure 23: Results of dilution experiments to determine the detection limit of the multiplex PCR assays.....	64
Figure 24: Correlation of tumor cell contents and AI ratios of all resected tumors analyzed for AI and CIN.....	65
Figure 25: Detection of an EBV-associated tumor using standardized methods .....	66
Figure 26: Examples of microsatellite patterns using a multiplex PCR assay with two mononucleotide and three dinucleotide markers .....	66
Figure 27: EBV and MSI classification of patients with tumor biopsies before CTx and resected tumors without or after CTx.....	67
Figure 28: Frequencies of instability at five microsatellite markers included for MSI analysis. ....	67
Figure 29: Examples of microsatellite patterns of tumors with AI at a respective microsatellite locus compared to the corresponding non-tumorous tissues .....	70
Figure 30: Frequency of AI in the tumor biopsy and resected tumor cohort .....	71
Figure 31: Frequency of AI in the resected tumors without and after neoadjuvant CTx .....	72
Figure 32: Frequency of AI in paired tumor biopsies before and resected tumors after CTx .....	72
Figure 33: Frequency of AI in the resected tumor cohort and association with tumor localization and Laurén histological subtypes .....	74

Figure 34: Frequency of AI and association with response to neoadjuvant CTx.....	75
Figure 35: Survival and AI-status of the tumor biopsy cohort at respective chromosomal regions.....	76
Figure 36: Survival and AI-status of the resected tumor cohort at respective chromosomal regions .....	77
Figure 37: Survival and AI-status of the non-CTx tumor cohort at respective chromosomal regions .....	77
Figure 38: Distribution of AI ratios and CIN classification according to TCGA.....	79
Figure 39: CIN according to TCGA and significant associations with clinical parameters.....	81
Figure 40: CIN according to TCGA and association with response to neoadjuvant CTx.....	81
Figure 41: CIN classification according to TCGA and survival of the patients.....	82
Figure 42: Distribution of AI ratios and modified CIN classification according to the quartiles of AI ratios .....	84
Figure 43: Modified CIN classification and significant associations with clinical parameters.....	86
Figure 44: Modified CIN classification according to the quartiles of AI ratios and survival of the patients .....	87
Figure 45: CIN-status in two subgroups according to the quartiles of AI ratios and survival of the patients .....	88
Figure 46: High CIN-M subgroup and significant associations with clinical parameters.....	91
Figure 47: Modified CIN-subgroups and frequency of AI at single microsatellite loci.....	92
Figure 48: Modified CIN classification according to the quartiles of AI ratios and association with response to neoadjuvant CTx.....	93
Figure 49: Modified CIN-subgroups and association with response to neoadjuvant CTx.....	93
Figure 50: CIN-H subgroup and survival of patients with tumor biopsies before CTx .....	94
Figure 51: CIN-subgroups defined by a restricted number of microsatellite markers and association with response to neoadjuvant CTx.....	95
Figure 52: Overall survival discriminated by the CIN-status in four subgroups according to chromosomal regions covering common fragile sites .....	96
Figure 53: Overall survival discriminated by the CIN-status in four subgroups according to chromosomal regions covering common tumor suppressor genes .....	96
Figure 54: Alterations in the four modified CIN-subgroups between corresponding tumor biopsies before CTx and resected tumors after CTx .....	97
Figure 55: Alterations in the four modified CIN-subgroups between pre-therapeutic tumor biopsies and corresponding post-therapeutic resected tumors .....	98
Figure 56: Molecular classification algorithm according to TCGA and frequency of molecular subgroups.....	99
Figure 57: Modified molecular classification system and frequency of molecular subgroups .....	100
Figure 58: Molecular classification systems and association with Laurén histological subtypes and tumor localization.....	103
Figure 59: Molecular classification according to TCGA and association with response to neoadjuvant CTx ...	104
Figure 60: Modified molecular classification system based on TCGA with high CIN-M or CIN-H as own subgroup and association with response to neoadjuvant CTx.....	105
Figure 61: Molecular classification according to TCGA and survival of the patients .....	106
Figure 62: Modified molecular classification based on TCGA and survival of the patients.....	109
Figure 63: Frequency of gastric carcinomas with mutations in selected gastric cancer related genes .....	113
Figure 64: Identified TP53 mutations in gastric carcinomas.....	114
Figure 65: Identified CDH1 mutations in gastric carcinomas.....	115
Figure 66: Mutation profile of gastric carcinomas and association with the MSI-L subtype.....	116
Figure 67: Mutation profile of gastric carcinomas and association with the CIN-status .....	117
Figure 68: Occurrence of TP53 and CDH1 mutations in the four CIN-subgroups .....	118
Figure 69: Clustering of GC related genes according to their function in cellular processes during carcinogenesis .....	118

Figure 70: Clustering of mutated gastric cancer related genes and association with CIN .....	119
Figure 71: Main characteristics of the molecular subgroups.....	120

## LIST OF TABLES

Table 1: Number of analyzed patients and respective TRG classification .....	23
Table 2: Primer pairs for EBV analysis .....	24
Table 3: Primer pairs for MSI analysis .....	25
Table 4: Primer mix for MSI analysis.....	25
Table 5: Primer pairs for preparation of length marker for agarose gel electrophoresis .....	25
Table 6: PCR primer mix for the preparation of length marker for agarose gel electrophoresis.....	25
Table 7: Microsatellite markers included in multiplex PCR assays for the detection of CIN.....	26
Table 8: PCR reaction mix for detection of EBV positivity .....	33
Table 9: PCR reaction mix for MSI analysis .....	34
Table 10: Composition of the multiplex PCR assays for the detection of AI and CIN.....	36
Table 11: Marker specific threshold values for the definition of AI.....	37
Table 12: Reaction mix for microsatellite based multiplex PCRs for AI and CIN analysis .....	37
Table 13: PCR reaction mix for the preparation of length marker for agarose gel electrophoresis .....	40
Table 14: Cycling protocol for the preparation of length marker for agarose gel electrophoresis.....	40
Table 15: Gastric cancer related genes included in the GC-panel.....	45
Table 16: PCR reaction mix for quantification of amplifiable DNA by qPCR.....	48
Table 17: PCR reaction mix for DNA library preparation per primer pool .....	49
Table 18: PCR reaction mix for quantification of DNA libraries by qPCR.....	51
Table 19: Clinical-pathological characteristics of the patient cohorts.....	56
Table 20: Types of discrepancy between the multiplex PCR assays and OncoScan analysis.....	60
Table 21: Classification of CIN according to the definition of TCGA based on results of the OncoScan and microsatellite based multiplex PCR analysis .....	62
Table 22: Patient characteristics for AI and CIN analysis .....	69
Table 23: Frequency of AI in the tumor biopsy and resected tumor cohort.....	71
Table 24: Comparison of the frequencies of AI in 38 paired tumor biopsies before and resected tumors after CTx .....	73
Table 25: AI at a single chromosomal locus and association with patient characteristics of the resected tumor cohort .....	74
Table 26: Significant pre-and post-therapeutically available factors for survival in association with the AI-status in multivariable analysis in the non-CTx patient group .....	78
Table 27: CIN according to TCGA and association with patient characteristics .....	80
Table 28: Cut-off values for CIN classification in four subgroups according to the quartiles of AI ratios.....	83
Table 29: Modified CIN classification and association with patient characteristics .....	85
Table 30: Survival data of the patient cohorts and subgroups in association with high CIN-M versus the remaining CIN-groups .....	89
Table 31: High CIN-M subgroup and association with patient characteristics .....	90
Table 32: Modified CIN classification according to quartiles of the AI ratios and association with response to neoadjuvant CTx.....	94

Table 33: Molecular classification according to TCGA and association with patient characteristics.....	101
Table 34: Modified molecular classification based on TCGA and association with patient characteristics .....	102
Table 35: Survival data of the different patient cohorts in association with the molecular classification system according to TCGA.....	107
Table 36: Survival data of the different patient cohorts in association with the modified molecular classification system based on TCGA.....	110
Table 37: Multivariable analysis of survival including pre- and post-therapeutically available clinical factors and modified molecular classification system based on TCGA in the resected non-CTx cohort .....	111
Table 38: Multivariable analysis of survival including pre- and post-therapeutically available clinical factors and modified molecular classification system based on TCGA in the resected CTx cohort .....	112

## ABBREVIATIONS

### A

<b>A230</b>	absorption at 230 nm
<b>A260</b>	absorption at 260 nm
<b>A280</b>	absorption at 280 nm
<b>ACRG</b>	Asian Cancer Research Group
<b>AEG</b>	adenocarcinoma of the esophageal junction
<b>AI</b>	allelic imbalance
<b>APC</b>	APC regulator of WNT signaling pathway
<b>ARID1A</b>	AT-rich interaction domain 1A

### B

<b>BAF</b>	B-allele frequency
<b>BCOR</b>	BCL6 Corepressor
<b>bp</b>	base pair

### C

<b>CagA</b>	cytotoxin-associated gene A
<b>Cap</b>	capecitabine
<b>CCNE1</b>	cyclin E1
<b>CD274</b>	CD274 molecule
<b>CDH1</b>	cadherin 1
<b>CDK</b>	cyclin dependent kinase
<b>CDKN2A</b>	cyclin dependent kinase inhibitor 2A
<b>CDS</b>	coding DNA sequence
<b>CDX2</b>	caudal type homeobox 2
<b>CGH</b>	comparative genomic hybridization
<b>Chr</b>	chromosome
<b>CI</b>	confidence interval
<b>CIMP</b>	CpG island methylator phenotype
<b>CIN-H</b>	high chromosomal instability
<b>CIN-L</b>	low chromosomal instability
<b>CIN-M</b>	medium chromosomal instability
<b>Cis</b>	cisplatin
<b>CN</b>	copy number
<b>CN-LOH</b>	copy neutral loss of heterozygosity
<b>CNV</b>	copy number variation
<b>COSMIC</b>	catalog of somatic mutations in cancer
<b>cT</b>	clinical tumor stage
<b>CTx</b>	chemotherapy

### D

<b>DDI</b>	DNA degradation index
<b>Del.</b>	deletion
<b>dNTP</b>	deoxyribonucleoside triphosphate
<b>Doc</b>	docetaxel
<b>dsDNA</b>	double-stranded desoxyribonucleic acid

### E

<b>EAC</b>	esophageal adenocarcinoma
<b>EBV(-)</b>	Epstein-Barr virus negative
<b>EBV(+)</b>	Epstein-Barr virus positive
<b>EGFR</b>	epidermal growth factor receptor
<b>EMAST</b>	elevated microsatellite alterations at selected tetranucleotide repeats
<b>EMT</b>	epithelial-mesenchymal transition
<b>Epi</b>	epirubicin
<b>ERBB2/HER2</b>	erb-B2 receptor tyrosine kinase 2
<b>ERBB3</b>	erb-B2 receptor tyrosine kinase 3
<b>et al.</b>	and others

### F

<b>FAM</b>	6-carboxyfluorescein
<b>FAP</b>	familial adenomatous polyposis
<b>FFPE</b>	formalin fixed paraffin-embedded
<b>FISH</b>	fluorescence in situ hybridization
<b>FLOT</b>	fluorouracil, leucovorin, oxaliplatin and docetaxel
<b>5-FU</b>	5-fluorouracil

<b>G</b>		
	<b>GATA4</b>	GATA binding protein 4
	<b>GC</b>	gastric carcinoma
	<b>gDNA</b>	genomic deoxyribonucleic acid
	<b>GEJ</b>	gastro-esophageal junction
	<b>GIST</b>	gastrointestinal stromal tumors
	<b>GRCh37</b>	Genome Reference Consortium Human Build 37
	<b>GS</b>	genomic stable
<b>H</b>		
	<b>HEX</b>	hexachloro-fluorescein
	<b><i>H. pylori</i></b>	<i>Helicobacter pylori</i>
	<b>HDGC</b>	hereditary diffuse gastric carcinoma
	<b>hMSH2</b>	human mutS homolog 2
	<b>hMSL1</b>	human mutL homolog 1
	<b>HNPCC</b>	hereditary non-polyposis colorectal cancer
	<b>HR</b>	hazard ratio
<b>I</b>		
	<b>IARC</b>	International Agency for Research on Cancer
	<b>ICGC</b>	International Cancer Genome Consortium
	<b>IGSR</b>	International Genome Sample Resource
	<b>IHC</b>	immunohistochemistry
<b>K</b>		
	<b>KRAS</b>	KRAS proto-oncogene
<b>L</b>		
	<b>Log2R</b>	log intensity ratio
	<b>LOH</b>	loss of heterozygosity
<b>M</b>		
	<b>MANEC</b>	mixed adenoneuroendocrine carcinomas
	<b>MAPK</b>	mitogen-activated protein kinase
	<b>MIP</b>	molecular inversion probe
	<b>MMR</b>	mismatch repair
	<b>MSI</b>	microsatellite instability
	<b>MSI-H</b>	high microsatellite instability
	<b>MSI-L</b>	low microsatellite instability
	<b>MSS</b>	microsatellite stable
	<b>MTOR</b>	mechanistic target of rapamycin kinase
	<b>MYC</b>	MYC proto-oncogene
<b>N</b>		
	<b>N.</b>	non-tumorous tissue
	<b>n or No.</b>	number of patients
	<b>n/a</b>	no data available
	<b>NARD</b>	normalized amplicon read depth
	<b>NCBI</b>	National Center for Biotechnology Information
	<b>NCI</b>	National Cancer Institute
	<b>NGS</b>	next-generation sequencing
	<b>nr</b>	not reached
	<b>NSCLC</b>	non-small lung cell lung carcinomas
<b>O</b>		
	<b>OS</b>	overall survival
	<b>Ox</b>	oxaliplatin
<b>P</b>		
	<b>Pac</b>	paclitaxel
	<b>PCR</b>	polymerase chain reaction
	<b>PD-1</b>	programmed cell death-1 receptor
	<b>PDCD1LG2</b>	programmed cell death 1 ligand 2
	<b>PD-L1</b>	programmed cell death ligand 1
	<b>PI3K</b>	phosphatidylinositol-3 kinase
	<b>PK</b>	proteinase K
	<b>PIK3CA</b>	phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit alpha
	<b>pM</b>	metastasis category
	<b>pN</b>	nodal status
	<b>pT</b>	primary tumor status
	<b>PTEN</b>	phosphatase and tensin homolog

<b>R</b>		
	<b>ref.</b>	reference
	<b>Ref. Seq.</b>	Reference Sequence
	<b><i>RHOA</i></b>	Ras homolog family member A
	<b>RKI</b>	Robert Koch Institut
	<b>rpm</b>	revolutions per minute
	<b>RTK</b>	receptor tyrosine kinase
<b>S</b>		
	<b>SCNA</b>	somatic copy-number aberrations
	<b>SD</b>	sStandard deviation
	<b><i>SMAD4</i></b>	SMAD family member 4
	<b>SNP</b>	single nucleotide polymorphism
	<b>SRCC</b>	signet ring cell carcinomas
	<b>Subst.</b>	substitution
<b>T</b>		
	<b>T.</b>	tumorous tissue
	<b>TCGA</b>	The Cancer Genome Atlas
	<b>TGF</b>	transforming growth factor
	<b>TMB</b>	tumor mutation burden
	<b><i>TP53</i></b>	tumor protein P53
	<b>TRG</b>	tumor regression grade
<b>U</b>		
	<b>UCSC</b>	University of California Santa Cruz
	<b>UICC</b>	Union for International Cancer Control
<b>V</b>		
	<b>v/v</b>	volume by volume
<b>W</b>		
	<b>w/v</b>	weight by volume
	<b>WHO</b>	World Health Organization

### Physical values

<b>%</b>	percentage
<b>°C</b>	degree centigrade
<b>µl</b>	microliter
<b>µM</b>	micromolar
<b>µm</b>	micrometer
<b>h</b>	hour
<b>L</b>	liter
<b>mg</b>	milligram
<b>min</b>	minute
<b>ml</b>	milliliter
<b>mo</b>	month
<b>ng</b>	nanogram
<b>nm</b>	nanometer
<b>pmol</b>	picomol
<b>pM</b>	picomolar [mol/L]
<b>sec</b>	second
<b>yr(s)</b>	year(s)

## ABSTRACT

The Cancer Genome Atlas (TCGA) project identified four molecular subgroups of gastric carcinomas (GCs) by the use of high-throughput technologies and comprehensive genome-wide analysis. Correspondingly, the tumors were classified as Epstein-Barr virus positive (EBV+), microsatellite unstable (MSI), genomically stable (GS) and chromosomally unstable (CIN). However, the clinical significance of these molecular subgroups for patients with GCs still remains unclear. For this reason, the aim of this doctoral thesis was to categorize a large GC cohort according to the TCGA classification and to clarify the predictive and prognostic implications of the molecular subgroups in the context of preoperative platinum/5-fluorouracil based chemotherapy (CTx). Comprehensive analysis was performed for a total of 759 GCs consisting of tumor biopsies before CTx and resected tumors without or after preoperative CTx. As the EBV- and MSI-status can be easily and reliably determined using standard methods, no generally diagnostic method was available for the detection of CIN in gastric cancer until then. Moreover, CIN-tumors represent a very heterogeneous group and no consistent definition exists in the literature.

In the present doctoral thesis, the detection of allelic imbalance (AI) by microsatellite based multiplex PCR assays was used as surrogate for chromosomal instability (CIN). Microsatellite analysis was compared to the genome-wide single nucleotide polymorphism (SNP)-based OncoScan method on a subset of the tumors to define a classification of CIN based on both methods. A concordance of 84% was found between the two methods regarding the individual microsatellite regions. The genome-wide extent of alterations was determined in analogy to the definition of TCGA and a positive correlation of both methods was shown ( $r = 0.88$ ). A CIN classification based on both methods was defined according to TCGA and the tumors were categorized into CIN and GS. The microsatellite based multiplex test was performed on a total of 759 GCs to determine the CIN-status. Besides, the CIN-assay was analyzed in relation to intra-tumor heterogeneity and revealed that in 93% of the tumors a concordant CIN classification in all tumor areas per patient were found. The limit of detection of CIN by the microsatellite-based assay was determined by dilution experiments using corresponding non-tumorous and tumorous DNA. A stable CIN classification was given at a minimal tumor cell content of 10%.

Analogous to the molecular classification in the four original TCGA subgroups, 4% of the patients with resected tumors before or after CTx were classified as EBV (+), 10% as MSI-H, 9% as GS and 77% as CIN. A prognostic relevance of the TCGA classification was found in the resected patient cohort ( $p = 0.021$ ). Thereby, patients with EBV (+) and MSI-H tumors showed the best survival and those with CIN-associated tumors the worst. No significant association was shown regarding response to preoperative CTx.

A more detailed modified classification system based on TCGA was developed as part of this doctoral thesis and GCs were categorized into the following five subgroups: EBV (+), MSI-H, low MSI (MSI-L), high CIN-medium (-M) and remaining CIN-groups. Thereby, the CIN classification was based on the calculation of quartiles of the AI ratios irrespectively of the definition of TCGA.

Accordingly, 4% of the patients with resected tumors before or after CTx were classified as EBV (+), 10% as MSI-H, 4% as MSI-L, 25% as high CIN-M and 58% of the tumors were summarized in the remaining CIN-groups. The modified classification system proposed in this study revealed significant associations with prognosis and prediction. Considering overall survival of patients with resected tumors before or after preoperative CTx concerning the modified subgroups, demonstrated that high CIN-M was associated with worse survival, which was particularly evident for patients after preoperative CTx ( $p = 0.043$ ). In contrast, MSI-L was significantly associated with worse prognosis of resected patients without CTx ( $p = 0.022$ ). The two subgroups MSI-L and CIN-H were significantly associated with response to preoperative CTx ( $p = 0.016$ ).

Tumor cell plasticity was determined by comparing paired tumor biopsies before and resected tumors after preoperative CTx with regard to the modified CIN classification. The analysis revealed that 50% of the tumors were initially classified in a higher CIN group and after receiving CTx in a lower group. This may reflect that tumor cells harboring increased AI were mainly eliminated by chemotherapeutic agents compared to those with lower AI. In contrast, the EBV- and MSI-status was the same for all corresponding tumors.

Through additional analysis of a subset of the tumors by next-generation sequencing, the most frequent mutations were detected in *TP53*, *CDH1* and *TGFBR2* using a GC related gene panel. Mutations in *CDH1* and *RHOA* were primarily detected in genomically stable tumors and were associated with the non-intestinal subtype. *TP53* mutations occurred more frequently in the high CIN-M subtype and correlated with proximal tumor localization. Mutations in genes which are involved in PI3K/AKT signaling and in cell cycle processes were also detected more frequently in the high CIN-M group.

In conclusion, high CIN-M and MSI-L were identified as two new molecular subgroups with potential predictive and prognostic impact in the context of preoperative CTx through comprehensive analysis of a large cohort of GCs and adenocarcinomas of the esophageal junction. This may be of great relevance for alternative therapy strategies and for the application of new therapeutic approaches especially for patients with tumors harboring these subtypes.

## ZUSAMMENFASSUNG

Durch die Entwicklung von Hochdurchsatztechnologien und umfassenden genomweiten Analysen von Magenkarzinomen ist es dem Cancer Genome Atlas (TCGA) Projekt gelungen vier molekulare Subgruppen zu identifizieren. Tumoren wurden entsprechend in Epstein-Barr Virus positiv (EBV+), mikrosatelliteninstabil (MSI), genomisch stabil (GS) und chromosomal instabil (CIN) eingeteilt. Die klinische Bedeutung dieser molekularen Subgruppen für Patienten mit Magenkarzinomen ist jedoch bisher noch weitgehend unklar. Aus diesem Grund war das Ziel dieser Doktorarbeit eine große Magenkarzinomkohorte entsprechend der TCGA-Klassifikation zu kategorisieren und die prädiktive und prognostische Relevanz der molekularen Subgruppen in Hinblick auf eine präoperative Platin/5-Fluoracil basierte Chemotherapie (CTx) zu überprüfen. Dazu wurden umfassende Analysen an insgesamt 759 Magenkarzinomen bestehend aus Tumorbiopsien vor einer CTx und Resektaten ohne und nach einer präoperativen CTx durchgeführt. Während sich der EBV- und MSI-Status relativ einfach und zuverlässig mit Standardmethoden bestimmen lässt, gab es bis dato keine allgemein diagnostische Methode für die Detektion von CIN in Magenkarzinomen. Zudem stellen die Karzinome mit CIN eine sehr heterogene Gruppe dar und es existiert keine einheitliche Definition in der Literatur.

In der hier vorliegenden Doktorarbeit wurde die Detektion von allelischen Imbalancen (AI) anhand eines mikrosatelliten-basierten multiplex Tests als Surrogat für CIN verwendet. Die Mikrosatellitenanalyse wurde mit der genomweiten Einzelnukleotid-Polymorphismus (SNP)-basierten OncoScan-Methode an einer Teilmenge von Magenkarzinomen verglichen, um eine Klassifikation von CIN auf der Grundlage beider Methoden zu definieren. Bei Betrachtung der einzelnen Mikrosatellitenregionen wurde eine Übereinstimmung von 84% der beiden Methoden gefunden. Das genomweite Ausmaß der Alterationen wurde in Analogie zur Definition nach TCGA bestimmt und eine positive Korrelation wurde beim Vergleich der beiden Methoden aufgezeigt ( $r = 0.88$ ). Eine CIN Klassifikation basierend auf beiden Methoden wurde entsprechend der Einteilung nach TCGA definiert und die Tumoren als CIN und GS klassifiziert. Zur Bestimmung des CIN-Status wurde der mikrosatelliten-basierte multiplex Test insgesamt an 759 Magenkarzinomen angewandt. Der CIN-Test wurde zudem in Hinblick auf intra-tumorale Heterogenität analysiert und ergab, dass in 93% der Fälle eine Übereinstimmung der CIN-Klassifikation innerhalb der verschiedenen Tumorareale pro Fall bestand. Das Detektionslimit von CIN des mikrosatelliten-basierten Tests wurde über Mischungsversuche von korrespondierender Normal- und Tumor-DNA bestimmt. Eine stabile CIN-Klassifikation konnte bei einem minimalen Tumorzellgehalt von 10% erreicht werden.

Entsprechend der molekularen Einteilung in die vier klassischen TCGA Subgruppen wurden 4% der Patienten mit resezierten Tumoren vor oder nach einer CTx als EBV(+) klassifiziert, 10% als MSI-H, 9% als GS und 77% als CIN. Eine prognostische Relevanz der TCGA-Klassifikation wurde bei der resezierten Patientenkohorte gezeigt ( $p = 0.021$ ). Patienten mit EBV(+) und MSI-H Tumoren zeigten dabei das beste Überleben und diejenigen mit CIN-assoziierten Tumoren das schlechteste. Hinsichtlich dem Ansprechen auf eine präoperative CTx konnte keine signifikante Assoziation gezeigt werden.

Ein detaillierteres modifiziertes Klassifikationssystem in Anlehnung an TCGA wurde im Rahmen dieser Doktorarbeit entwickelt und die Magenkarzinome wurden in folgende fünf Subgruppen unterteilt: EBV(+), MSI-H, niedriggradig MSI (MSI-L), *high* CIN-medium (-M) und zusammengefasste restliche CIN-Gruppen. Die CIN-Klassifikation basierte dabei auf der Berechnung von Quartilen der AI-Ratios unabhängig von der Definition von TCGA. Entsprechend wurden 4% der Patienten mit resezierten Tumoren vor oder nach einer CTx als EBV(+) klassifiziert, 10% als MSI-H, 4% als MSI-L, 25% als *high* CIN-M und 58% der Tumoren wurden in die restliche CIN-Gruppen zusammengefasst. In Hinblick auf Prognose und Prädiktion wurden in dem hier vorgeschlagenen modifizierten System signifikante Korrelationen gefunden. Bei Betrachtung des Gesamtüberlebens der Patienten mit resezierten Tumoren vor oder nach einer präoperativen CTx hinsichtlich der modifizierten Subgruppen wurde gezeigt, dass *high* CIN M mit einem schlechten Überleben assoziiert war, was sich vor allem in Patienten nach einer präoperativen CTx zeigte ( $p = 0.043$ ). MSI-L hingegen war signifikant mit einer schlechten Prognose in resezierten Patienten ohne einer CTx assoziiert ( $p = 0.022$ ). Die beiden Subgruppen MSI-L und CIN-H waren zudem signifikant mit dem Ansprechen auf eine präoperative CTx assoziiert ( $p = 0.016$ ). Die Tumorzellplastizität wurde anhand von korrespondierenden Tumorbiopsien vor und Resektaten nach einer CTx in Hinblick auf die modifizierte CIN-Klassifikation untersucht und ergab, dass 50% der Tumoren zunächst in eine höhere CIN-Gruppe klassifiziert wurden und nach einer CTx in eine niedrigere Gruppe. Dies könnte möglicherweise widerspiegeln, dass Tumorzellen mit hoher allelischer Imbalanz hauptsächlich von den Chemotherapeutika eliminiert werden im Vergleich zu denjenigen mit niedrigeren AI-Raten. Im Gegensatz dazu war der EBV- und MSI-Status bei allen korrespondierenden Tumoren gleich.

Durch begleitende Analysen mit massiver paralleler Gensequenzierung einer Subgruppe der Tumoren unter Verwendung eines Magenkarzinom-spezifischen Genpanels wurden die häufigsten Mutationen in *TP53*, *CDHI* und *TGFBR2* detektiert. In genomisch stabilen Tumoren wurden vor allem *CDHI* und *RHOA* Mutationen detektiert, die mit dem nicht-intestinalen Subtyp assoziiert waren. *TP53* Mutationen traten gehäuft im *high* CIN-M Subtyp auf und waren mit einer proximalen Tumorlokalisation assoziiert. Mutationen in Genen, die im PI3K/AKT Signalweg und im Zellzyklus beteiligt sind, wurden ebenfalls häufiger in der *high* CIN-M Gruppe nachgewiesen.

Zusammenfassend ist zu sagen, dass *high* CIN-M und MSI-L in dieser Studie als zwei neue molekulare Subgruppen mit potenzieller prädiktiver und prognostischer Bedeutung im Kontext einer präoperativen CTx durch umfassende Analysen von Magenkarzinomen und Adenokarzinomen des ösophagealen Übergangs identifiziert wurden. Dies könnte von großer Bedeutung für alternative Therapiestrategien und für die Anwendung neuer Therapieansätze speziell bei Patienten mit Tumoren dieser Subtypen sein.

# 1. INTRODUCTION

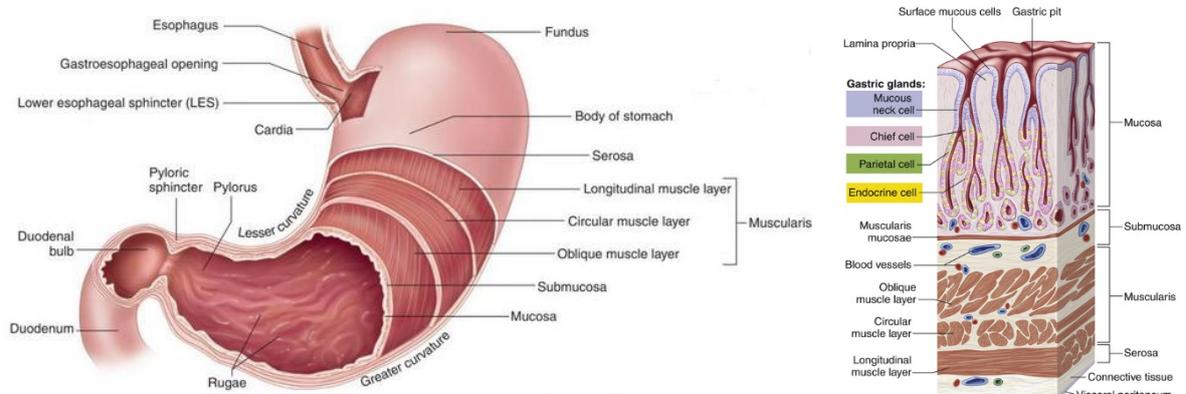
## 1.1. Gastric cancer

### 1.1.1. Anatomy and histology of the stomach

The human stomach is a muscular and hollow organ of the gastro-intestinal tract which is found in the upper left area of the abdominal cavity. The stomach is involved in the digestive system by temporary storage of assimilated food, partial chemical and mechanical digestion by muscle contracting and transfer of the chyme in portions to the following digestive organs (Patten and Thibodeau 2016).

The stomach is located between the anterior esophagus and the duodenum and is divided into the following five main regions: cardia, fundus, body/corpus, antrum and pylorus (Kugler 2017). The cardia is directly connected to the esophagus and is followed by the corpus, which is the large central portion of the stomach containing the fundus and antrum (Figure 1). The junction between the distal esophagus and the proximal cardia region is called gastro-esophageal junction (GEJ) (Fried et al. 2013).

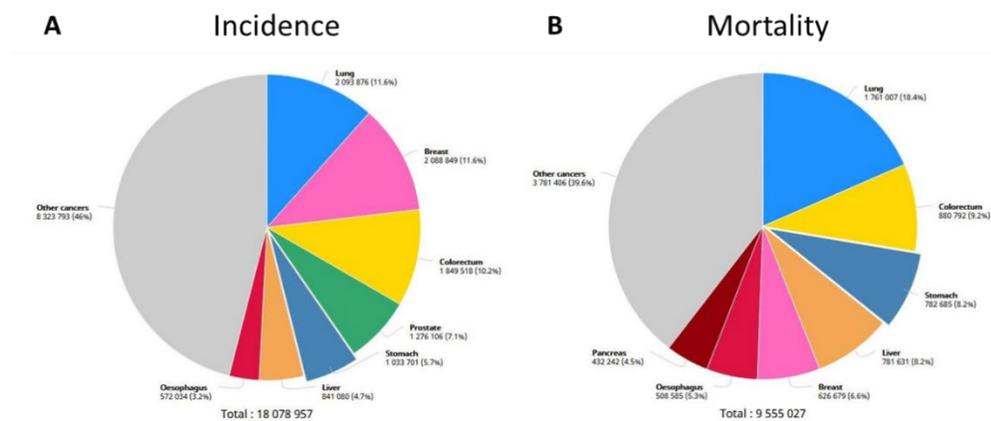
The wall of the stomach is comprised of four layers and is divided in the luminal located mucous membrane (Tunica mucosa) followed by the submucosa, muscularis propria and a serosa delimiting it to the abdomen (Figure 1). In the inner epithelium, numerous gastric pits are formed by invagination of the surface and are connected to gastric glands which secrete gastric juice containing enzymes as pepsin and muriatic acid for chemical food digestion. The secretory cells are especially located in glands of the fundus and body of the stomach. A mucinous secrete is mainly produced by the surface mucous cells of pyloric and cardiac glands to prevent self-digestion of the stomach wall (Fried et al. 2013, Patten and Thibodeau 2016).



**Figure 1: Anatomy of the stomach and layers of the stomach wall**  
(Patten and Thibodeau 2016)

## 1.1.2. Epidemiology of stomach cancer

Stomach cancer, also known as gastric cancer, is the fifth most common malignancy worldwide in both sexes with more than 1 million new cases in 2018, which represented 5.7% of all new cancer cases and the third leading cause of cancer-related mortality with an estimated number of 783.000 deaths in 2018 (Figure 2). Gastric carcinomas (GCs) are more prevalent in men and incidence rates are nearly 2-fold higher in men (7.2% of all new cancer cases) than in women (4.1% of all new cancer cases). Furthermore, 9.5% of all cancer deaths in men are attributable to GC compared to 6.5% in women (Bray et al. 2018).



**Figure 2: Incidence and mortality of the most common cancers worldwide for both sexes and all ages**

The estimated numbers of new cases (A) and of deaths (B) in 2018 are shown according to GLOBOCAN 2018 (Global Cancer Observatory, <https://gco.iarc.fr/>).

According to the GLOBOCAN data, gastric cancer was the tenth most frequently diagnosed cancer in Germany in both sexes in 2018 with an estimated number of around 14.000 new cases which represented 2.3% of all new cancer cases. Despite steadily decline in incidence rates, gastric cancer is the sixth leading cause of cancer-related death in Germany with a mortality rate of 3.8% in both sexes with an estimated number of 9.480 deaths in 2018 (Bray et al. 2018). In 2016, the median age at diagnosis for men was 72 years compared to 76 years for women. The prognosis of patients with GC is still unfavorable with a 5-year survival rate of all stages for both sexes between 32% and 34% (Cancer in Germany 2015/2016, RKI).

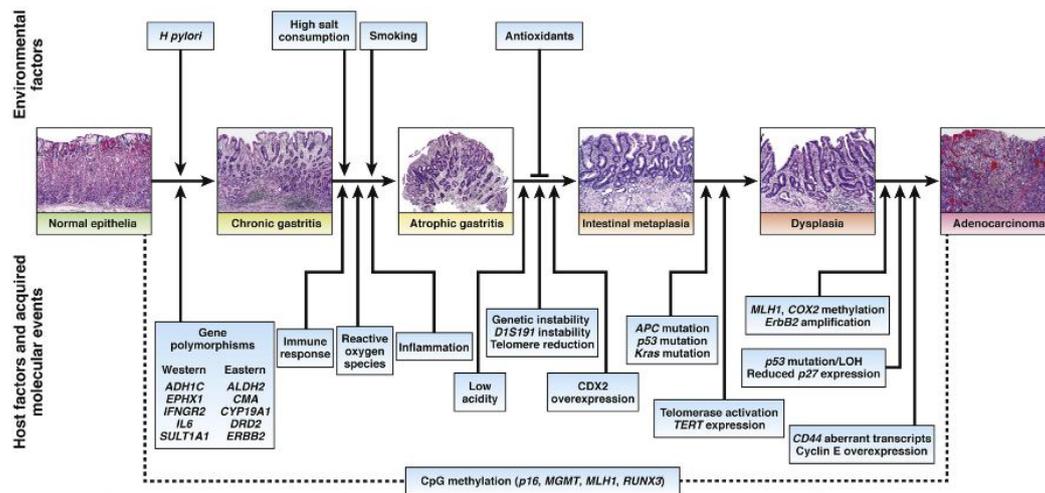
The incidence of gastric cancer varies geographically worldwide. The highest incidence rates were mainly recorded in eastern and central Asia and Latin America, whereas the rates were rather lower in northern America, Europe and the African regions. The Republic of Korea reached the highest rates worldwide in both sexes with an average incidence of 57.8 per 100.000 new cases among men and 23.5 per 100.000 among women. In several western Asian nations, gastric cancer is the leading oncological cause of death among men (Bray et al. 2018, Rawla and Barsouk 2019).

### 1.1.3. Etiology and risk factors

Different combinations of environmental and infectious risk factors and hereditary genetic predisposition can cause the pathogenesis of gastric cancer. The consumption of smoked meat with high nitrate content and salt-preserved or strongly salted foods are potential risk factors for the development of GC whereas a healthy dietary and lifestyle including high-intake of fresh foods and vegetables reduce the risk (Kono and Hirohata 1996, Wang et al. 2009, Bertuccio et al. 2013, Massarrat and Stolte 2014). Smoking tobacco and high alcohol consumption can also predispose to GC (Nishino et al. 2006, Ladeiras-Lopes et al. 2008, Ma et al. 2017). There is also strong evidence that obesity is associated with cardia subtypes (Lagergren et al. 1999, Lin et al. 2014). Through improved food conservation, dietary and lifestyle changes and the eradication of *Helicobacter pylori* (*H. pylori*), the incidence of stomach cancer steadily declines (Balakrishnan et al. 2017).

The most significant risk factor for stomach cancer is an infection with the bacterium *H. pylori* which induces chronic inflammation of the gastric mucosa and can lead to gastritis (Warren and Marshall 1983, Forman et al. 1991, Stolte and Meining 1998). According to the Correa-hypothesis, the first step of the multifactorial pathogenesis of GC is an acute inflammation of the gastric mucosa by *H. pylori* infection also depending on different environmental factors which can lead via an atrophic gastritis and intestinal metaplasia up to intestinal GC (Correa 1992). Not only environmental factors and infection with *H. pylori* play a crucial role in the pathogenesis of GC, also various molecular genetic events are associated (Figure 3). The World Health Organization (WHO) identified *H. pylori* as a definite (class I) carcinogen (WHO 1994). Around 60% of the world population has an infection with *H. pylori*, but only 1%-2% of the infected people will develop GC due to cytotoxin-associated gene A (CagA) positive *H. pylori* strains (Peleteiro et al. 2014). However, infected patients with *H. pylori* have a 3- to 6-fold increased risk of developing GC (Stolte and Meining 1998). Distal located (non-cardia) tumors of the stomach are strongly associated with chronic *H. pylori* infection (Mukaisho et al. 2015). The decline of the incidence of this GC subtype is attributable to successful prevalence and treatment strategies of *H. pylori* infections (Balakrishnan et al. 2017). In contrast, incidence rates of the cardia GC have an increasing tendency (Marqués-Lespier et al. 2016, Rawla and Barsouk 2019). An infection with *H. pylori* and interaction with polymorphisms of pro-inflammatory interleukin genes are associated with an increased risk of developing GC especially in Asian populations (Qinghai et al. 2014).

Infection with the human herpes virus, Epstein-Barr virus (EBV), is also associated with an increased risk of developing GC (Iizasa et al. 2012). EBV-associated GCs are found in a range between 5%-10% and are predominantly associated with male patients and younger age (Nishikawa et al. 2014). EBV positive tumors can be seen as own subtype of gastric cancer, due to good prognosis and distinct clinical-pathological characteristics (van Beek et al. 2004, TCGA 2014).



**Figure 3: Pathogenesis of intestinal gastric carcinoma including environmental and genetic factors** (Tan and Yeoh 2015)

### 1.1.4. Hereditary syndromes of gastric carcinoma

Although the majority of GCs arose sporadically, hereditary genetic predisposition is also a potential risk factor for developing gastric cancer. First-degree relatives of a patient with stomach cancer have a 2- to 3-fold increased risk to develop GC (Foschi et al. 2008, Moehler et al. 2011). The hereditary diffuse gastric carcinoma (HDGC) is an autosomal dominant disorder and occurs very rarely among all GCs (~1%). HDGC is associated with germline mutations of the tumor suppressor gene *CDH1*, which encodes E-cadherin a protein involved in cell-cell adhesion (Oliveira et al. 2009, Fitzgerald et al. 2010). The lifetime risk of patients carrying inherited *CDH1* mutations to develop diffuse GC is 56%-70% by age 80 years and woman have additionally a higher risk (40%-50%) for the development of lobular breast cancer (Pharoah et al. 2001, Kaurah et al. 2007, van der Post et al. 2015). Patients harboring an identified *CDH1* germline mutation are recommended for prophylactic gastrectomy (Huntsman et al. 2001, Moehler et al. 2011).

Patients with hereditary non-polyposis colorectal cancer (HNPCC), also known as Lynch syndrome, have a higher risk for the development of GC. HNPCC is associated with microsatellite instability (MSI) which is caused by mutations in one of the mismatch repair genes (Velho et al. 2014). The most frequent alterations occur in the genes human mutS homolog 2 (*hMSH2*) and human mutL homolog 1 (*hMLH1*) (Mitchell et al. 2002). Furthermore, GC is also associated with a range of other hereditary cancer syndromes such as familial adenomatous polyposis (FAP) which is caused by *APC* germline mutation, Li-Fraumeni or Peutz-Jeghers syndrome (Keller 2002, Oliveira et al. 2015).

### 1.1.5. Staging and histological classification

The most common malignancies of the stomach are gastric adenocarcinomas (~90%) which originate from the epithelial gastric mucosa (Brenner et al. 2009). The remaining malignant tumors of the stomach include rare types like primary lymphomas, gastrointestinal stromal tumors (GIST), gastrointestinal neuroendocrine neoplasms (carcinoid tumors) or mixed adenoneuroendocrine carcinomas (MANEC). Gastric lymphomas often arose in the context of chronic *H. pylori* gastritis (Juarez-Salcedo et al. 2018).

Gastric cancer is a heterogeneous disease which exhibits multiple morphological and genetic phenotypes and therefore a variety of different classification systems exist (Cislo et al. 2018). Adenocarcinomas of the stomach are primarily classified according to the latest TNM-staging system (UICC 2017, 8<sup>th</sup> edition). Thereby, the size and extent of the primary tumor (T-category), the involvement of regional lymph nodes (N-category) and the occurrence of distant metastasis (M-category) are recorded. The clinical staging determines the pretreatment extent of a tumor and the pathological staging system supplements detailed information after surgical treatment. The staging is a crucial factor for prognosis of a patient and the decision for treatment strategies (Amin et al. 2017).

Gastric adenocarcinomas are additionally described according to their proximal or distal localization in the stomach. Proximal tumors are located at the junction between the esophagus and cardia region of the stomach and are usually called cardia carcinomas whereas distal located tumors were found in the antrum region of the stomach and are called non-cardia carcinomas (Hu et al. 2012, Rawla and Barsouk 2019). Adenocarcinomas of the esophageal junction (AEG) are classified into the three subtypes AEG I, AEG II and AEG III according to the Siewert-classification (Siewert and Stein 1998). Type I tumors are classified as adenocarcinomas of the distal esophagus (or Barrett carcinoma) and type II tumors as cardia carcinomas. The type III tumors infiltrate the esophageal junction as well as the distal esophagus and are classified as GCs whereas AEG I and AEG II are assigned and treated like esophageal carcinoma (German S3-guidelines 2019).

The growth behavior of tumor cells is determined histologically according to the Laurén classification which divides tumors into intestinal, diffuse and indeterminate subtypes (Lauren 1965). The intestinal subtype is the most frequent type of GC and is histopathologically characterized by a high grade of glandular differentiation and expansive growth pattern. Furthermore, the intestinal subtype occurs more frequently in European countries and is more often associated with distal located tumors (Sitarz et al. 2018). The progression of intestinal GC is preceded by a cascade of multistep precancerous lesions (Correa 1992). Various genetic factors as overexpression of CDX2 can induce intestinal metaplasia and mutations in several driver genes (*APC*, *TP53*, *KRAS*) and MLH1 promoter methylation can promote the formation of intestinal adenocarcinomas (Figure 3) (Correa 1992, Yuasa 2003, Marqués-Lespier et al. 2016).

In contrast, the diffuse subtype has a more aggressive growth pattern with cells showing no cohesiveness and poor differentiation. This subtype is usually diagnosed in younger individuals and overall has a worse prognosis. The development of diffuse GC differs from the intestinal subtype and is more based on hereditary and genetic factors lacking defined precursor lesions (Marqués-Lespier et al. 2016, Cisto et al. 2018).

Alternatively, GCs are classified into more detailed histological subtypes by the WHO in papillary, mucinous, tubular, mixed tumors and uncommon variants with minor clinical relevance (WHO 2019). The WHO classification is the most detailed method of classifying GC and includes gastric adenocarcinomas as well as less common gastric tumors such as the very rare MANEC and the signet ring cell carcinomas (SRCC) (Gurzu et al. 2015). Tubular, papillary and mucinous carcinomas are comparable with the intestinal histological subtype according to Laurén whereas SRCC are poorly cohesive and are more likely classified as diffuse type tumors (Berlth et al. 2014).

### **1.1.6. Diagnostic and multimodal treatment**

After precise tumor staging and histopathological diagnosis, an individual therapy concept follows depending on the degree of tumor spread and the physical condition of the patient (Meyer and Wilke 2011). Endoscopic examination of the esophagus and stomach is the standard procedure for the detection of precursor lesions and intraepithelial neoplasia. Endoscopic resection is performed as curative intent according to the Japanese treatment guidelines in very early staged GC in which the tumor invasion is limited to the mucosa and clinically diagnosed as T1a (Takahashi et al. 2013, Japanese Gastric Cancer Association 2017). Population-based endoscopic screening is only common in high-risk areas such as Japan or Korea and for now not provided in Germany (German S3-guidelines 2019).

Most of the GCs are often diagnosed at an advanced tumor stage mostly with metastasis disease due to late or unspecific development of physical symptoms. Thus, the late medical outcome is associated with a poor prognosis (Moehler et al. 2011). The standard therapy for patients with advanced GC in European countries is preoperative CTx followed by surgical resection, which depicted the only curative treatment option (Smyth et al. 2016). Depending on the tumor stage and localization, subtotal or partial gastrectomy is performed including abdominal D2 lymphadenectomy (German S3-guidelines 2019). To shrink the tumor mass prior to the surgical resection, usually preoperative (or neoadjuvant) chemotherapy (CTx) is necessary. Patients with inoperable or recurrent tumors and advanced metastatic spread were treated with palliative CTx regimens to improve survival and obtain livability.

According to the recommendations of the German S3-guidelines (2019), patients with GC with advanced clinical tumor stage cT3 and cT4 should undergo perioperative CTx whereas for patients with cT2 categorized tumors only a “can” recommendation is provided. A perioperative multimodal treatment consists of neoadjuvant CTx given prior to surgery and continued in an adjuvant setting.

A first randomized phase III study (MAGIC study) demonstrated benefits of perioperative CTx in patients with advanced gastric cancer and the CTx regimen containing epirubicin, cisplatin and 5-fluorouracil (ECF) was proposed as therapy standard (Cunningham et al. 2006). According to the latest German S3-guidelines published in 2019, the platinum/5-fluorouracil (5-FU) based CTx regimens were replaced by a new triplet combination consisting of fluorouracil plus leucovorin, oxaliplatin and docetaxel (FLOT) due to the findings of the FLOT4-study (Al-Batran et al. 2016, Al-Batran et al. 2019).

Patients with human epidermal growth factor receptor 2 (HER2)-positive advanced GCs can additionally be targeted with the monoclonal antibody Trastuzumab in combination with standard CTx regimens (Bang et al. 2010). Overexpression or gene amplification of HER2 is detected in around 13%-20% of stomach cancer and acts a negative prognostic factor (Gravalos and Jimeno 2008, Bang et al. 2010, Park and Chun 2013). Immune checkpoint inhibition is another new possibility for targeted therapy especially in metastatic GC with high MSI (MSI-H) and positive programmed cell death-ligand 1 (PD-L1) expression (Kelderman et al. 2015, Le et al. 2015, Ma et al. 2016). The monoclonal antibody Pembrolizumab inhibits the activity of the programmed cell death-1 (PD-1) receptor and is already accepted in first line therapy in non-small lung cell lung carcinomas (NSCLC) (Brar and Shah 2019). Currently, various immune checkpoint inhibitors are tested in clinical trials for targeted therapy in GCs (Fontana and Smyth 2016, Goode and Smyth 2016, Smyth and Moehler 2019).

Tumor regression is not yet evaluated consistently worldwide in gastrointestinal carcinomas after neoadjuvant treatment and various classification systems exist (Thies and Langer 2013, Langer and Becker 2018). In Germany, the response to preoperative CTx is usually determined histopathologically according to the Becker-classification and categorized into three tumor regression grades (TRG) (Becker et al. 2011). The grading is based on the percentage of residual tumor tissue in relation to the macroscopically identifiable tumor bed after neoadjuvant CTx (Becker et al. 2011, Thies and Langer 2013). Accordingly, patients with tumors showing complete and subtotal regression (TRG1) were classified as responders whereas patients with tumors showing partial, minimal or no regression (TRG2/3) as non-responders. The prognostic relevance of this classification system has been demonstrated in several studies in GCs and esophagogastric adenocarcinomas (Becker et al. 2011, Lorenzen et al. 2013, Schmidt et al. 2014).

## 1.2. Molecular classification systems of gastric carcinomas

Improvements in high-throughput technologies enable the determination and characterization of comprehensive genetic profiles of GC. Based on this, two international consortia, the Cancer Genome Atlas (TCGA) project and the Asian Cancer Research Group (ACRG) identified distinct molecular subgroups of GC which revealed correlations with clinical-pathological characteristics (TCGA 2014, Cristescu et al. 2015).

### 1.2.1. TCGA classification and characteristics of molecular subgroups

In the context of the TCGA study a comprehensive genome-wide analysis of 295 gastric adenocarcinomas of patients treated without pre-therapeutic CTx or radiotherapy was performed on genetic and protein level and the following four distinct molecular subtypes were identified: Epstein-Barr virus positive (EBV+), microsatellite unstable (MSI), genomically stable (GS) and chromosomally unstable (CIN) tumors (TCGA 2014). According to their classification algorithm, 9% of the tumors were categorized as EBV(+) and 22% as MSI. The remaining tumors were classified according to the degree of their aneuploidy into 20% as GS and 50% as CIN tumors (Figure 5A).

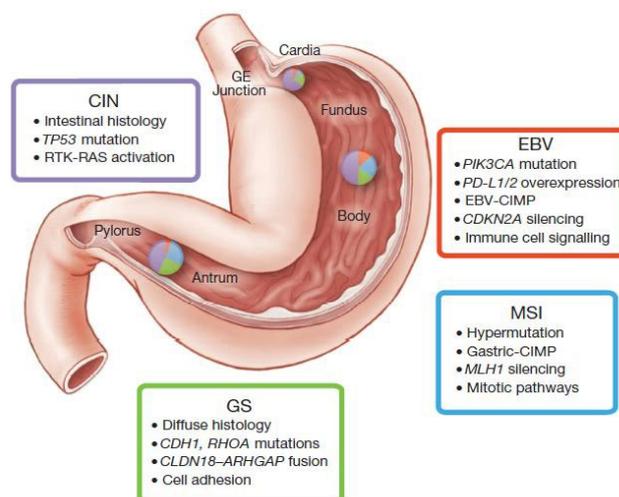
The EBV(+) subtype was mainly localized in the fundus or body of the stomach and more frequently found in male patients (Figure 4). All EBV(+) tumors exhibited the CpG island methylator phenotype (CIMP) which is usually characterized by hypermethylation of promoter island sites rich with CpG dinucleotides (Nazemalhosseini Mojarad et al. 2013). Especially CDKN2A promoter hypermethylation occurred in the EBV positive subtype. The main characteristics of the MSI subtype were high mutation rates and hypermethylation of the promoter of MLH1. Furthermore, MSI tumors were associated with older age.

The remaining carcinomas, microsatellite stable (MSS) and EBV negative (-), were further classified according to their status of somatic copy-number aberrations (SCNA). The GS tumors were associated with the diffuse histological subtype whereas the CIN subgroup was more frequently found in intestinal type tumors and in tumors localized in the gastric cardia or gastroesophageal junction (Figure 4). The chromosomally unstable tumors with a high number of copy alterations showed deletions and amplifications of specific chromosomal regions targeting oncogenes and tumor suppressor genes. The regions 17q12 (*ERBB2*), 19q12 (*CCNE1*), 12p12.1 (*KRAS*), 8q24.21 (*MYC*), 7q21.1 (*CDK6*), 8p23.1 (*GATA4*), and 7p12 (*EGFR*) were frequently amplified whereas focal deletions were identified at the chromosomal regions of tumor suppressors such as 18q21.2 (*SMAD4*), 9p21 (*CDKN2A*), 10q23.3 (*PTEN*), and 1p35.3 (*ARID1A*). They also identified a recurrent amplification at 9p24.1 containing *CD274* and *PDCD1LG2* which encode for the immunosuppressant proteins PD-L1 and PD-L2. These 9p amplifications were found in 15% of the EBV(+) tumors.

Mutation profiling of non-hypermethylated tumors was performed by whole-genome sequencing and revealed that in EBV(+) tumors, mutations preferentially occurred in *PIK3CA*, *ARID1A*, *BCOR*, whereas *TP53* mutations were rarely found. The MSI tumors showed hypermutation in genes encoding targetable proteins such as *PIK3CA*, *ERBB3*, *ERBB2* and *EGFR*. *CDH1* somatic mutations were most frequently found in the GS subtype and *RHOA* mutations were enriched exclusively in this subgroup. *TP53* mutations were identified in 71% of the CIN tumors (TCGA 2014).

Survival analysis of the TCGA patient cohort revealed no discernible differences in prognosis among the four molecular subgroups due to limited available follow-up data of the patients by the time of analysis (TCGA 2014).

In another study, a prediction model based on gene expression has demonstrated a correlation between the four molecular TCGA subgroups and clinical outcome of the patients in an adjuvant CTx setting (Sohn et al. 2017). However the overall clinical relevance of this molecular classification is still poorly characterized. An analysis of various GC cohorts encompassing considerably high number of patients and addressing different clinically relevant questions is necessary to estimate the full potential value of this “omics” based classification system.



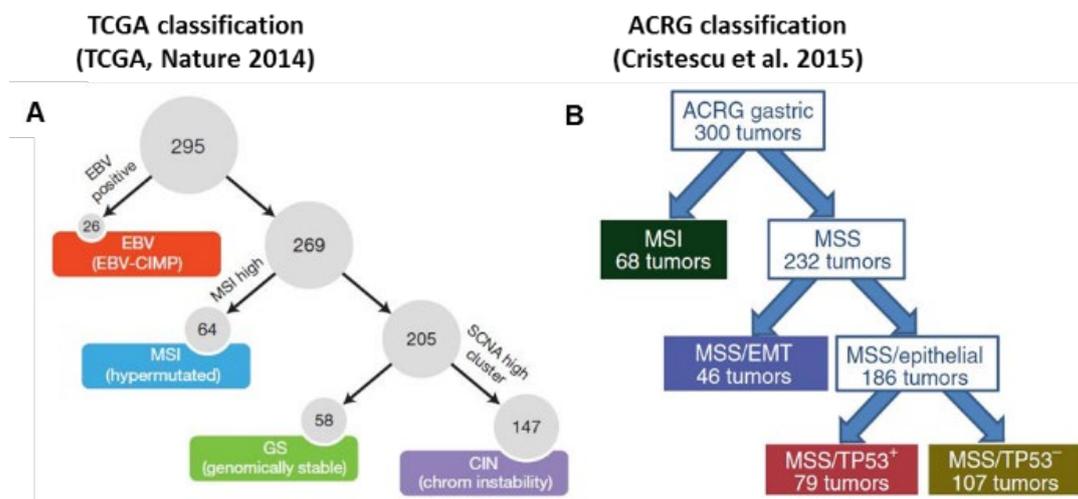
**Figure 4: Characteristics of the four gastric cancer subtypes according to the TCGA classification system (TCGA 2014)**

### 1.2.2. ACRG classification

Another molecular classification system was proposed by the Asian Cancer Research Group (Cristescu et al. 2015). A total of 300 primary resected GCs from patients treated without neoadjuvant CTx were analyzed by mRNA expression profiling, somatic copy number and targeted gene sequencing. According to their generated expression signatures, a classification system with four molecular subtypes was defined (Figure 5B). One aspect of the ACRG classification, which is different from the TCGA study, is that the molecular subtypes correlate significantly with prognosis.

MSI tumors with a loss of MLH1 expression and occurrence of a high number of mutations in *KRAS*, *ALK*, *ARID1A* or genes involved in the PI3K-PTEN-mTOR pathway represented the first subgroup and were clearly distinguished from the MSS tumors. The MSS tumors showed a typical gene expression pattern for epithelial-to-mesenchymal transition (EMT) with a loss of CDH1 expression and were therefore classified as a second subtype. MSI tumors showed the best prognosis, whereas patients with MSS/EMT tumors the worst. The remaining MSS tumors were further characterized according to their status of *TP53* activation on the basis of a two-gene expression analysis of *CDKN1A* and *MDM2*.

Tumors with intact *TP53* activity were classified as MSS/TP53+ and those with a *TP53* functional loss as MSS/TP53-. Patients with tumors of those two subgroups had an intermediate prognosis whereas the *TP53* active (wild type) subgroup showed better overall survival (OS). The highest prevalence of *TP53* mutations were found in the MSS/TP53- subgroup and recurrent amplifications in *ERBB2*, *EGFR*, *CCNE1*, *CCND1* and *MYC* were enriched especially in this subgroup. An infection with EBV occurred more frequently in the MSS/TP53+ subgroup (Cristescu et al. 2015).



**Figure 5: Molecular classification algorithms according to TCGA and ACRG**

According to the TCGA classification, tumors were categorized as Epstein-Barr virus positive (EBV+), microsatellite unstable (MSI), genomically stable (GS) and chromosomally unstable (CIN) (TCGA 2014) (A). The ACRG group identified the following molecular subgroups: MSI; microsatellite stable (MSS) and epithelial-to-mesenchymal transition (EMT) phenotype (MSS/EMT), MSS and intact *TP53* activity (MSS/TP53+) and MSS and *TP53* functional loss (MSS/TP53-) (Cristescu et al. 2015) (B).

### **1.3. Genomic instability in gastric carcinomas**

Genomic instability is a hallmark of distinct subclasses of tumors with potential clinical relevance. Genetic alterations in chromosomes and genetic instability are one of the main driving forces for cancer development and progression (Saeki et al. 2011, Ryland et al. 2015). Most of the tumors cells harboring genetic alterations show abnormalities in the number of chromosomes or structural changes of genes (Jefford and Irminger-Finger 2006). Genomically unstable tumors can be characterized by high rates of accumulating mutation events (Lengauer et al. 1998).

In general, genomic instability can be subdivided into MSI and CIN whereas both types of instability indicate mutator phenotypes in tumors (Loeb 2001, Geigl et al. 2008). In the context of the TCGA project, an integrative genomic analysis of GCs identified MSI and CIN as distinct molecular subgroups (TCGA 2014).

#### **1.3.1. Microsatellite instability**

Microsatellites are short repeating DNA motifs widespread throughout the human genome. MSI is mostly found in patients with Lynch syndrome and is associated with a defective mismatch repair (MMR) system caused by germline mutations in one of the genes coding for MMR proteins (Velho et al. 2014, Hause et al. 2016, Ratti et al. 2018, Yamamoto and Imai 2019). MSI occurred also sporadically in 10%-22% of gastric cancer or other gastrointestinal tumors due to epigenetic silencing by methylation of the DNA mismatch repair protein MLH1 (TCGA 2014, Ratti et al. 2018, Yamamoto and Imai 2019). MSI leads to the accumulation of various deletions and insertions of nucleotides in the repetitive microsatellite sequences and results in genomic hypermutability (Hause et al. 2016).

MSI can be detected by two standard methods, polymerase chain reaction (PCR) using panels of microsatellite markers and immunohistochemistry (IHC) using antibodies against MMR proteins. New approaches to detect MSI with next-generation sequencing (NGS) are currently under development (Salipante et al. 2014, Baudrin et al. 2018, Vanderwalde et al. 2018, Yamamoto and Imai 2019). MSI is a favorable prognostic and predictive marker in colorectal cancer (Goldstein et al. 2014, Ryan et al. 2017). However, the prognostic relevance of MSI in GC remains ambiguous (An et al. 2012, Fang et al. 2012, Smyth et al. 2017). Recent studies predict that patients with MMR deficient tumors and positive PD-L1 expression have a benefit for treatment with immune checkpoint inhibitors (Le et al. 2015, Le et al. 2017).

According to the National Cancer Institute (NCI) workshop in 1997, a marker panel of five microsatellite repeats was recommended for the evaluation of MSI in patients with Lynch syndrome and the definition of MSI was standardized (Boland et al. 1998). Accordingly, tumors were classified as having high-frequency MSI (MSI-H) if at least two of the five markers showed instability and as having low-frequency MSI (MSI-L) if only one of the five markers was unstable. Tumors without any instability were classified as microsatellite stable.

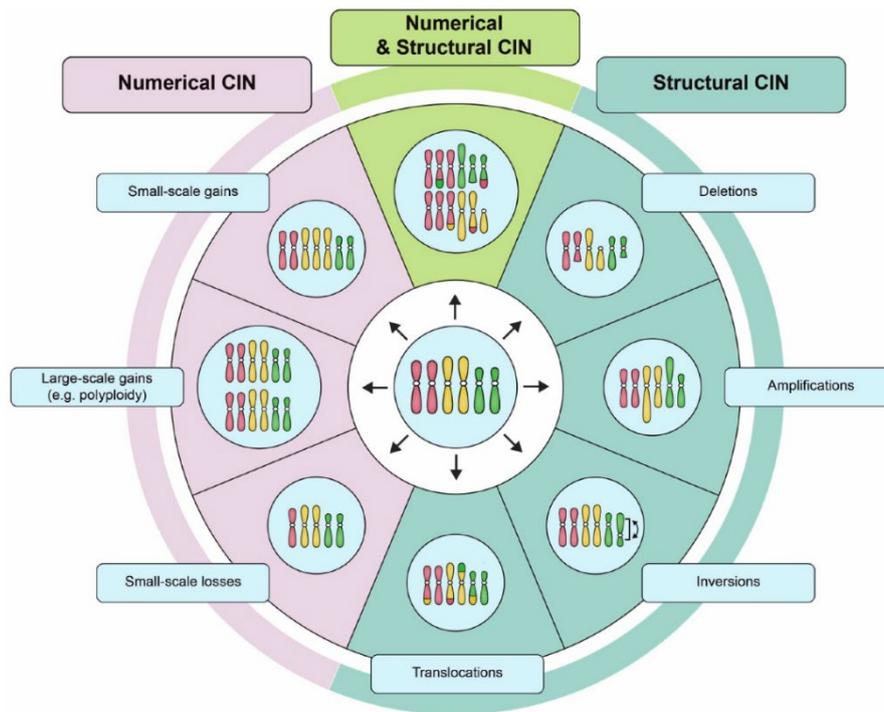
MSI-L tumors are usually not distinguished with MSS tumors and in contrast to the known clinical relevance of the MSI-H phenotype; the biological significance of MSI-L remains unclear or is not defined yet (Lee et al. 2015, Nazemalhosseini Mojarad et al. 2016).

### **1.3.2. Chromosomal instability**

CIN is considered to be a main driving force for the observed aneuploidy in many tumor types, but the term CIN is still controversially discussed and not clearly defined. CIN is suggested to describe a rate of losses and gains of whole or parts of chromosomes whereas other studies defined CIN as chromosomal rearrangements and frequent change in the chromosome number (Lengauer et al. 1998, Geigl et al. 2008). In general, chromosomal alterations can be distinguished between numerical and structural CIN (Figure 6). While partial deletions, amplifications, inversions or translocations of a single gene locus or chromosomal arms lead to structural CIN, numerical CIN includes gains and losses of whole chromosomes which can lead to a polyploid karyotype (Lepage et al. 2019).

CIN associated tumors express a very heterogeneous phenotype due to the occurrence of various and complex karyotypes in different parts of the tumor (Maleki and Röcken 2017). CIN can be detected by many techniques like comparative genomic hybridization (CGH), microsatellite-based PCR assays, fluorescence in situ hybridization (FISH) or single nucleotide polymorphisms (SNP) arrays (Geigl et al. 2008, Lepage et al. 2019). However, to date no consistent method exists for the reliable detection of CIN. Chromosomal instability in the TCGA study of GCs and also of esophageal adenocarcinomas (EAC) refers to the description of the extent of SCNA in the tumors mainly determined by genome-wide SNP based microarray technology (TCGA 2014, TCGA 2017, Liu et al. 2018). According to TCGA (2014), a chromosomal arm was considered to be altered if at least 66% of the arm was lost or gained. A more precise classification of CIN tumors was given from the TCGA consortium in a further molecular classification system of esophageal carcinoma (TCGA 2017). Accordingly, chromosomal arms were considered to be altered if at least 80% of the arm was lost, gained or demonstrated AI. Tumors were classified as having high chromosomal instability (SCNA-High), if they showed at least one altered chromosomal arm, except for chromosome 18 and 21q, which were altered recurrently in tumors with low and high CIN. Comparison of the molecular subgroups revealed that EAC and chromosomal unstable GCs could be defined in one subgroup distinct from EBV(+), MSI and GS tumors (TCGA 2014, TCGA 2017).

Various altered genes were identified in the carcinogenesis of GC. However, a signature of genes which were especially involved in the development of CIN are not yet defined and identification of molecular markers is required to determine CIN consistently (Maleki and Röcken 2017).



**Figure 6: Examples of various karyotypes as results of numerical and structural CIN**  
(Lepage et al. 2019)

## 1.4. Mutational landscape and signaling pathways in gastric cancer

Massive parallel sequencing, also known as next-generation sequencing, enables the extensive identification of recurrent and novel somatic alterations of gastric cancer related genes and dysregulated signaling pathways to recognize potential therapeutic targets (Wang et al. 2014, Lin et al. 2015, Cai et al. 2019). In addition, global tumor mutation burden (TMB) can be assessed by whole-exome sequencing. Thereby, high TMB can be a potential biomarker of sensitivity to immune therapy (Cai et al. 2019, Wang et al. 2019).

Sporadic events during carcinogenesis like somatic mutations, rearrangements of chromosomes or hypermethylation of promoter regions can lead to inactivation of tumor suppressor genes by loss of heterozygosity (LOH) or activation of oncogenes. In sporadic carcinogenesis both alleles of a tumor suppressor gene have to be inactivated by recessive mutation events according to the Knudson's two-hit hypothesis. In hereditary carcinomas one allele is already inactivated through germline mutation and the second one has to be inactivated sporadically (Knudson 1971). LOHs without affecting a reduction of the copy number are defined as copy neutral LOHs (CN-LOH) and can occur through duplication of a maternal or paternal chromosomal region and coincident deletion of the other allele or mitotic recombination (O'Keefe et al. 2010, Saeki et al. 2011). In a systematic analysis of a large cohort of tumors from various cancer types, CN-LOH was found in classical tumor suppressor genes such as *TP53*, *BRCA1*, *BRCA2* or *PTEN* (Kandoth et al. 2013).

One of the most important driver genes which is involved in the progression of diverse types of cancer and also of GC especially in highly chromosomal unstable tumors is the *TP53* gene encoding a tumor suppressor protein localized on chromosome 17p13.1 (Fenoglio-Preiser et al. 2003, TCGA 2014). *TP53* plays a key role in the genome integrity, different cellular processes and associated signaling pathways like control in cell proliferation, cell cycle, apoptosis, DNA repair and metabolism (Lane 1992). Germline mutations of *TP53* are associated with hereditary carcinomas such as the Li-Fraumeni syndrome which predisposes to various early-onset tumors (Malkin et al. 1990, Petitjean et al. 2007). According to the TP53 database of the International Agency for Research on Cancer (IARC) in 2019, the prevalence of somatic *TP53* mutations is the highest in colorectal, head and neck and esophagus carcinomas with a rate between 41%-43%. In stomach cancer, somatic *TP53* mutations occur in 32% of the analyzed cases (Bouaoun et al. 2016). The most frequent alterations of *TP53* are missense substitutions (73%) causing single amino acid changes at different positions which result in loss of function of the protein (Olivier et al. 2010, Bouaoun et al. 2016). The single nucleotide substitutions occur mostly at the DNA-binding domain in exons five to eight at well-known hotspot regions including codons 175, 220, 245, 248, 273 and 282 (Bouaoun et al. 2016).

Different signaling pathways play a crucial role in cellular processes, development and homeostasis of normal cells. In the pathogenesis of various cancer types and GCs, numerous genes and proteins which were involved in complex signaling cascades are mutated or overexpressed causing deregulation of cellular signal transduction (Harvey 2019).

The aberrant activation and overexpression of protein members of the ERBB-family such as the epidermal growth factor receptor (EGFR) or HER2 promote tumor progression and metastasis of various tumor entities including stomach cancer (Hynes and Lane 2005, Arienti et al. 2019). The activation of the ERBB receptor tyrosine kinases (RTK) stimulates downstream components via the phosphatidylinositol-3 kinase (PI3K)/AKT/mTOR pathway, Ras/mitogen-activated protein kinase (MAPK) pathway or other complex intracellular signaling cascades (Scaltriti and Baselga 2006, Moghbeli et al. 2019). Overexpression of the HER2 protein is used as biomarker for targeted therapy in GC as described previously in paragraph 1.6 (Gravalos and Jimeno 2008, Bang et al. 2010, Dang et al. 2012).

Related genes of the RTK/Ras/MAPK signaling pathway like *FGFR2*, *KRAS*, *EGFR*, *HER2* and *MET* are often amplified in GC especially in highly chromosomal unstable tumor subtypes (Deng et al. 2012, TCGA 2014). Alterations in genes involved in the PI3K/AKT/mTOR signaling pathway play a crucial role in the pathogenesis of GC. The pathway acts as a regulator in cell growths, proliferation, metabolism and angiogenesis and is frequently found to be activated in GC (Matsuoka and Yashiro 2014, Tapia et al. 2014, Riquelme et al. 2015). Overexpression and activating mutations of related components of the pathway like *PIK3CA* (phosphatidylinositol 3-kinase) or *AKT* (serine/threonine kinase) can trigger the aberrant activation of the signal transduction (Cinti et al. 2008, Welker and Kulik 2013, Riquelme et al. 2015).

Cyclin dependent kinases (*CDK4*, *CDK6*) and genes of the cyclin family (*CCND1*, *CCNE1*) which function as regulators of CDK kinases are involved in different cell cycle processes and are often activated or dysregulated in GC leading to cell cycle progression and tumor growth (Collins et al. 1997, Foster 2008, Zhang et al. 2018). Alterations in genes related to stabilization of the cytoskeleton and cell adhesion lead to progression of diffuse type GC. The two most prominent genes which were frequently mutated in diffuse GC are *CDH1* encoding the cell adhesion protein E-cadherin and *RHOA*, a member of the Rho family of small GTP-binding proteins. Different studies identified recurrent somatic mutations of *CDH1* between 32%-37% and of *RHOA* between 14%-23% in diffuse GC patients (Kakiuchi et al. 2014, TCGA 2014, Wang et al. 2014).

Alterations of genes involved in the WNT/ $\beta$ -catenin pathway like *APC* or *CTNNB1* and transforming growth factor (TGF- $\beta$ ) pathway like *SMAD4* or *SMAD2* are also frequently found in GC (Clevers and Nusse 2012, TCGA 2014, Luo et al. 2019). Further identified signaling cascades which are known to be activated and deregulated in GC such as the Notch pathway and JAK/STAT pathway promote tumor cell proliferation (Riquelme et al. 2015). Besides the identification of numerous genetic alterations, a deeper understanding of molecular mechanisms of GC progression and development is necessary to identify novel molecular agents for targeted therapy.

## 1.5. Objectives and study design

The primary aim of this study was to categorize a large cohort of GCs into the four molecular subgroups EBV(+), MSI, GS and CIN according to the classification system of TCGA (2014) and to analyze the clinical relevance of these molecular subgroups in the context of platinum/5-FU based preoperative CTx. Comprehensive molecular analysis of a study population encompassing three different patient cohorts consisting of tumor biopsies before CTx, resected tumors of patients without (non-CTx) and after neoadjuvant CTx was performed to address the following goals:

**(A)** The prognostic relevance should be clarified for non-CTx and CTx patients, **(B)** the predictive impact for CTx patients should be determined using pre-therapeutic tumor biopsies before neoadjuvant platinum/5-FU based CTx, **(C)** tumor cell plasticity should be analyzed by comparing corresponding tumor biopsies before and resected tumors after neoadjuvant CTx and **(D)** additional genetic aspects and characteristics of the molecular subgroups should be obtained by mutational profiling of a subset of the tumors.

The following steps were performed to reach the goals and an overview about the study design is summarized in Figure 7.

1. Assessment of EBV- and MSI-status:

The molecular TCGA subgroups, EBV(+) and MSI, were determined using standardized methods applicable in routine diagnostic settings.

2. Determination of CIN-status:

CIN in the TCGA study (2014) refers to the genome-wide description of SCNA using a cost intensive SNP based microarray technology, which is not suitable for the analysis of large GC cohorts in routine diagnostics. In addition, to date no consistent method exists for the detection of CIN. Therefore, a solid and reliable classification system of CIN should be implemented using previously established microsatellite based multiplex PCR assays which represent a simple and cost-efficient diagnostic tool for the analysis of large GC cohorts suitable in routine diagnostic settings. Results of the multiplex PCR assays were compared to a genome-wide analysis of chromosomal alterations in a subset of the tumors to define a reliable classification of CIN based on both methods. Comprehensive analysis of the different patient cohorts was performed to determine the CIN-status using the microsatellite based multiplex PCR assays and to identify distinct CIN-subgroups.

3. Molecular classification systems and clinical relevance:

The tumors were classified into four molecular subgroups according to the TCGA classification system (TCGA 2014). The prognostic and predictive relevance of the identified molecular subgroups was clarified in the context of perioperative CTx for the different patient cohorts including tumor biopsies before CTx. A detailed modified molecular classification system more suitable for the patient cohorts used in this study was additionally proposed.

4. Tumor cell plasticity:

For the determination of tumor cell plasticity, pre- and post-therapeutic alterations of the results in corresponding tumor biopsies before and resected tumors after CTx were analyzed for allelic imbalance (AI) and CIN.

5. Mutation profiling:

In addition, mutation profiling of a subset of the tumors was performed by next-generation sequencing using a gastric cancer related gene panel to further characterize the molecular subgroups.

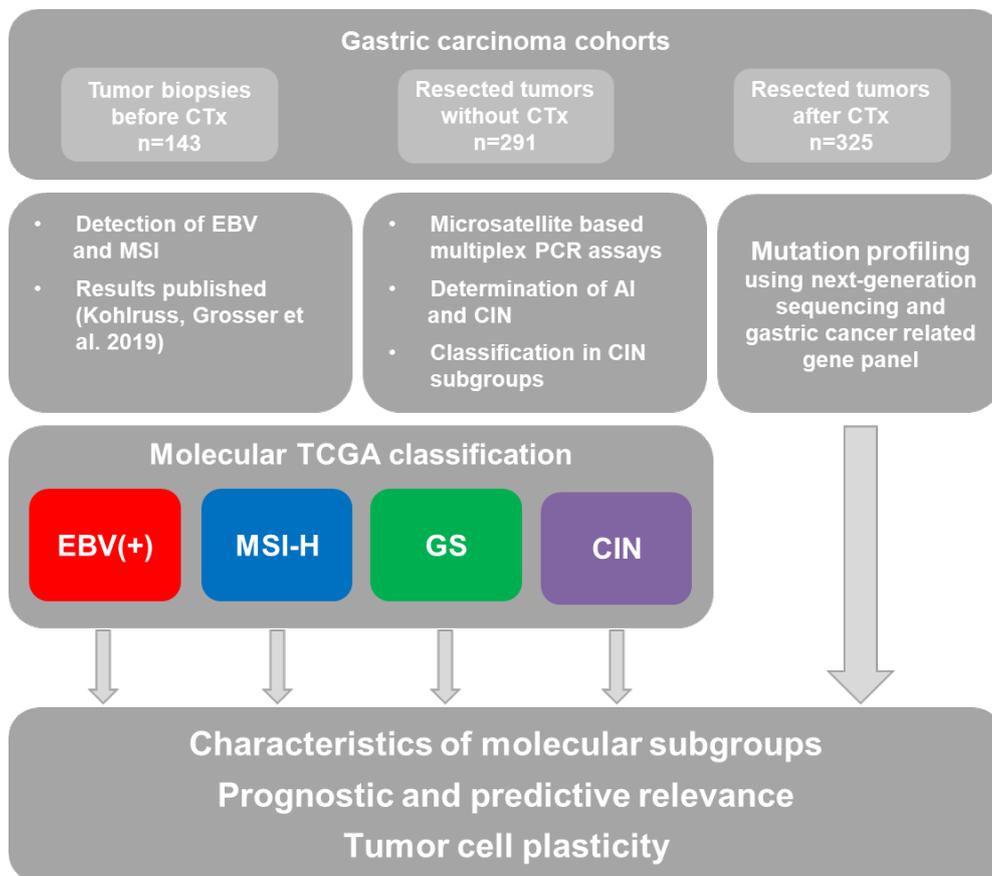


Figure 7: Study design and objectives of the study

## **2. MATERIAL AND METHODS**

### **2.1. Patients and tumor material**

#### **2.1.1. Patients**

The study population comprised a total of 871 patients with gastric carcinomas consisting of 704 patients with resected tumors without or after neoadjuvant CTx and 167 patients with tumor biopsies before neoadjuvant CTx as described previously (Kohlruss et al. 2019). Tumorous and corresponding non-tumorous tissue material from resected tumors from patients with gastric adenocarcinomas including tumors of the gastroesophageal junction (GEJ) with AEG II and AEG III (Siewert and Stein 1998), originated from two clinical institutions. The patients with resected tumors were treated between 2001 and 2013 at the Department of Surgery of the University of Heidelberg and between 2001 and 2012 at the Department of Surgery, Klinikum rechts der Isar, of the Technical University of Munich. The 167 patients with tumor biopsies before CTx were treated between 1993 and 2013 at the Department of Surgery of the Technical University of Munich. For the microsatellite analysis, corresponding non-tumorous tissues for each tumor were required and only paired samples could be used for analysis. The non-tumorous tissues comprised histologically normal mucosa of the stomach or tumor free lymph nodes.

Criteria for inclusion in the molecular analysis was the availability of sufficient DNA or paraffin blocks with tumor and non-tumorous tissues and tumor cell contents with at least 10%. Tumor biopsies of 24 patients and resected tumors from 88 patients had to be excluded consequently from the analysis. The final cohorts consisted of 143 patients with tumor biopsies before CTx and 616 patients with resected tumors comprising 291 tumors from patients treated with surgery alone (non-CTx group) and of 325 tumors from patients after neoadjuvant CTx treatment (CTx group).

The clinical-pathological characteristics of the tumor biopsy cohort were determined within the scope of a master thesis (Krenauer 2018) and published recently (Kohlruss et al. 2019). The clinical data of the resected patient cohort were already described in a previous study (Bauer et al. 2018). The clinical-pathological characteristics of the patient cohorts were summarized in the Results part in Table 19.

The use and analysis of the tissue samples was approved by the local Institutional Review Boards at the Technical University of Munich (reference: 502/15s) and at the University of Heidelberg (reference: 301/2001).

## 2.1.2. Neoadjuvant chemotherapy and surgery

The patients receiving neoadjuvant CTx were treated with platinum (cis- or oxaliplatin) and 5-FU based chemotherapeutic regimens as shown in detail in Supplementary Table 1. The resection of GCs and tumors of the gastroesophageal junction were performed according to the tumor localization and standard approaches (Ott et al. 2009, Blank et al. 2013, Schmidt et al. 2014). Total or subtotal distal gastrectomy with D2 lymphadenectomy was performed for GCs and for tumors with AEG II and AEG III gastrectomy with transhiatal distal abdominal oesophagectomy plus D2 lymphadenectomy was performed as described previously (Kohlruss et al. 2019).

## 2.1.3. Response evaluation

The response to neoadjuvant CTx was determined histopathologically according to a standardized protocol (Becker et al. 2011) and classified into three tumor regression grades (TRG). TRG1 corresponded to total (1a) or subtotal (1b) tumor regression with less than 10% residual tumor cells in the tumor bed; TRG2 to partial tumor regression with 10 to 50% residual tumor cells in the tumor bed and TRG3 to minimal or no tumor regression with more than 50% residual tumor cells in the tumor bed. Patients with TRG1 were classified as responders and those with TRG2 and TRG3 as non-responders. The prognostic relevance of this classification system in GCs was demonstrated in previous studies (Becker et al. 2011, Schmidt et al. 2014).

All three tumor regression grades in the resected specimen were present among the 143 analyzed patients with tumor biopsies before CTx and 45 patients showed TRG1, 34 patients showed TRG2 and 64 patients showed TRG3. Among the 325 analyzed patients with resected tumors after CTx, 153 revealed tumors with TRG2 and 172 with TRG3. Tumors with TRG1 were not included among resected tumors after CTx due to the total absence of residual tumor cells or very low tumor cell contents (<10%) which makes an analysis of these samples impossible. An overview about the different patient cohorts and the respective TRG classification is given in Table 1.

**Table 1: Number of analyzed patients and respective TRG classification**

	Number of analyzed patients	
	<i>n</i>	%
<b>Tumor biopsies before CTx</b>		
TRG1 (Responder)	45	31
TRG2 (Non-responder)	34	24
TRG3 (Non-responder)	64	45
Total	143	100
<b>Resected specimens</b>		
Without neoadjuvant CTx	291	47
After neoadjuvant CTx		
TRG2	153	25
TRG3	172	28
Total	616	100

## 2.1.4. Tumor cell plasticity

Corresponding tumor biopsies before and resected tumors after CTx of 42 patients were included in this study to determine tumor cell plasticity. Pre- and post-therapeutic alterations of the results were analyzed for AI and CIN.

## 2.1.5. Overall survival and follow-up time

The clinical endpoint of the study was overall survival (OS), which was defined as the time between the date of operation and death by any cause. In the tumor biopsy cohort, two patients were not operated and therefore the date of start of CTx was used. The follow-up period was calculated by the inverse Kaplan-Meier method as described previously (Bauer et al. 2018).

## 2.2. Materials

### 2.2.1. Primer and microsatellite marker

All primer pairs were purchased from Eurofins Genomics (Ebersberg, Germany). The primers were purified using high performance liquid chromatography with a synthesis scale of 0.01  $\mu\text{mol}$  and were dissolved in pure  $\text{H}_2\text{O}$  according to the synthesis report of the manufacturer. The concentration of the primer stock solutions was 50  $\mu\text{M}$  or 100  $\mu\text{M}$ . The forward primers were labeled fluorescently at the 5'-end with one of the three dyes HEX, FAM or ATO550. Primers for EBV analysis are shown in Table 2, primers for MSI testing in Table 3 and Table 4 and primers for the preparation of length marker for agarose gel electrophoresis in Table 5 and Table 6. The microsatellite markers included in the multiplex PCR assays for the detection of AI and CIN are listed in Table 7.

#### 2.2.1.1. Primer for EBV analysis

**Table 2: Primer pairs for EBV analysis**

Primer <sup>1</sup>	Forward sequence (5'→3')	Reverse sequence (5'→3')	Size [bp]
BamHI-W	TCGGGCCAGAGGTAAGTGGACTTTAAT	ACCGGTGCCTTCTTAGGAGCTGTC	126
BamHI-K	TCTCTTTTAGTGTGAATCATGTCTGACGA	GCAGCCAATGCAACTTGGACG	264

<sup>1</sup>According to Huber et al. (2002)

A primer mix was prepared containing the two primer pairs each with 10  $\mu\text{M}$  for a total volume of 100  $\mu\text{l}$  in RNase-free  $\text{H}_2\text{O}$ .

### 2.2.1.2. Primer for MSI analysis

**Table 3: Primer pairs for MSI analysis**

Primer <sup>1</sup>	Forward sequence (5'→3')	Dye <sup>2</sup>	Reverse sequence (5'→3')	Size range [bp]
<b>Mononucleotide repeats</b>				
BAT25	TACCAGGTGGCAAAGGGCA	FAM	TCTGCATTTTAACTATGGCTC	153
BAT26	TGACTACTTTTGACTTCAGCC	HEX	AACCATTCAACATTTTAAACCC	116
<b>Dinucleotide repeats</b>				
D2S123 <sup>3</sup>	AAACAGGATGCCTGCCTTTA	FAM	GGACTTTCCACCTATGGGAC	197-227 <sup>4</sup>
D5S346 <sup>3</sup>	ACTCACTCTAGTGATAAATCGGG	FAM	AGCAGATAAGACAGTATTACTAGTT	96-122 <sup>4</sup>
D17S250 <sup>3</sup>	GGAAGAATCAAATAGACAAT	HEX	GCTGGCCATATATATATTTAAACC	151-169 <sup>4</sup>

<sup>1</sup>According to Boland et al. (1998); <sup>2</sup>Forward primers were labeled fluorescently at the 5'-end; <sup>3</sup>Included for AI and CIN analysis; <sup>4</sup>According to human genetic map (Broman et al. 1998)

The stock concentration of the primer pairs was 50 µM for the markers D5S346, D2S123, D17S250, BAT25 and 100 µM for the marker BAT26. A primer mix was prepared containing the five primer pairs each with different concentrations as listed in Table 4 for a total volume of 125 µl in RNase-free H<sub>2</sub>O.

**Table 4: Primer mix for MSI analysis**

Primer pairs	Concentration [µM]	Volume [µl] <sup>1</sup>
D2S123 <sup>2</sup>	1 or 4 <sup>2</sup>	2.5 or 10
D5S346	0.4	1
D17S250	3.2	8
BAT25	1	2.5
BAT26	0.4	0.5

<sup>1</sup>Ad 125 µl RNase-free H<sub>2</sub>O; <sup>2</sup>For DNA derived from highly degraded non-tumorous and tumorous FFPE tissues, the concentration of the primer D2S123 was increased up to 4 µM due to weak amplification intensities

### 2.2.1.3. Primer for preparation of length marker for agarose gel electrophoresis

**Table 5: Primer pairs for preparation of length marker for agarose gel electrophoresis**

Primer <sup>1</sup>	Forward sequence (5'→3')	Reverse sequence (5'→3')	Size [bp]
100	GTTCCAATATGATTCCACCC	CTCCTGGAAGATGGTGATGG	100
BRCA2	TCCTTTCGCCACACTGAGAAATA	CTCTGCCGCCTAGTTTCAGAAG	140
200	AGGTGGAGCGAGGCTAGC	TTTTGCGGTGGAAATGTCCT	200
Globin PC04/GH20	CCACTTCATCCACGTTTACC	GAAGAGCCAAGGACAGGTAC	270

<sup>1</sup>According to Diaz-Cano and Brady (1997)

**Table 6: PCR primer mix for the preparation of length marker for agarose gel electrophoresis**

Primer pairs	Concentration [µM] <sup>1</sup>	Volume [µl]
100	4.4	9
BRCA2	4.4	9
200	6.4	13
Globin PC04/GH20	9.8	20

<sup>1</sup>Primer stock concentration: 100 µM; <sup>2</sup>Ad 204 µl RNase-free H<sub>2</sub>O

## 2.2.1.4. Microsatellite markers included in multiplex PCR assays for the detection of CIN

Table 7: Microsatellite markers included in multiplex PCR assays for the detection of CIN

Marker	Locus	Forward sequence <sup>1</sup> (5'→3')	Dye	Reverse sequence <sup>1</sup> (5'→3')	Size range [bp] <sup>1</sup>
D4S423	4q22	TTGAGTAGTTCCTGAAGCAGC	FAM	CAAAGTCCTCCATCTTGAGTG	103-125
D5S624	5q11	CTATGTAACAAACCTGCATGTTGTG	HEX	ATTTGCTGAACGAATGACCC	146-166
D6S1617	6p25	TGCAAAAACAGGCACACATAC	FAM	TTAATCAATTTTCTGCAAAGATAAA	101-123
D7S492	7q21	ATCTTGGATTTAGGGTTGGC	FAM	GGCTGTGCTCCATCTTCATA	145-155
D7S486	7q31	AAAGGCCAATGGTATATCCC	ATTO550	GGGACTTTTTGAAGGTTTG	133-146
D8S552	8p23	AGGATTGTAATTCCTTGC	FAM	GCCAGGTGATTGATAGTGC	168-182
D8S1793	8q24	TGAGCCGAGTCTTACCAC	FAM	AACAAGTCCAGCTTGTATGAG	113-147
D8S1720	8q24	GTGCCACCTGCCTGAA	FAM	CCACTACATTTTAGAGAGGCCA	130-144
D9S157	9p21	AGCAAGGCAAGCCACATTC	FAM	TGGGGATGCCAGATAACTATATC	133-149
D9S171	9p21	AGCTAAGTGAACCTCATCTGTCT	FAM	ACCCTAGCACTGATGGTATAGTCT	159-177
D12S1682	12p12	GGGACAAGAGTGAGACTTGG	HEX	CCTTTATTGAAGTAAACTGTGAAGC	133-151
D12S1631	12p12	TGGGCTCATCTGGAAA	FAM	GGAGGCAAACACTGATAACTTAC	161-185
D16S507	16q23	GCAGGGGCTAGAAGGTG	HEX	TGTTCCGCTCTTGCAGT	175-195
D17S796	17p13	CAATGGAACCAAATGTGGTC	HEX	AGTCCGATAATGCCAGGATG	144-174
D17S1872	17q12	CCAACCTAGGACTGGGG	HEX	AATTGGGTCCAGAGAGCA	108-140
D17S1861	17q12	AGGGGCAGCAGTCCTGTA	FAM	ACATCATCCTGAAATCTAATGGG	94-116
D18S487	18q21	ACAATCAGAAACCCGCCA	HEX	AGCTGACTTAGGTAGATTTTCCTCG	115-127
D18S1119	18q21	CCTATCGTACATGGTGAGTG	FAM	CTTGATTTGAACCTAATGACG	156-170
D19S875	19q12	TGGTCTGTGATGACTACTACATGC	HEX	AACTTGGTTTATGATGTCTCTTGC	95-123

<sup>1</sup>Primer sequences and size ranges were obtained from the human genetic linkage map (Dib et al. 1996)

## 2.2.2. Commercial kits

Agencourt™ AMPure XP Kit #A63881	Beckman Coulter, Krefeld
EBER1 DNP Probe #760-1209	Ventana, Roche, Basel, Switzerland
GeneScan™ 500 ROX™ dye Size Standard #401734	Applied Biosystems, Thermo Fisher Scientifics, Darmstadt
Ion 510™ & Ion 520™ & Ion 530™ Kit – Chef	Ion Torrent, Thermo Fisher Scientific, Darmstadt
Ion 520™, 530™ or 540™ Chip Kit <sup>1</sup> #A27762, #A27764, #A27766	Ion Torrent, Thermo Fisher Scientific, Darmstadt
Ion AmpliSeq™ Library Kit 2.0 <sup>2</sup> #4475345	Ion Torrent, Thermo Fisher Scientifics, Darmstadt
Ion Library TaqMan™ Quantitation Kit #4468802	Ion Torrent, Thermo Fisher Scientifics, Darmstadt
Ion Xpress™ Barcode Adapters 1–96 <sup>3</sup> #4474517	Ion Torrent, Thermo Fisher Scientifics, Darmstadt
iViEW Blue Plus Detection Kit #760-097	Ventana, Roche, Basel, Switzerland
Maxwell® RSC Blood DNA Kit #AS1610	Promega, Mannheim
QIAamp® DNA FFPE Tissue Kit #56404	Qiagen, Hilden
Qubit™ dsDNA HS Assay Kit #Q32854	Invitrogen, Thermo Fisher Scientifics, Darmstadt
TaqMan™ RNase P Detection Reagents Kit #4316831	Applied Biosystems, Thermo Fisher Scientifics, Darmstadt
TaqMan™ Universal PCR Master Mix #4324018	Applied Biosystems, Thermo Fisher Scientifics, Darmstadt
Type-it® Microsatellite PCR Kit <sup>4</sup> #206243	Qiagen, Hilden

<sup>1</sup>Number of million reads per chip: 3-6 (520 Chip), 15-20 (530 Chip), 60-80 (540 Chip)

<sup>2</sup>Kit contents: 5x Ion AmpliSeq™ HiFi Mix, FuPa Reagent, Switch Solution, DNA Ligase, Low TE buffer

<sup>3</sup>Kit contents: Ion Xpress™ P1 Adapter, Ion Xpress™ Barcode Adapters 1–96

<sup>4</sup>Kit contents: Type-it Multiplex PCR Master Mix (2x), HotStarTaq® Plus DNA Polymerase, Type-it Microsatellite PCR Buffer with 6mM MgCl<sub>2</sub>, dNTPs

### 2.2.3. Instruments

20- and 30-well comb	Bio-Rad, München
Benchtop vortex mixer	VF2, IKA, Staufen
Centrifuge 5417R	Eppendorf, Hamburg
Centrifuge 5427R	Eppendorf, Hamburg
DynaMag™-2 Magnet	Thermo Fisher Scientific, Darmstadt
DynaMag™-96 Side Magnet	Thermo Fisher Scientific, Darmstadt
Gel documentation instrument E-BOX VX2	Vilber, Eberhardzell
Genetic Analyzer 3130, 4-capillary	Applied Biosystems, Darmstadt
Hybex microsample incubator	SciGene, Sunnyvale, USA
Incubator	Memmert, Schwabach
Ion Chef instrument	Ion Torrent, Thermo Fisher Scientific, Darmstadt
Ion S5XL instrument	Ion Torrent, Thermo Fisher Scientific, Darmstadt
Laboratory scale	Sartorius, Göttingen
Magnetic stirrer with heating, IKAMAG Ret	IKA, Staufen
Maxwell® RSC Instrument	Promega, Mannheim
MicroAmp® Adhesive Film Applicator	Thermo Fisher Scientific, Darmstadt
Microscope Labovert FS	Leica, Wetzlar
Microscope M410	Wild Heerbrugg, Heerbrugg, Switzerland
Microwave ME82V	Samsung, Seoul, South Korea
Mini centrifuge	Biozym, Oldendorf
Minishaker MS1	IKA, Staufen
PCR Plate Spinner	VWR, Radnor, USA
Pipettes (0.1-2.5µl, 2-20µl, 10-100µl, 20-200µl, 100-1000µl)	Research® plus, Eppendorf, Hamburg Pipetman, Gilson, Middleton, USA
Pipetting aid accu-jet® pro	Brand, Wertheim
Power Pac 300 Electrophoresis Power Supply	Bio-Rad, München
Qubit™ 3.0 Fluorometer	Thermo Fisher Scientific, Darmstadt
Repetitive pipette HandyStep® electronic	Brand, Wertheim
Spectrophotometer Nanodrop™ 2000c	PeqLab Biotechnologie, Erlangen
StepOnePlus™ Real-Time PCR System	Thermo Fisher Scientific, Darmstadt
Sub-Cell GT UV-Transparent Gel Tray	Bio-Rad, München
Thermal Cycler Mastercycler® gradient	Eppendorf, Hamburg
Thermal Cycler Mastercycler® nexus	Eppendorf, Hamburg
Thermal cycler TRIO	Biometra, Göttingen
Thermomixer compact	Eppendorf, Hamburg
Wide Mini-Sub cell GT electrophoresis cell	Bio-Rad, München

## 2.2.4. Chemicals and reagents

2'-Deoxythymidine 5'-Triphosphate (dTTPs, 100 mM)	Thermo Fisher Scientific, Darmstadt
AmpliTaq Gold® DNA Polymerase with Buffer I, #N8080244	Applied Biosystems, Darmstadt
Biozym LE Agarose, #840004	Biozym, Oldendorf
Bromphenol Blue Xylene Cyanole Dye, #B3269	Sigma-Aldrich, Hamburg
Ethanol (70%, 96%, absolute)	Merck, Darmstadt
Ethanol, absolute, molecular biology grade ( $\geq 99.5\%$ )	Thermo Fisher Scientific, Darmstadt
Ethylenediaminetetraacetic acid (EDTA) pH 8.0, 1 M	Sigma-Aldrich, Hamburg
Gel Loading Dye, Purple 6x, # B7024S	New England Biolabs, Frankfurt a.M.
Gelatin	Merck, Darmstadt
GelRed™ Nucleic Acid Gel Stain, 10.000x in Water, #41003	Biotium, Hayward, USA
Hi-Di™ Formamide, #4311320	Thermo Fisher Scientific, Darmstadt
Isopropyl alcohol (100%)	Merck, Darmstadt
Magnesium chloride (MgCl <sub>2</sub> )	Merck, Darmstadt
Methylene blue solution (Löffler's)	Merck, Darmstadt
Nuclease free H <sub>2</sub> O, #AM9937	Thermo Fisher Scientific, Darmstadt
PCR Nucleotide Mix Plus, # 11888412001	Sigma-Aldrich, Hamburg
Polyethylene glycol sorbitan monolaurate (Tween®-20)	Sigma-Aldrich, Hamburg
POP-7 polymer for DNA analyzers, #4363929	Thermo Fisher Scientific, Darmstadt
Potassium chloride (KCl)	Merck, Darmstadt
Proteinase K (20 mg/ml)	Roche, Basel, Switzerland
RNase, 10 mg/ml	Sigma-Aldrich, Hamburg
RNase/DNase free H <sub>2</sub> O	Qiagen, Hilden
Rotiphorese® 50x TAE Buffer, #CL86.1	Carl Roth, Karlsruhe
Running Buffer 10x with EDTA, #402824	Thermo Fisher Scientific, Darmstadt
Sodium dodecyl sulfate (SDS)	Sigma-Aldrich, Hamburg
Tris(hydroxymethyl)aminomethane hydrochloride (Tris-HCl)	Merck, Darmstadt
UltraPure 10x TBE Puffer, pH 8.4, #15581028	Thermo Fisher Scientific, Darmstadt
Water for chromatography (LC-MS Grade)	Merck, Darmstadt
Xylene	Merck, Darmstadt

## 2.2.5. Buffers and solutions

### 50x TAE running buffer for DNA electrophoresis

Tris	2 M
Acetic acid	1 M
EDTA	50 mM
in distilled deionized H <sub>2</sub> O, pH 8.5, storage at 20°C	

The electrophoresis chamber was filled with 1x TAE running buffer (20 ml 50x TAE buffer in a total volume of 1 L distilled H<sub>2</sub>O).

### 10x DNA loading buffer for agarose gel electrophoresis

Formamid	80 % (v/v)
Bromphenolblau	0,05% (w/v)
Xylencyanol	0,5 % (v/v)

### Proteinase K (PK) buffer for DNA isolation

Tris-HCl, pH 8.3	50 mmol/l
EDTA, pH 8.0	1 mmol/l
Tween-20	0.5%
Proteinase K	0.2 mg/ml

in distilled H<sub>2</sub>O, filtered sterile,  
storage at room temperature

### 10x PCR buffer

Tris-HCl, pH 8.3	100 mM
KCl	500 mM
Gelatin	0.1%
MgCl <sub>2</sub>	30 mM

in distilled H<sub>2</sub>O, storage at -20°C

## 2.2.6. Consumption materials

60-well Microtest Plate, #653190

Filtered pipette tips  
(10 µl, 20 µl, 100 µl, 200 µl, 1000 µl)

MicroAmp® Clear Adhesive Film, #4306311

MicroAmp® Optical 96-Well Reaction Plate,  
#4316813

Microcentrifuge tubes, DNA LoBind  
(0.5 ml, 1.5 ml, 2.0 ml, 5.0 ml)

Microcentrifuge tubes, Safe-Lock PCR clean  
(0.5 ml, 1.5 ml, 2.0 ml)

Microscope slides Superfrost

PCR Tube Strip, 8 Wells, 0.2 ml, #4ti-0792

Precision wipes Kimtech Science

Quali-PCR-plate, 96 well, for ABI 3100,  
#G060/H/1E-OA

Qubit™ Assay Tubes, #Q32856

Scalpel blades, carbon steel

Septa mat for ABI 3100, #G060/S-3100

Serological pipettes, 5ml

TipOne® Repeat Dispenser Tips (0.5 ml, 1 ml)  
#S4751-0050, #S4751-0100

Greiner bio-one, Kremsmünster, Austria

TipOne® Starlab, Hamburg

Corning® DeckWorks, Wiesbaden

Thermo Fisher Scientific, Darmstadt

Thermo Fisher Scientific, Darmstadt

Eppendorf, Hamburg

Eppendorf, Hamburg

Thermo Fisher Scientific, Darmstadt

4titude, Wotton, UK

Kimberley-Clark, Koblenz

Kisker Biotech, Steinfurt

Thermo Fisher Scientific, Darmstadt

Heinz Herenz, Hamburg

Kisker Biotech, Steinfurt

Greiner bio-one, Kremsmünster, Austria

Starlab, Hamburg

## 2.2.7. Software and programs

Adobe Illustrator CC 2018	Adobe Inc., San José, USA
Gene Mapper Software v5	Applied Biosystems, Waltham, USA
Maxwell® RSC software	Promega, Mannheim
Microsoft Office 2010	Microsoft Corporation, Redmond, USA
Nanodrop 2000c	PeqLab Biotechnologie, Erlangen
Nexus Express Software for OncoScan v3.1	Biodiscovery, El Segundo, USA
Prism v8	GraphPad Software, San Diego, USA
R v3.1.0	R Foundation for Statistical Computing, Vienna, Austria
SPSS Statistics v23-25	IBM, Armonk, USA
StepOne™ Software v2.3	Thermo Fisher Scientific, Darmstadt
Torrent Suite Software v5.10.1	Thermo Fisher Scientific, Darmstadt
Build-in plugins:	
Coverage Variant Caller v5.10.1.19	
Coverage Analysis v5.10.0.3	

## 2.2.8. Online platforms

1000 Genomes Project Consortium, The International Genome Sample Resource, (IGSR)	<a href="http://www.internationalgenome.org/">http://www.internationalgenome.org/</a> (Auton et al. 2015)
ANNOVAR	<a href="http://www.openbioinformatics.org/annovar/">http://www.openbioinformatics.org/annovar/</a> (Wang et al. 2010)
Association of Population-based Cancer Registries in Germany, Robert Koch Institute (RKI)	<a href="http://www.krebsdaten.de/english">www.krebsdaten.de/english</a>
cBioPortal	<a href="http://www.cbioportal.org/">http://www.cbioportal.org/</a> (Cerami et al. 2012, Gao et al. 2013)
COSMIC database	<a href="https://cancer.sanger.ac.uk/cosmic">https://cancer.sanger.ac.uk/cosmic</a> (Tate et al. 2019)
Genome browser Ensembl	<a href="http://www.ensembl.org">http://www.ensembl.org</a> (Zerbino et al. 2018)
Global Cancer Observatory GLOBOCAN 2018, IARC, WHO	<a href="https://gco.iarc.fr">https://gco.iarc.fr</a> (Bray et al. 2018)
Human Gene Database GeneCards	<a href="http://www.genecards.org/">http://www.genecards.org/</a> (Stelzer et al. 2016)
IACR TP53 database	<a href="http://p53.iarc.fr/">http://p53.iarc.fr/</a> (Bouaoun et al. 2016)
ICGC Data Portal 6.0.4	<a href="https://dcc.icgc.org/">https://dcc.icgc.org/</a> (Hudson et al. 2010)
Integrative Genomics Viewer (IGV)	<a href="https://www.broadinstitute.org/">https://www.broadinstitute.org/</a> (Robinson et al. 2011, Robinson et al. 2017)
Ion AmpliSeq Designer for NGS panel design	<a href="https://www.ampliseq.com">https://www.ampliseq.com</a> Thermo Fisher Scientific, Darmstadt
NCBI Genome Data Viewer	<a href="https://www.ncbi.nlm.nih.gov/genome/gdv/">https://www.ncbi.nlm.nih.gov/genome/gdv/</a> NCBI, Bethesda, USA
NCBI Map Viewer	<a href="http://www.ncbi.nlm.nih.gov/mapview/">http://www.ncbi.nlm.nih.gov/mapview/</a> NCBI, Bethesda, USA
Single Nucleotide Polymorphism Database: dbSNP (NCBI)	<a href="https://www.ncbi.nlm.nih.gov/snp/">https://www.ncbi.nlm.nih.gov/snp/</a> NCBI, Bethesda, USA
UCSC genome browser	<a href="https://genome.ucsc.edu/">https://genome.ucsc.edu/</a> (Kent et al. 2002)

## **2.3. Methods**

### **2.3.1. DNA isolation and quantification from formalin-fixed paraffin-embedded tissues**

Corresponding tumorous and non-tumorous DNA was isolated from formalin-fixed paraffin-embedded (FFPE) tissue sections. Therefore, tumorous and non-tumorous areas of each patient sample were microscopically identified and the areas of interest were marked on hematoxylin-eosin (H&E) stained slides. All cases were reviewed by a pathologist and the tumor cell content of each tumor sample was estimated. According to the respective size of the tumorous or non-tumorous area, five to ten thin tissue sections with 8 µm thicknesses were prepared and dried at 60°C overnight on an objective slide.

The paraffin was solubilized of the unstained tissue sections by immersion in a xylene bath twice for 10 min and the deparaffinized tissues were washed for rehydration for 5 min in a series of alcohol solutions with decreasing concentrations as follows: 100% isopropyl alcohol, 96% ethanol and 70% ethanol. The tissue sections were stained in methylene blue for 5 sec and were finally washed in a water bath for 2 min to remove xylene and ethanol and kept there till manual microdissection.

In general, DNA isolation was performed using the Maxwell DNA purification system and the QIAamp DNA FFPE Tissue Kit was used for selected cases. Tissue samples with only small amounts, like in the tumor biopsies, were re-suspended after manual microdissection in 200 µl Proteinase K (PK) buffer and were incubated for 3 h at 55°C. Proteinase K was inactivated by boiling for 10 min and 1:5 to 1:10 dilutions were directly used for PCR as described (Keller et al. 1995). Isolated DNA from 278 cases was available from the cooperation partners in Heidelberg and also DNA from tumor biopsies and resected tumors was partially available from previous studies (Ott et al. 2003, Bauer et al. 2018).

#### **2.3.1.1 DNA isolation using the Maxwell DNA purification system**

After deparaffinization the tissue sections were manually microdissected with a scalpel in incubation and lysis buffer (each 100 µl) which were provided in the Maxwell RSC Blood DNA Kit. For protein digestion, 30 µl proteinase K solution (20 mg/ml) was added to each sample tube, incubated in a heating block at 56°C and shaken at 300 rpm overnight. If there was still tissue material left in the solution, 30 µl of proteinase K was added for complete digestion. After complete tissue lysis, DNA was isolated using the Maxwell® RSC automated DNA purification system according to the instructions of the manufacturer (Promega). In brief, the extraction method is based on MagnaCel™ particles and cartridges prefilled with purification reagents. After sample addition, the DNA was bound to the paramagnetic particles and processed further through multiple steps of washes before the DNA was finally eluted. According to the amount of microdissected tumor or non-tumorous tissue areas, DNA was eluted in 60-80 µl elution buffer. If there were still magnetic particles left in the elution tubes after completion of the extraction, the samples were centrifuged for 4 min at 1300 rpm and the supernatant, containing the eluted DNA, was transferred into a new PCR tube.

### **2.3.1.2. DNA isolation using the QIAamp DNA FFPE Tissue Kit**

DNA isolation using the QIAamp DNA FFPE Tissue Kit was required for the genome-wide Oncoscan analysis of 30 samples. The kit uses the QIAamp MinElute technology which is based on silica-based membranes. After deparaffinization of the tissue samples and proteinase K digestion, the DNA was bound to the membrane of the MiniElute column and contaminants were washed away by several washing and centrifugation steps to the final elution of the bound DNA. For this extraction method, the tissue sections were manually microdissected in 180  $\mu$ l ATL tissue lysis buffer, mixed for 10 sec by vortexing and incubated at 98°C for 15 min. After cooling down for 15 min at room temperature, the samples were centrifuged and 20  $\mu$ l proteinase K (20 mg/ml) was added for protein digestion. The samples were incubated at 56°C while shaking till the tissue was completely lysed. If there was still tissue material left in the solution, 20  $\mu$ l of proteinase K was added for complete digestion. The completely lysed samples were incubated at 90°C for 1 h to inactivate proteinase K. To obtain a RNA-free solution, 2  $\mu$ l RNase A (10 mg/ml) was added to each sample and incubated for 2 min at room temperature. The samples were further processed according to the protocol of the QIAamp DNA FFPE Tissue Kit (Qiagen). All centrifugation steps were carried out at room temperature and the bound DNA was finally eluted in 50-100  $\mu$ l ATE buffer.

### **2.3.1.3. Quantification of genomic DNA**

The concentrations of the eluted DNA were determined by a spectrophotometer (Nanodrop 2000c) or a fluorometer using the Qubit™ 3.0 DNA quantitation assay.

For the spectrophotometric measurement, 1.5  $\mu$ l DNA of each sample was pipetted on the measurement pedestal of the Nanodrop instrument and the adsorption was measured at 260 nm and 280 nm. The spectrophotometer provided an assessment of nucleic acids purity and the 260/280 ratio of the adsorption values exhibited the purity of a sample, at which a ratio of around 1.8 was accepted for pure DNA. Absorbance at 230 nm was caused by different reagents used during DNA isolation as carbohydrates, phenol or guanidine and therefore the 260/230 values were a second indicator for pure DNA when the values were between 2.0-2.2 (Wilfinger et al. 1997).

The measurements on the benchtop fluorometer (Qubit® 3.0) were performed using the Qubit dsDNA HS Assay Kit and the samples were prepared according to the manufacturers protocol (Thermo Fisher Scientific). The assay is highly selective for double-stranded DNA. For every new measurement of samples concentration, two standards (0 ng/ $\mu$ l and 10 ng/ $\mu$ l in TE buffer) were prepared for calibrating the fluorometer. Fluorescence versus concentration graphs were displayed and the data points of the two standards were connected. The concentrations of each sample were generated based on the relationship between the two standards using a curve-fitting algorithm and were calculated automatically by the instrument in ng/ml. The concentration values measured with the fluorometer were usually lower than those measured with the spectrophotometer, because specific fluorescence dyes were selectively used for binding at double stranded DNA.

For microsatellite analysis, the DNA stock solutions were diluted to a consistent concentration at 10-20 ng/μl according to the concentration values determined by the Nanodrop instrument and according to Formula 1. For next-generation sequencing, the concentration values determined by the Qubit 3.0 DNA quantitation assay were used.

**Formula 1: Calculation of dilutions with known stock concentration**

$$c_1 \times V_1 = c_2 \times V_2$$

$c_1$ : Concentration of DNA stock solution (ng/μl);  $V_1$ : Start volume of solution (μl);  $c_2$ : Final concentration (ng/μl);  $V_2$ : Final volume of solution (μl)

### 2.3.2. Methods for determination of molecular classification

#### 2.3.2.1. Detection of EBV

Screening for EBV positivity was performed in all tumors by a standardized PCR assay using primers for amplification of EBV specific DNA in the BamHI-W and BamHI-K regions of the virus (Huber et al. 2002). The respective forward and reverse sequences of the primer pairs are listed in Table 2. For the performance of the PCR assay, a reaction mix was prepared according to Table 8. For each tumor sample, the undiluted DNA stock solution and a 1:10 dilution with RNase-free H<sub>2</sub>O was added in a final volume of 12.5 μl, respectively. DNA derived from an EBV associated tumor was used as a positive control and RNase-free H<sub>2</sub>O as a negative control. PCR was performed in a thermal cycler using following temperature steps: after an initial activating step for 10 min at 95°C, 37 cycles were performed consisting of denaturation at 94°C, primer annealing at 65°C and elongation at 72°C each step for 1 min and a final elongation step at 72°C for 7 min.

**Table 8: PCR reaction mix for detection of EBV positivity**

Reagents	Volume per reaction [μl]
RNase-free H <sub>2</sub> O	9
dUTPs/dTTPs <sup>1</sup>	0.625
10x PCR buffer (30 mM MgCl <sub>2</sub> )	1.25
Primer mix (each 10 μM)	0.5
Taq DNA polymerase	0.125
Tumorous DNA <sup>2</sup>	1
Total volume	12.5

<sup>1</sup>Nucleotide mix was composed of dTTPs (100 mM), dATPs, dCTPs, dGTPs (each 10 mM) and dUTPs (30 mM);

<sup>2</sup>Stock concentration of DNA and a 1:10 dilution with RNase-free H<sub>2</sub>O per tumor sample was added, respectively.

The PCR fragments were visualized on a GelRed stained 2% agarose gel using ultraviolet light. For those tumors which revealed positive signals in the PCR assay, chromogenic in situ hybridization using the EBV Early RNA Probe and the iViEW Blue detection kit was additionally performed on an automated system (Ventana Medical System, Roche) according to the instructions of the manufacturer. The in situ hybridization was processed by a technical assistant in the routine-diagnostic laboratory of the Institute of Pathology. Tumors were classified as EBV positive (+) when positive staining after in situ hybridization was present in the nuclei of the tumor cells.

### 2.3.2.2. Analysis for MSI

MSI was determined using the standardized Bethesda marker panel recommended by the National Cancer Institute (Boland et al. 1998). Amplification of tumorous and corresponding non-tumorous DNA was performed for two mononucleotide repeats BAT25 and BAT26 and three dinucleotide repeats D17S250, D2S123 and D5S346 using the Type-it Microsatellite PCR kit (Qiagen). The forward primers were labelled at the 5'-end with the fluorescent dyes FAM or HEX. The respective forward and reverse sequences are listed in Table 3 and the composition of the primer mix in Table 4.

A reaction mix was prepared according to Table 9 and 20 ng of tumorous or non-tumorous DNA of a paired sample was added in a final volume of 25µl. RNase-free H<sub>2</sub>O was used as negative control. PCR cycling was performed in a thermal cycler using following temperature steps: after an initial step of 95°C for 5 min, 32 cycles were performed consisting of denaturation at 95°C for 30 sec, annealing at 58°C for 90 sec and extension at 72°C for 30 sec and final extension at 60°C for 30 min. Amplification control was performed by gel electrophoresis and the PCR products were separated and detected in a capillary sequencer instrument.

**Table 9: PCR reaction mix for MSI analysis**

Reagents	Volume per reaction [µl]
2x Type-it Multiplex PCR Master Mix	12.5
Primer mix <sup>1</sup>	2.5
RNase-free H <sub>2</sub> O	8
Non-tumorous or tumorous DNA (10-20 ng/µl)	2
Total volume	25

<sup>1</sup>Preparation of primer mix for MSI analysis according to Table 4

Tumors with additional alleles at specific microsatellite markers compared to the corresponding normal tissue were classified as MSI. According to a standardized definition, MSI-High (-H) was defined if at least two of the five markers showed MSI and as MSI-Low (-L) if one of the five markers showed MSI (Boland et al. 1998). Tumors without any MSI were classified as microsatellite stable (MSS). All MSI-L cases were confirmed by a second independent PCR.

### 2.3.2.3. Analysis for AI and CIN using microsatellite based multiplex PCR assays

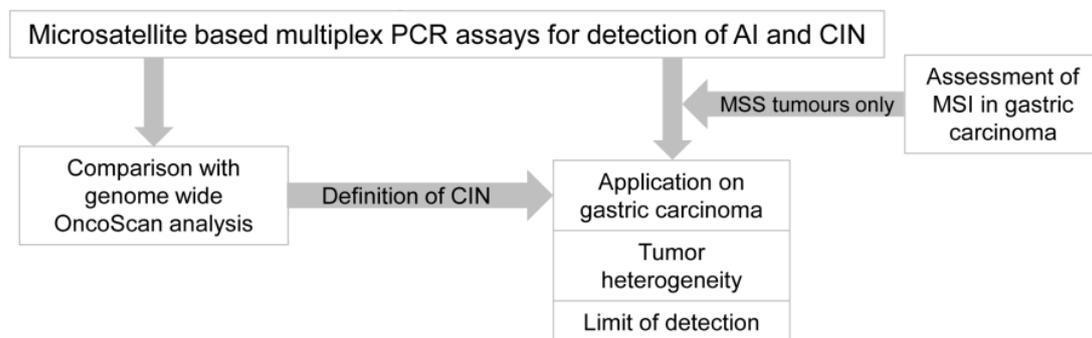
#### 2.3.2.3.1. Study design of the multiplex PCR assays

AI was determined using the microsatellite based multiplex PCR assays established during my master thesis at the Institute of Pathology at the Technical University of Munich (2016). The results of the establishment of the multiplex PCR assays are part of the publication in Scientific Reports (Kohlruss et al. 2018).

The assessment of the limit of detection of AI by the microsatellite based multiplex PCRs and the comparison of the multiplex PCR assays with the genome-wide OncoScan analysis for the determination of CIN were performed during this doctoral thesis and results were jointly published with the results of the establishment of the multiplex PCR assays in Kohlruss et al. (2018).

In the following, criteria for the selected microsatellite markers and the determination of individual cut-off values are summarized in brief. Chromosomal regions which showed various frequencies of arm-level copy number alterations in chromosomal unstable tumors were selected according to TCGA (2014). The chromosomal loci were associated with genes which play a crucial role in various cellular processes during carcinogenesis and were linked to tumor suppressor genes or oncogenes. The selected chromosomal regions and an exemplary tumor-related gene which is located at the respective loci are shown in Supplementary Table 2. Microsatellite markers were selected for each chromosomal locus and multiplex PCR reactions were designed using the Multiplex Manager 1.2 (Holleley and Geerts 2009) and optimized. By analyzing non-tumorous tissues, individual threshold values for every marker were defined for the determination of AI essentially as described (Frigerio et al. 2007).

The multiplex PCR assays and the adaption for the determination of CIN provided a basis for this doctoral thesis for an extensive analysis of AI and CIN and finally for the molecular classification of large cohorts of GCs. Within the scope of this doctoral thesis, the performance of the multiplex PCR assays were initially tested on a small subset of tumors treated with surgery alone and the results determined with the multiplex PCR assays were then compared to a genome-wide analysis of chromosomal alterations to define a classification of CIN based on both methods. Additionally, the established PCR assays were analyzed in relation to tumor heterogeneity and the limit of detection was assessed for reliable CIN detection in the context of the tumor cell content. The workflow of the performance of the microsatellite based multiplex PCRs is shown in Figure 8.



**Figure 8: Workflow for the detection of AI and CIN using microsatellite based multiplex PCR assays** (modified according to Kohlruss et al. (2018))

### 2.3.2.3.2. Composition of the microsatellite based multiplex PCR assays

The multiplex PCR assays encompassed 19 microsatellite markers covering 14 chromosomal regions and were combined in four multiplex PCR reactions. The composition of the four multiplex PCRs and the covered chromosomal regions are summarized in Table 10. The chromosomal regions 8q24, 9p21, 12p12, 17q12 and 18q21 were covered with two microsatellite markers and AI was counted when at least one of the both markers detected AI.

**Table 10: Composition of the multiplex PCR assays for the detection of AI and CIN**

PCR reaction	Chromosomal region	Microsatellite marker	Primer concentration [ $\mu$ M]
multiplex PCR 1	12p12 <sup>1</sup>	D12S1682	4
	16q23	D16S507	6
	8q24 <sup>1</sup>	D8S1793	4
	9p21 <sup>1</sup>	D9S171	1
multiplex PCR 2	19q12	D19S875	2
	17p13	D17S796	2
	4q22	D4S423	2
	9p21 <sup>1</sup>	D9S157	2
	18q21 <sup>1</sup>	D18S1119	8
multiplex PCR 3	18q21 <sup>1</sup>	D18S487	2
	5q11	D5S624	2
	17q12 <sup>1</sup>	D17S1861	2
	8q24 <sup>1</sup>	D8S1720	5
	8p23	D8S552	2
	7q31	D7S486	2
multiplex PCR 4	17q12 <sup>1</sup>	D17S1872	4
	6p25	D6S1617	3
	7q21	D7S492	1
	12p12 <sup>1</sup>	D12S1631	4

<sup>1</sup>Chromosomal regions covered with two microsatellite markers

### 2.3.2.3.3. Marker specific threshold values for definition of AI

DNA from non-tumorous tissues was amplified for the respective microsatellite markers to determine individual cut-off values for the definition of AI as described previously (Kohlruss 2016, Kohlruss et al. 2018). The variation range of the amplification of the alleles of each marker was determined referring to Frigerio et al. (2007). The marker specific cut-off values were estimated from the lower and upper bounds of the bootstrapped two-sided 95% confidence intervals of the 2.5% and 97.5% quantiles, respectively (Efron and Tibshirani 1994, Chernik 2008).

The three dinucleotide markers D2S123, D5S346 and D17S250 of the Bethesda panel used for the analysis of MSI were included in the analysis of AI and individual threshold values were also determined. AI-threshold values of a total of 21 microsatellite markers were calculated in the range between 0.64 and 1.56. The upper and lower cut-offs for the definition of AI of the respective microsatellite markers are listed in Table 11.

**Table 11: Marker specific threshold values for the definition of AI**

Chromosomal region	Microsatellite marker	Lower threshold value <sup>2</sup>	Upper threshold value <sup>2</sup>
2p21	D2S123 <sup>1</sup>	< 0.82	> 1.29
4q22	D4S423	< 0.77	> 1.43
5q11	D5S624	< 0.67	> 1.34
5q21	D5S346 <sup>1</sup>	< 0.67	> 1.18
6p25	D6S1617	< 0.81	> 1.33
7q21	D7S492	< 0.68	> 1.26
7q31	D7S486	< 0.65	> 1.56
8p23	D8S552	< 0.66	> 1.24
8q24	D8S1793	< 0.71	> 1.43
8q24	D8S1720	< 0.73	> 1.34
9p21	D9S157	< 0.88	> 1.16
9p21	D9S171	< 0.83	> 1.18
12p12	D12S1682	< 0.80	> 1.17
12p12	D12S1631	< 0.77	> 1.26
16q23	D16S507	< 0.67	> 1.39
17p13	D17S796	< 0.69	> 1.30
17q12	D17S1861	< 0.78	> 1.30
17q12	D17S1872	< 0.80	> 1.37
17q21	D17S250 <sup>1</sup>	< 0.64	> 1.45
18q21	D18S487	< 0.81	> 1.20
18q21	D18S1119	< 0.74	> 1.38
19q12	D19S875	< 0.75	> 1.49

<sup>1</sup>Dinucleotide markers of the Bethesda panel used for MSI analysis (Boland et al. 1998); <sup>2</sup>Lower and upper bounds of the bootstrapped two-sided 95% confidence intervals of the 2.5% and 97.5% quantiles.

#### 2.3.2.3.4. Microsatellite based multiplex PCRs

Amplification of the microsatellite markers was performed in a thermal cycler using the Type-it Microsatellite PCR kit (Qiagen). A reaction mix was prepared per sample according to the protocol of the manufacturer and as shown in Table 12. 20 ng of the template DNA was added to each PCR reaction to a final volume of 25 µl. RNase-free H<sub>2</sub>O was used as negative control.

The cycle conditions for the amplification of the microsatellite markers were as follows: after an initial activation step of 95°C for 5 min to activate the HotStar Taq *Plus* DNA Polymerase, 32 cycles were performed consisting of denaturation at 95°C for 30 sec, primer annealing at 58°C for 90 sec and extension at 72°C for 30 sec, followed by a final extension step at 60°C for 30 min. Amplification control was performed by agarose gel electrophoresis and the PCR products were further analyzed using a capillary sequencer instrument.

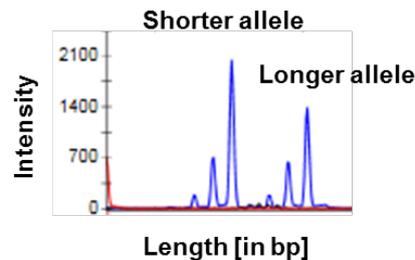
**Table 12: Reaction mix for microsatellite based multiplex PCRs for AI and CIN analysis**

Reagents	Volume per reaction [µl]	Concentration
2x Type-it Multiplex PCR Master Mix (Qiagen)	12.5	1x
Primer mix	2.5	Variable <sup>1</sup>
RNase-free H <sub>2</sub> O	8	-
Template DNA	2	20 ng
Total volume	25	

<sup>1</sup>Primer concentrations per multiplex PCR reaction are listed in Table 10

### 2.3.2.3.5. Detection and evaluation of AI

The microsatellite based multiplex PCR assays were utilized for the analysis of AI and CIN in the tumor cohorts using DNA of matched non-tumorous and tumorous tissues from one patient. Allele ratios of every heterozygous marker of the normal (N) and tumor DNA (T) were calculated per patient by dividing the peak area of the shorter allele by the peak area of the longer allele (Figure 9). According to Formula 2, AI-values were calculated by dividing the allele ratios of the normal DNA by the matched tumor DNA as reported previously (Beckmann et al. 1996, Ott et al. 2003).



**Figure 9: Calculation of allele ratios of a heterozygous marker**

#### Formula 2: Determination of AI-values

$$AI - value = \frac{Peak\ area\ (N,\ shorter\ allele)}{Peak\ area\ (N,\ longer\ allele)} \times \frac{Peak\ area\ (T,\ longer\ allele)}{Peak\ area\ (T,\ shorter\ allele)}$$

The frequency of AI at a respective chromosomal region and the AI ratio per tumor was determined according to Formula 3 and Formula 4 as described previously (Foster et al. 2015, Kohlruss et al. 2018).

#### Formula 3: Determination of frequency of AI at a respective chromosomal region

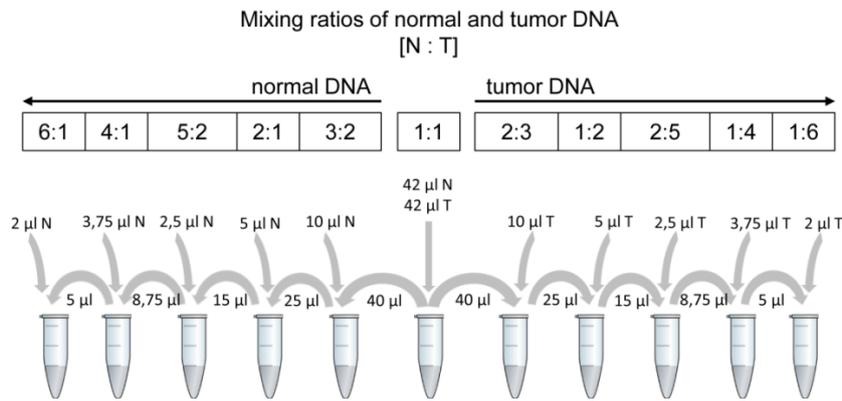
$$Frequency\ of\ AI\ (\%) = \frac{Number\ of\ tumors\ with\ AI}{Number\ of\ informative\ tumors} \times 100$$

#### Formula 4: Definition of AI ratio per tumor

$$AI\ ratio = \frac{Number\ of\ markers\ with\ AI}{Number\ of\ informative\ markers}$$

### 2.3.2.3.6. Limit assessment of AI detection by microsatellite based multiplex PCR assays

Tumors are usually a mixture of normal and cancer cells which can affect the detection of AI. Therefore, the limit of detection of the established multiplex PCR assays was determined by dilution experiments. The initial tumor cell contents of four primary resected tumors were determined by a pathologist prior to the analysis of AI and DNA from the tumors was mixed with the corresponding normal DNA in 11 dilution steps as shown in Figure 10. The resulting tumor cell contents of each mixture were estimated. Amplification of the mixtures of normal and tumor DNA was performed for the microsatellite markers and AI ratios were determined for every mixing ratio as described previously.



**Figure 10: Dilution experiments to assess the limitation of AI detection by the multiplex PCR assays**

#### 2.3.2.4. Gel electrophoresis for PCR quality control

After PCR cycling, amplification control was performed by 2%-agarose gel electrophoresis. For the preparation of the gel, agarose was dissolved in 1x concentrated TAE buffer by heating for 2-3 min and 5 µl of the nucleic acid dye GelRed was added to the solution while stirring the components. After cooling down for 15 min, the solution was poured in a gel tray with two installed 20- or 30-well combs. After 30 min the gel has set completely and the combs were removed. 4 µl of each PCR product was mixed with 1 µl of 10x loading buffer and pipetted into the respective wells on the gel. To identify the size of the separated DNA fragments, 5 µl of a DNA length marker with known fragments sizes was added into the first well. The electrophoresis was run for 25 min at 130V. The PCR fragments were visualized in a gel documentation instrument using ultraviolet light and were diluted with deionized water according to the intensity of the amplification for further analysis using a capillary sequencer in the range from 1:5 to 1:80.

The preparation of the length marker was performed using four primer pairs detecting fragments of 100 bp, 140 bp, 200 bp and 270 bp in a multiplex PCR. The respective forward and reverse sequences of the primers are listed in Table 5. DNA from blood lymphocytes was amplified for the four primers and a reaction mix was prepared according to Table 13. 20 ng DNA from blood lymphocytes was added to each PCR reaction to a final volume of 25 µl. The cycling conditions are shown in Table 14. After PCR cycling, the single reactions of amplicates (each 25 µl) were combined in one solution and 1 µl of Gel Loading 6x Purple dye was added to every 5 µl of the amplicate. The PCR quality was reviewed by gel electrophoresis.

**Table 13: PCR reaction mix for the preparation of length marker for agarose gel electrophoresis**

Reagents	Volume per reaction [ $\mu$ l]
RNase-free water	18
dUTPs/dTTPs <sup>1</sup>	1.25
10x PCR buffer (30 mM MgCl <sub>2</sub> )	2.5
Primer mix <sup>2</sup>	1
Taq DNA polymerase	0.25
DNA from blood lymphocytes (10 ng/ $\mu$ l)	2
Total volume	25

<sup>1</sup>Nucleotide mixture was composed of dTTPs (100 mM), dATPs, dCTPs, dGTPs (each 10 mM) and dUTPs (30 mM); <sup>2</sup>Preparation of primer mix according to Table 5

**Table 14: Cycling protocol for the preparation of length marker for agarose gel electrophoresis**

Cycling step	Temperature [ $^{\circ}$ C]	Time [min]	Cycles
Activation	95	10	
Denaturation	94	0.5	37
Annealing	56	0.5	
Extension	72	0.5	
Final extension	72	7	
Cooling step	10	$\infty$	

### 2.3.2.5. Fragment analysis by capillary electrophoresis

Separation and detection of the PCR products were performed in a capillary sequencer instrument equipped with a 36 cm four-capillary array loaded with Pop-7 polymer. Four samples were simultaneously injected into a capillary array. By applying a high voltage charge, the negatively charged DNA fragments were separated by their size while wandering through the polymer in the capillary to the positive electrode. The fluorescently labeled fragments were detected by a laser beam and the fluorescent signals were transformed into electropherograms. Several chromosomal regions could be detected simultaneously in one capillary injection, because each fluorescent dye emitted light at a different wavelength. ROX-500 Genescan was used as size standard. The PCR products were diluted with deionized water in the range of 1:5 to 1:80 according to their intensity of amplification visualized on the agarose electrophoresis gel.

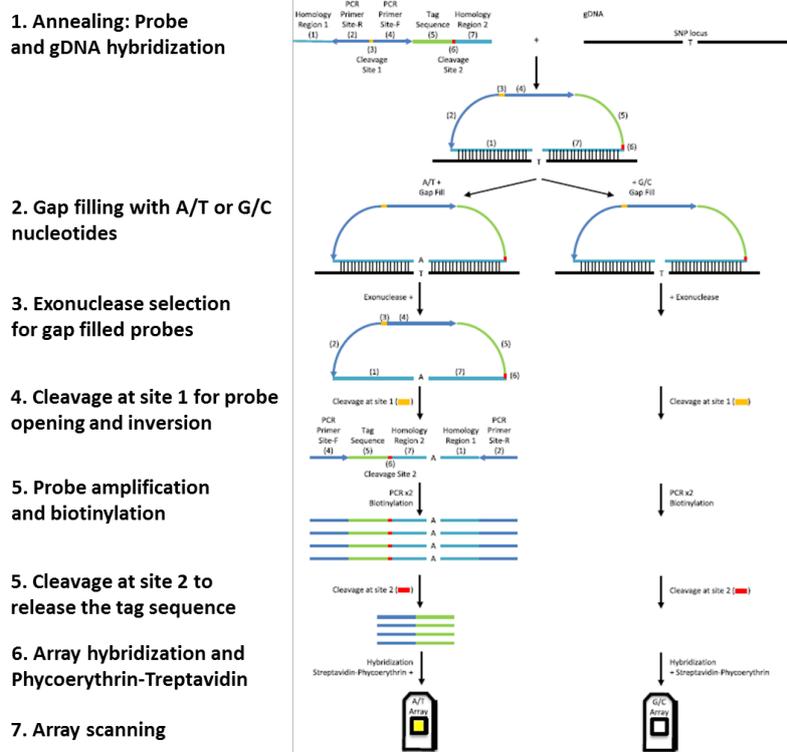
A mixture of formamid (15.25  $\mu$ l) and size standard (0.25  $\mu$ l) was prepared per sample and a total volume of 15.5  $\mu$ l was added per well into a 96-well plate. The diluted PCR products (5  $\mu$ l) were added to the respective wells and a denaturation step was performed for 2 min at 95 $^{\circ}$ C in formamid. All samples were analyzed with the GeneMapper Software v5.

## 2.3.3. Genome-wide OncoScan analysis and CIN classification

### 2.3.3.1. OncoScan technology

A subset of tumors of 30 primary resected patients was analyzed for genome-wide copy number variations (CNV) using the Affymetrix OncoScan® FFPE Assay, which also detects loss of heterozygosity (LOH), AI and cancer-related mutations (Foster et al. 2015). The single nucleotide polymorphism (SNP) based array technology is based on molecular inversion probes (MIP) which is optimized for highly degraded FFPE samples with a probe interrogation site of 40 bp. The tumor samples (80 ng input DNA) were processed by IMG M Laboratories GmbH (Martinsried) according to the protocol of the manufacturer (Affymetrix) and as described previously (Kohlruss et al. 2018).

In brief, MIPs were hybridized to complementary DNA fragments leaving a gap at the SNP position of interest (Figure 11). By addition of a gap filling enzyme, ligase and nucleotide complementary to the locus being interrogated, the MIPs were circularized. This procedure was performed in two separate reactions for A/T and G/C nucleotides. After digestion of linear DNA fragments, the circularized probes underwent an intra-molecular rearrangement after cleavage by *HaeIII* and were amplified by PCR using common primers. The resulting probes were biotinylated and hybridized to separate microarrays for A/T and G/C signals per sample as shown in Figure 11. The fluorescence intensity data were detected with the GeneChip scanner (Affymetrix) and processed to intensity (.cel) files which were converted to .OSCHP files using OncoScan Console 1.3 (Affymetrix).

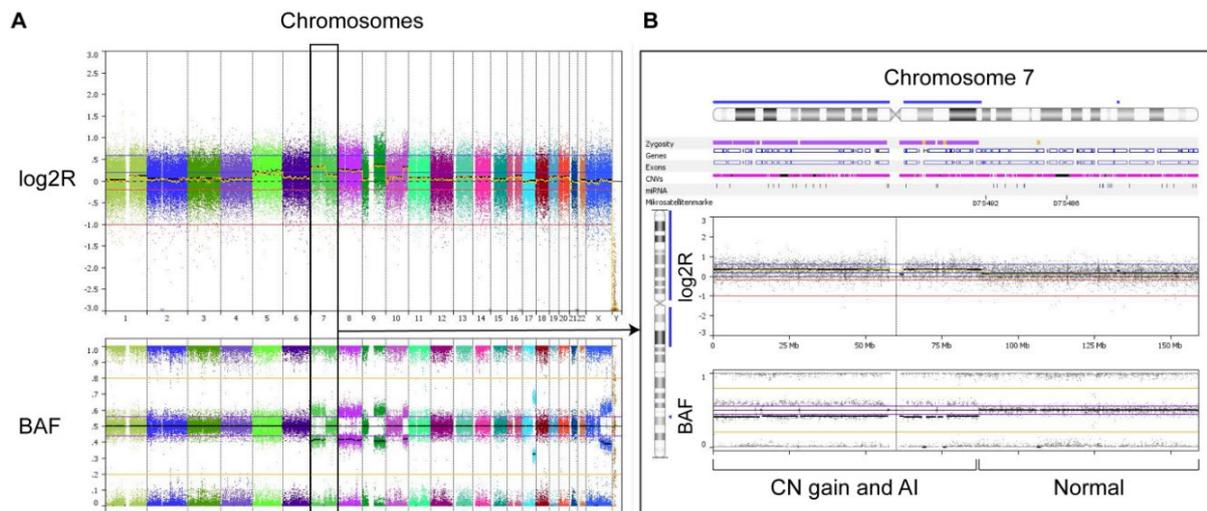


**Figure 11: Molecular inversion probe technology of the Affymetrix OncoScan® FFPE Assay** (modified according to Jung et al. (2017))

### 2.3.3.2. Analysis and evaluation of genome-wide copy number variants

A set of quality metric parameters, normalized log intensity ratios (sample/reference) and B-allele frequencies (BAF) were generated. Copy number aberrations of the samples were analyzed based on the log intensity ratio (log2R) using the SNP-FASST2 algorithm implemented in the Nexus Express Software for OncoScan v3.1. The algorithm generates segmentation calls based on both the log2R and BAF data. The significance threshold for segmentation was set according to Supplementary Table 3 and as described previously (Hardiman et al. 2016). By default, each sample was centered to the median log2R by the Nexus software automatically. Whole genome plots of all samples were visual inspected and manual re-centering was performed if log2R indicates losses or gains but the BAF plot showed a normal three band pattern.

The Nexus Express software displays copy numbers (CN) in log2R plots and allele frequencies in BAF plots per chromosome and sample (Figure 12A). In the log2R plot each dot represents copy number values which are calculated from signal intensities of a tumor sample compared to a normal reference and the BAF plot generates allelic information at a respective SNP position (Jung et al. 2017). For more detailed information, the software displays CN and BAF plots for each chromosomal arm (Figure 12B). In tumor samples which show zero values in the log2R and a normal three band pattern in the BAF plot at a specific chromosomal region, no AI and no CNV were indicated (Figure 12B and Figure 13A).

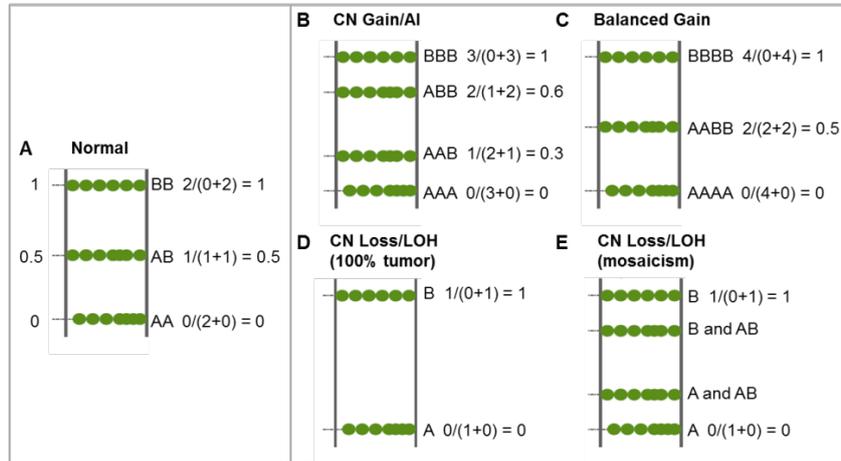


**Figure 12: Analysis of copy number variants and AI using Nexus Express software**

Whole genome plots (A) and detailed information at chromosome 7 (B) of copy numbers (log2R) and allele frequencies (BAF) are displayed.

For the purpose of this study, copy number gains (CNG) and losses (CNL) were evaluated and allelic calls were also included in the analysis due to the fact that AI can represent gains or losses which in some cases could not be unequivocally identified by the OncoScan assay and which cannot be distinguished per se by the microsatellite based multiplex PCR assays. BAF was calculated by dividing the number of minor [B] alleles by the sum of major [A] and minor [B] alleles as shown in Figure 13 and as described previously (Alkan et al. 2011).

Chromosomal regions with CNG or AI show a clear four-band pattern whereas regions with balanced gains reveal a normal three-band pattern (Figure 13B, C). A two-band pattern indicates a CNL in a sample with 100% tumor cells (Figure 13D). Due to the fact that tumors usually contain a mixture of cells with different copy number states, mosaicism BAF plots occur more frequently (Figure 13E).



**Figure 13: Determination of B-allele frequency**

BAF is calculated by dividing the number of minor [B] alleles by the sum of major [A] and minor [B] alleles. A normal chromosomal region (A), chromosomal region with copy number gain or AI (B), with balanced gain (C), with copy number loss of 100% tumor cells (D) and with a mosaicism copy number loss (E) is shown (modified according to the manufacturer Affymetrix).

First, AI and copy number gains and losses were estimated for every single chromosomal p- or q- arm per tumor. The frequency of alterations per chromosomal arm was calculated by dividing the length of the particular alteration of the p- or q-arm through the total length of the respective arm according to Formula 5.

**Formula 5: Calculation of frequency of alterations per chromosomal arm**

$$\text{Frequency of alteration per chromosomal arm} = \frac{\text{Length of alteration [bp]}}{\text{Total length [bp]}}$$

Next, the genome wide extent of alterations affecting all chromosomal arms was determined per tumor as indicated in Formula 6. The occurrence of chromosomal alterations determined by the OncoScan assay was compared to the results of AI at the respective microsatellite locus determined by the microsatellite based multiplex PCR assays.

**Formula 6: Ratio of chromosomal alterations per tumor**

$$\text{Ratio of chromosomal alterations per tumor} = \frac{\text{Number of altered chromosomal arms}}{\text{Total number of chromosomal arms}} \quad (n = 36)$$

### **2.3.3.3. OncoScan and TCGA based CIN classification system**

A determination of CIN according to the definition of TCGA and based on the genome wide OncoScan and microsatellite based multiplex PCR analysis was performed. In line with the TCGA data, a chromosomal arm was considered to be altered if at least 80% of the arm demonstrated gains, losses or AI (TCGA 2017). Tumors were classified as having high chromosomal instability (CIN-High) if they showed at least one altered chromosomal arm, except for chromosome 18 and 21q (TCGA 2017). Accordingly, the 30 tumors analyzed with the OncoScan assay were classified as CIN-Low or CIN-High and a corresponding threshold value of the AI ratios determined with the microsatellite assays was searched.

### **2.3.4. Assessment of intra-tumor heterogeneity**

The performance and robustness of the microsatellite based CIN classification was analyzed in relation to tumor heterogeneity. Therefore, four tumors of patients treated without CTx and five tumors of patients after neoadjuvant CTx were used and the intra-tumor variability was assessed of five tumor areas from each patient. AI ratios were determined by the microsatellite based multiplex PCR assays for every tumor area as described previously. A crossing probability was calculated from a data set comprising a total of 45 tumor areas according to Formula 8 shown in the part of statistical analysis. The five areas of each tumor were selected in a previous study and represented central parts of the tumor as well as areas located proximal, distal and/or lateral to the center near the respective tumor margins (Bauer et al. 2018).

### **2.3.5. Next-generation sequencing**

For mutation profiling and analysis of copy number variations, next-generation sequencing (NGS) using the Ion Torrent platform (Thermo Fisher Scientific) was performed in a subset of the tumors. The sequencing technology is based on the detection of released hydrogen ions ( $H^+$ ) when a complementary nucleotide is incorporated into a DNA strand while polymerization (Rothberg et al. 2011, Merriman and Rothberg 2012). The semiconductor chips are flooded sequentially with deoxyribonucleoside triphosphates (dNTPs) and an ion sensor detects the change of the pH of the solution when a dNTP is complementary to the template strand and a proton is released. To perform sequencing, the DNA templates were attached to a solid surface of a bead and amplified clonally by emulsion PCR (Garrido-Cardenas et al. 2017).

#### **2.3.5.1. Gastric cancer panel design for targeted sequencing**

An Ion Ampliseq™ custom panel was designed specifically for targeted sequencing in gastric carcinoma in close collaboration with Prof. W. Weichert, Dipl. Biol. N. Pfarr, Prof. G. Keller and Prof. B. Luber at the Institute of Pathology of the Technical University of Munich.

First, literature research was performed for the selection of the most common gastric cancer related genes on the basis of copy number alteration data.

The data referred either to the respective original study or to the published data available on the online platform cBio Portal (Cerami et al. 2012, Gao et al. 2013). A list of published studies which were considered for the selection of the gastric cancer related genes is given in Supplementary Table 4. The mutation rate of every single selected gene and the genes catalogued in the Cancer Gene Census in gastric carcinoma were reviewed using the COSMIC (catalog of somatic mutations in cancer) database (Sondka et al. 2018, Tate et al. 2019). Criteria for the selection of the gastric cancer related genes were mutation rates above 3-5% in most cases, alterations of chromosomal regions covering distinguished genes  $\geq 3\%$  and applicability of drugs. In total, 58 gastric cancer related genes were selected for panel design (Table 15). Next, relevant regions of the selected genes which exhibited mutations or chromosomal alterations were selected using the platform cBio Portal for validation and visualization. Two data sets of different studies with a total of 395 analyzed samples with gastric adenocarcinomas were used to review these regions and the respective amino acid position of the mutation was selected (TCGA 2014, Wang et al. 2014).

Additionally, recurrent hotspot mutations were checked in the COSMIC database and included for the selection. The positions of the chromosomal regions were exactly determined by the UCSC browser using the human reference genome assembly GRCh37/hg19 and the European server (Kent et al. 2002). The whole region of an exon was selected due to the high number of mutations. If only one GC related mutation or few mutations were found, this part of the exon was selected. Initially, 316 chromosomal regions with an average length of 70-80 bp were selected for amplicon design. The gastric cancer panel (GC-panel) was designed using the Ion AmpliSeq™ Designer (Thermo Fisher Scientific) and consisted of four primer pools encompassing a total of 525 amplicons of 58 gastric cancer related genes (Table 15).

**Table 15: Gastric cancer related genes included in the GC-panel**

Gene <sup>1</sup>	Gene description <sup>1</sup>	Chr	Frequency of genetic alterations [%] <sup>2</sup>	Number of amplicons	RefSeq Number
<i>ACVR2A</i>	Activin A Receptor Type 2A	2	5	6	NM_001616
<i>AKT1</i>	V-Akt Murine Thymoma Viral Oncogene Homolog 1	14	1.8	1	NM_005163
<i>AKT2</i>	V-Akt Murine Thymoma Viral Oncogene Homolog 2	19	5	5	NM_001616
<i>AKT3</i>	V-Akt Murine Thymoma Viral Oncogene Homolog 3	1	2.8	6	NM_005465
<i>ALK</i>	Anaplastic Lymphoma Receptor Tyrosine Kinase	2	4	6	NM_004304
<i>APC</i>	Adenomatous Polyposis Coli	5	14	10	NM_001127510
<i>ARID1A</i>	AT-Rich Interaction Domain 1A	1	29	17	NM_006015
<i>ATM</i>	ATM Serine/Threonine Kinase	11	10	14	NM_000051
<i>BRAF</i>	B-Raf Proto-Oncogene, Serine/Threonine Kinase	7	5	4	NM_004333
<i>CCND1</i>	Cyclin D1	11	5	5	NM_053056
<i>CCNE1</i>	Cyclin E1	19	9	4	NM_001238
<i>CD274 (PDL1)</i>	CD274 Molecule, PDL-1	9	6	4	NM_014143
<i>CDH1</i>	Cadherin 1	16	12	42	NM_004360
<i>CDK6</i>	Cyclin-Dependent Kinase 6	7	7	3	NM_001259
<i>CDKN2A</i>	Cyclin-Dependent Kinase Inhibitor 2A (P16-INK4A)	9	26	5	NM_000077
<i>CTNNA1</i>	Catenin Alpha 1	5	5	8	NM_001903

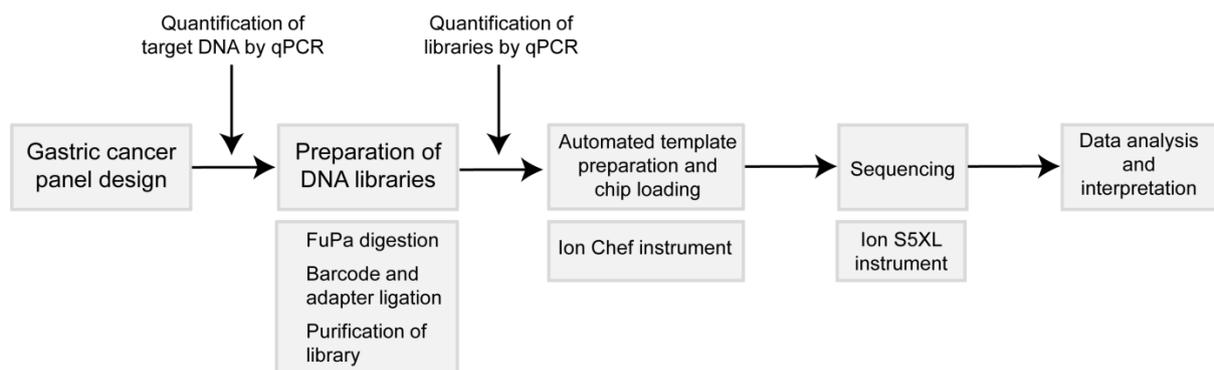
<i>CTNNB1</i>	Catenin Beta 1	3	7	6	NM_001904
<i>EGFR</i>	Epidermal Growth Factor Receptor, EGFR	7	9	13	NM_005228
<i>HER2</i>	Erb-B2 Receptor Tyrosine Kinase 2, HER2	17	14	10	NM_004448
<i>ERBB3</i>	Erb-B2 Receptor Tyrosine Kinase 3	12	12	10	NM_001982
<i>ERBB4</i>	Erb-B2 Receptor Tyrosine Kinase 4	2	11	9	NM_005235
<i>FBXW7</i>	F-Box And WD Repeat Domain Containing 7	4	10	10	NM_033632
<i>FGFR1</i>	Fibroblast Growth Factor Receptor 1	8	5	4	NM_023110
<i>FGFR2</i>	Fibroblast Growth Factor Receptor 2	10	7	5	NM_000141
<i>FGFR3</i>	Fibroblast Growth Factor Receptor 3	4	3	10	NM_000142
<i>FHIT</i>	Fragile Histidine Triad	3	7	4	NM_002012
<i>KRAS</i>	Kirsten Rat Sarcoma Viral Oncogene Homolog	12	13	8	NM_033360
<i>MAP2K7</i>	Mitogen-Activated Protein Kinase Kinase 7	19	6	12	NM_145185
<i>MAP3K4</i>	Mitogen-Activated Protein Kinase Kinase Kinase 4	6	5	8	NM_005922
<i>MAP3K6</i>	Mitogen-Activated Protein Kinase Kinase Kinase 6	1	3	8	NM_004672
<i>MDM2</i>	MDM2 Proto-Oncogene	12	4	4	NM_002392
<i>MET</i>	MET Proto-Oncogene, Receptor Tyrosine Kinase	7	5	9	NM_001127500
<i>MTOR</i>	Mechanistic Target Of Rapamycin	1	8	21	NM_004958
<i>MUC6</i>	Mucin 6	11	10	25	NM_005961
<i>MYC</i>	V-Myc Avian Myelocytomatosis Viral Oncogene Homolog	8	11	7	NM_002467
<i>NRAS</i>	Neuroblastoma RAS Viral Oncogene Homolog	1	2.8	3	NM_002524
<i>PIK3CA</i>	Phosphatidylinositol-4,5-Bisphosphate 3-Kinase Catalytic Subunit Alpha	3	18	16	NM_006218
<i>PIK3R1</i>	Phosphoinositide-3-Kinase Regulatory Subunit 1	5	5	7	NM_181523
<i>PREX2</i>	Phosphatidylinositol-3,4,5-Trisphosphate Dependent Rac Exchange Factor 2	8	16	12	NM_024870
<i>PTEN</i>	Phosphatase And Tensin Homolog	10	10	8	NM_000314
<i>PTPRT</i>	Protein Tyrosine Phosphatase, Receptor Type T	20	12	10	NM_133170
<i>RBI</i>	Retinoblastoma 1	13	6	6	NM_000321
<i>RAC1</i>	Rac Family Small GTPase 1	7	4	2	NM_005908
<i>RHOA</i>	Ras Homolog Family Member A	3	7	3	NM_001664
<i>ROS1</i>	ROS Proto-Oncogene 1, Receptor Tyrosine Kinase	6	7	9	NM_002944
<i>SMAD2</i>	SMAD Family Member 2	18	4	6	NM_005901
<i>SMAD4</i>	SMAD Family Member 4	18	11	6	NM_005359
<i>SMARCA4</i>	SWI/SNF Related, Matrix Associated, Actin Dependent Regulator Of Chromatin, Subfamily A, Member 4	19	8	7	NM_001128849
<i>SOX11</i>	SRY-Box 11	2	2.3	3	NM_003108
<i>SRC</i>	SRC Proto-Oncogene, Non-Receptor Tyrosine Kinase	20	2.8	4	NM_005417
<i>STK3</i>	Serine/Threonine Kinase 3	8	5	8	NM_006281
<i>TGFBR2</i>	Transforming Growth Factor Beta Receptor 2	3	8	12	NM_001024847
<i>TLR4</i>	Toll Like Receptor 4	9	8	11	NM_138554
<i>TP53</i>	Tumor Protein P53	17	49	22	NM_000546.5
<i>VEGFA</i>	Vascular Endothelial Growth Factor A	6	5	5	NM_001171623
<i>WWOX</i>	WW Domain Containing Oxidoreductase	16	7	4	NM_016373
<i>XIRP2</i>	Xin Actin Binding Repeat Containing 2	2	15	8	NM_152381.5
<i>RNF43</i>	Ring Finger Protein 43	17	16	30	NM_017763

<sup>1</sup>The gene symbols and descriptions were obtained from GeneCards, the human gene database (Stelzer et al. 2016);  
<sup>2</sup>Two data sets of different studies with a total of 395 analyzed samples with gastric adenocarcinomas were analyzed (TCGA 2014, Wang et al. 2014) using cBioPortal (Cerami et al. 2012, Gao et al. 2013). Mutations and copy number variants were included.

## 2.3.5.2. Massive parallel sequencing with 4-pool panel

### 2.3.5.2.1. Workflow for next-generation sequencing

Sequencing of 52 primarily resected tumors and 8 non-tumorous samples was performed using the designed GC-panel consisting of four primer pools yielding 525 amplicons as described previously (Pfarr et al. 2017a). The multiplex PCR based Ion AmpliSeq™ targeted sequencing technology (Thermo Fisher Scientific) was used for library construction. An overview of the workflow is shown in Figure 14. A first quality check prior to the library preparation was performed using a quantitative real-time PCR (qPCR) assay to quantify the amount of amplifiable DNA. After the amplification with the GC-panel, the amplicons were partially digested using the FuPa reagent, followed by barcode and adapter ligation and purification of the final libraries. A second quality check was performed for the quantification of the final libraries by qPCR. Automated template preparation and chip loading was performed using the Ion Chef System and libraries were further processed for sequencing using the semiconductor technology of the Ion S5XL instrument. Data were interpreted and visualized using different analysis tools and platforms.



**Figure 14: Next-generation sequencing workflow using the Ion Torrent technology**

### 2.3.5.2.2. Quantification of amplifiable DNA by qPCR using RNase P Detection Assay

DNA derived from FFPE tissue material, as used in this study, is often highly degraded and therefore the amount of amplifiable DNA was determined using a qPCR assay. This quality step was performed prior to the preparation of the DNA libraries. The detection and quantitation of genomic copies of the human RNase P gene was determined by qPCR with the TaqMan™ RNase P Detection Assay. A standard dilution series was prepared for the Human Genomic Control DNA (10 ng/μl), which was included in the qPCR kit, with a final concentration of 2 ng and 0.4 ng. The diluted standards were stored at -20°C for longer periods. The target DNA was freshly diluted with nuclease-free H<sub>2</sub>O to 1 ng/μl according to the concentration values determined by the fluorometric measurement with the Qubit™ quantification assay.

The DNA samples were run in duplicate and the three DNA standard dilutions and the negative template control (H<sub>2</sub>O) in triplicate, respectively. A PCR reaction mix was prepared according to Table 16. 20 µl of the reaction mix was added in wells of a MicroAmp™ Fast Optical 96-well reaction plate and 5 µl of each standard dilutions and DNA samples was added to the appropriate well for a total volume of 25 µl. The cycling conditions were as follows: Uracil-N glycosylase (UNG) incubation at 50°C for 2 min, enzyme activation at 95°C for 10 min and 40 cycles were performed at 95°C for 15 sec and 60°C for 1 min. The qPCR was run on a StepOne™ Real-Time PCR System and analyzed using the StepOne™ Software v2.3. The ROX™ Reference Dye was selected as passive reference dye and FAM™ dye and TAMRA as the TaqMan® probe reporter and quencher, respectively.

**Table 16: PCR reaction mix for quantification of amplifiable DNA by qPCR**

Reagents	Volume per reaction [µl]
2x Taqman™ Universal PCR Mix	12.5
20x RNase P Primer Probe	1.25
Nuclease-free H <sub>2</sub> O	6.25
Total volume	20

A standard curve plotting the threshold cycle (C<sub>T</sub>) versus concentration (ng/µl) of the DNA standards was generated by the StepOne software. The average concentration of the duplicates of each diluted tumor sample was determined and multiplied by the respective dilution factor to obtain the amount of amplifiable DNA of each sample. The DNA degradation index (DDI) was computed by dividing the quantification of amplifiable DNA calculated by the RNase P assay by the concentration of the total amount of DNA quantified with the Qubit dsDNA assay per tumor sample according to Formula 7.

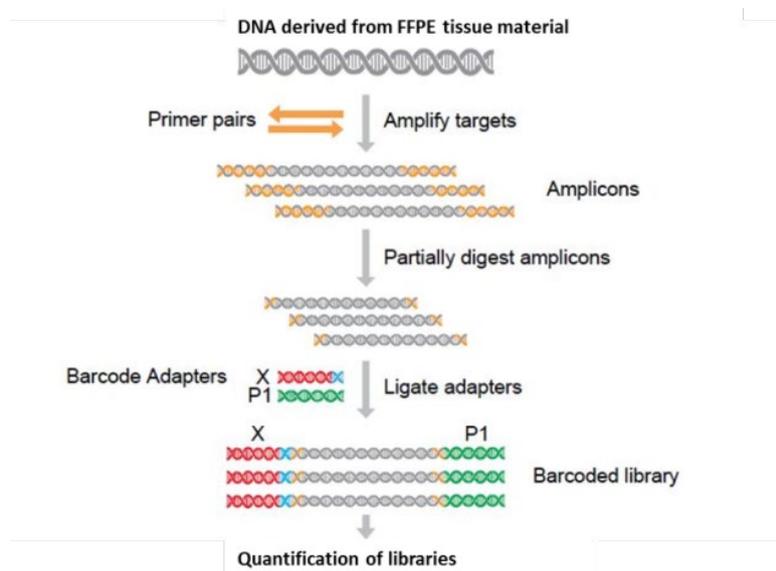
If the calculated DDI-value was > 0.3, 10 ng DNA was added as input to the PCR reaction mixture and if the DDI-value was between 0.2 and 0.3, 15 ng DNA was added. Tumors with DDI-values < 0.2 were not used for sequencing.

**Formula 7: Calculation of the DNA degradation index (DDI)**

$$DDI = \frac{\text{amplifiable DNA, RNase P [ng/}\mu\text{l]}}{\text{DNA concentration, Qubit [ng/}\mu\text{l]}}$$

### 2.3.5.2.3. Preparation of DNA libraries using a 4-pool GC-panel

After the quality check for amplifiable DNA, library preparation was performed using the Ion AmpliSeq™ targeted sequencing technology. The target regions were amplified using the Ion AmpliSeq™ Library Kit v2.0 and the designed GC-panel consisting of four primer pools. An overview about the workflow is given in Figure 15.



**Figure 15: Preparation of DNA libraries using the multiplex PCR based Ion AmpliSeq™ targeted sequencing technology**

Workflow modified according to the user's guide provided by the manufacturer (Thermo Fisher Scientific)

According to the concentration quantified by qPCR and to the calculated DDI-value, 10 ng or 15 ng DNA was added to the PCR reaction mixtures. For library preparation with a 4-pool primer panel, 2.5 ng or 3.75 ng DNA was diluted freshly with nuclease-free H<sub>2</sub>O and added to each primer pool. A separate PCR preparation mix was performed per primer pool according to Table 17 and all amplification reactions were performed in 0.2 ml PCR reaction tubes.

**Table 17: PCR reaction mix for DNA library preparation per primer pool**

Reagents	Volume per reaction and primer pool [μl]
Nuclease-free H <sub>2</sub> O	1-0
2.5 (or 3.75) ng DNA	0.5 – 1.5
2x Ion Ampli Seq Primer Pool	2.5
5x Ion Ampli Seq Hifi Master Mix	1
Total volume	5

PCR cycling was performed in a thermal cycler using following temperature steps: after an initial step at 99°C for 2 min to activate the enzyme, 21 cycles were performed consisting of denaturation at 99°C for 15 sec, annealing and extension at 60°C for 4 min and infinite holding stage at 10°C. After amplification, the four PCR reactions corresponding to one sample were combined in one reaction to a total volume of 20 μl by pipetting up and down and processed further on as one reaction mixture.

The combined amplicons were partially digested by adding 2  $\mu$ l of FuPa reagent to a total volume of 22  $\mu$ l. The samples were mixed thoroughly, centrifuged and loaded into a thermal cycler with following conditions: 50°C for 10 min, 55°C for 10 min, 60°C for 20 min and held for 24 h at 10°C.

Due to the fact that multiple libraries were sequenced on a single chip, barcoded adapters were ligated to the amplicons of each sample. A mixture of Ion P1 Adapter and Ion Xpress™ Barcode was prepared for each barcode at a final dilution of 1:4 and stored at -20°C. As a next step, the ligation reaction was performed adding the following components to the digested amplicons. First, 4  $\mu$ l of Switch Solution was added followed by 2  $\mu$ l of the diluted Ion Xpress™ barcode adapter mix and 2  $\mu$ l Ligase to a total volume of 30  $\mu$ l per sample. The cycle program was as follow: 22°C for 30 min, 68°C for 5 min, 72°C for 5 min and held for up to 24 h at 10°C. Finally, the purification of the barcoded libraries was performed using AMPure XP magnetic beads. Therefore, 45  $\mu$ l of the beads were mixed thoroughly for 2 min and pipetted to low binding tubes. The respective libraries were added and the bead-library mixture was incubated at room temperature for 5 min, then placed in a DynaMag™ magnet rack and incubated for at least 2 min, till the magnet beads had adsorbed to the inner surface of the tube forming a pellet. The tubes were opened individually and the supernatant was discarded without disturbing the pellet.

For a washing step, 150  $\mu$ l of freshly prepared 70% ethanol was added to each reaction tube, briefly centrifuged and placed in the magnet rack for 2 min till the beads formed pellets. While incubating for 2 min, the tubes were rotated completely to wash and shift the beads. The supernatant was removed and discarded without disturbing the pellet. The washing step with ethanol was repeated once. After the second washing step, the supernatant was removed in two pipette steps (2x100  $\mu$ l) to ensure that all ethanol droplets were removed. The reaction tubes were kept in the magnet rack and the caps were opened. The beads were air-dried at room temperature for a maximum of 5 min or alternatively dried in a heating block at 37°C for a maximum of 2 min.

To disperse the beads, 50  $\mu$ l Low TE buffer was added to the dried pellets, mixed thoroughly and briefly centrifuged. The reaction tubes were placed in the magnetic rack and incubated for 2 min till the beads had adsorbed completely to the wall of the tube. 40-45  $\mu$ l of the supernatant, containing the purified library, was transferred into a new PCR tube (0.2 ml) without touching the bead pellet and stored at -20°C for longer periods.

#### **2.3.5.2.4. Quantification of DNA libraries by qPCR using Ion Library TaqMan Quantitation Kit**

The libraries were quantified by qPCR using the Ion Library TaqMan™ Quantitation Kit running on a StepOne™ Real-Time qPCR system. The kit contained a recombinant Taq DNA polymerase and a Uracil DNA glycosylase (UDG) for prevention of re-amplification of carryover products between single qPCR reactions. The sample libraries were processed according to the protocol of the manufacturer (Thermo Fisher Scientific).

A standard dilution series was prepared freshly for the *E.coli* DH10B Control Library (68 pM stock concentration) at 0.68 pM and 0.068 pM. A 100-fold dilution was prepared for the libraries using nuclease-free H<sub>2</sub>O. The DNA standard dilutions, diluted libraries and the negative template control (H<sub>2</sub>O) were run in triplicates, respectively. A PCR reaction mixture was prepared according to Table 18 and aliquots of 5.5 µl were dispensed into the appropriate wells of a 96-well reaction plate. For a total reaction volume of 10 µl per well, 4.5 µl of the three standard dilutions, negative control and diluted libraries were added, respectively. The plate was sealed with a MicroAmp™ Clear Adhesive Film, mixed thoroughly and centrifuged briefly. The cycling conditions were as follows: UDG incubation at 50°C for 2 min, polymerase activation and template denaturation at 95°C for 20 sec and 40 cycles were performed at 95°C for 1 sec and 60°C for 20 sec and infinite holding stage at 10°C.

**Table 18: PCR reaction mix for quantification of DNA libraries by qPCR**

Reagents	Volume per reaction [µl]
2x Ion Library Taq Man qPCR Mix	5
20x Ion Library Taq Man Quantitation Assay	0.5

The ROX™ Reference Dye was selected as passive reference dye and FAM™ dye and NFQ-MGB as the TaqMan® probe reporter and quencher, respectively. The concentrations of the diluted sample libraries were calculated from the standard curve generated from serial dilutions of the control library by the StepOne™ Software v2.3. The average concentration of the triplicates of each diluted library was determined and by multiplying the determined concentration values with qPCR by the library dilution (1:100), the quantity of the undiluted library was calculated. Only those DNA samples were further processed for sequencing which achieved a library concentration > 100 pM as previously defined (Pfarr et al. 2016).

#### **2.3.5.2.5. Automated template preparation and sequencing**

Prior to sequencing, the DNA libraries were attached to a solid surface of a bead and amplified clonally (templated) by emulsion PCR. Therefore, the libraries were diluted to a final concentration of 25 pmol and the diluted libraries were combined in a barcoded Ion Chef Library sample tube. Automated template preparation and semiconductor chip loading with the clonally amplified beads was performed on an Ion Chef instrument according to the protocol of the manufacturer (Thermo Fisher Scientific). Following this, the chips were loaded successively onto an Ion S5XL instrument and sequencing was performed. The loading of the Ion Chef Instrument and the start of sequencing runs were conducted by a technical assistant and/or N. Pfarr in the routine-diagnostic laboratory of the Institute of Pathology. Criteria after sequencing were loading of total reads per chip between 80-95%, coverage of each amplicon at 2.000 and an average read value at least at 100.000.

#### **2.3.5.2.6. Data analysis of sequencing with GC-panel**

The raw sequencing data were processed using the Torrent Suite Software (v5.10.1). Coverage data and mutation analysis were generated by the two plugins *Torrent Variant Caller* and *Coverage Analysis* according to Pfarr et al. (2017b). The software tool ANNOVAR was used to annotate single nucleotide variants, insertions and deletions detected from human genome GRCh37/hg19 (Wang et al. 2010). The software performed a filter-based annotation to identify variants based on comparison to other variant databases, for example if a variant is already reported in the database of single nucleotide polymorphisms (dbSNP), in the 1000 Genome Project or Exome Aggregation Consortium (ExAC). The software also calculated SIFT, PolyPhen, MutationTaster and FATHMM scores to predict the probability that an amino acid substitution is damaging or affects protein function. The output files of ANNOVAR contained multiple columns including chromosomal region, start and end position, reference and observed nucleotides of a respective variant.

The sequencing reads were visualized by the Integrative Genomics Viewer (IGV) and tumorous and non-tumorous (reference) tissues were compared to each other (Robinson et al. 2017). Identified mutations were reviewed using the COSMIC database (Tate et al. 2019), Genome Browser Ensembl (Zerbino et al. 2018) and cBioPortal (Cerami et al. 2012, Gao et al. 2013). The tumorous samples were validated with eight non-tumorous tissues.

#### **2.3.5.2.7. Prediction of copy number variations**

Copy number variants (CNVs) were identified additionally by the Torrent Suite Software. Therefore, the summary of the coverage data of each tumor sample and amplicons were used for prediction of amplifications and deletions using a 4-step algorithm as previously reported (Endris et al. 2013, Pfarr et al. 2017b). Accordingly, a gene amplification was considered to be true, if the normalized amplicon read depth (NARD) of all amplicons differed by  $> 2$  standard deviation (SD) from the median value and deletions were considered to be true, if the SD of all amplicons covering one gene is  $< 0.5$  (Pfarr et al. 2017b).

### **2.3.6. Statistical analysis**

Comparison between the AI ratios determined by the microsatellite based multiplex PCR assays and the ratio of chromosomal alterations of the OncoScan method was performed using the Pearson correlation coefficient ( $r$ ). For statistical evaluation of the microsatellite based CIN classification in relation to tumor heterogeneity, a crossing probability was calculated from a data set comprising 45 tumor areas from nine patients. The probability that patients would be allocated to a different CIN classification due to intra-tumor variability of AI was computed by Formula 8 as previously shown (Kohlruss et al. 2018).

**Formula 8: Assessment of crossing probability for intra-tumor variability**

$$\begin{aligned}
\text{Crossing probability CP} &= P(\text{"patient changes CIN classification"}) \\
&= \int_{-\infty}^{\infty} P(\text{"patient with AI ratio } x \text{ changes CIN classification"}) dx \\
&= \int_{-\infty}^{\infty} P(\text{"patient changes CIN classification"} | \text{"AI ratio } x \text{"}) \cdot P(\text{"AI ratio } x \text{"}) dx \\
&= \int_{-\infty}^{\infty} \left( 1 - \Phi \left( \frac{|x - c|}{s_w / \sqrt{r}} \right) \right) \cdot \varphi \left( \frac{x - \bar{x}}{s_b} \right) dx.
\end{aligned}$$

Here,  $\varphi$  and  $\Phi$  denote the density and cumulative distribution function of the standard normal distribution. The parameters  $s_w$  and  $s_b$  refer to the empirical estimates of the standard deviation within repeated measurements (= intra-tumor variability) and between the patients' AI ratios (= inter-tumor variability). The cut-off value (AI ratio  $\geq$  or  $<$  0.2) of concern is denoted by  $c$ . The intra-tumor variability decreases by a multiplicative factor equal to the inverse square root of the number of  $r$  (repeated measurements) on a patient's AI analysis when an average AI ratio is used for risk prediction. Thus, the reliability of a prediction can be increased through the number of measurements made on the AI analysis of a patient.

The following tests were used for the statistical analysis regarding OS and clinical characteristics of the patients. Kaplan-Meier estimates of survival rates were compared by log rank tests. Relative risks were estimated by hazard ratios (HRs) from univariable Cox proportional hazard models or from Firth's corrected Cox-regression. Two-sided Chi-squared tests or Fisher's exact tests were used for hypothesis testing of differences between the relative frequencies.

The Wilcoxon Signed-Rank test was used for the comparison of the matched samples with paired biopsies before CTx and resected tumors after neoadjuvant CTx.

A multivariable Cox proportional hazard model was built by stepwise forward variable selection using likelihood-ratio tests of pre-therapeutically and post-therapeutically available clinical factors. The pre-therapeutically available factors were: gender, age (continuous variable), histological type according to Laurén (intestinal versus non-intestinal), tumor localization (proximal, middle, distal, total) and clinically determined tumor stage (cT2 versus cT3/cT4). The post-therapeutically available factors were: gender, age, histological type according to Laurén, tumor localization, depth of tumor invasion (pT2 versus pT3/pT4), lymph node involvement (pN0 versus pN+), metastasis status (M0 versus M+) and resection category (R0 versus R+).

Overall, exploratory 5% significance levels (two-tailed) were used for hypothesis testing. All statistical analyses were performed using IBM SPSS Statistics 25 and R version 3.1.0 and were supported by the statistician Dr. A. Hapfelmeier (Institute of Medical Informatics, Statistics and Epidemiology, Technical University of Munich). All survival graphs were generated using GraphPad Prism 8.

## **3. RESULTS**

### **3.1. Study enrolment and patient characteristics**

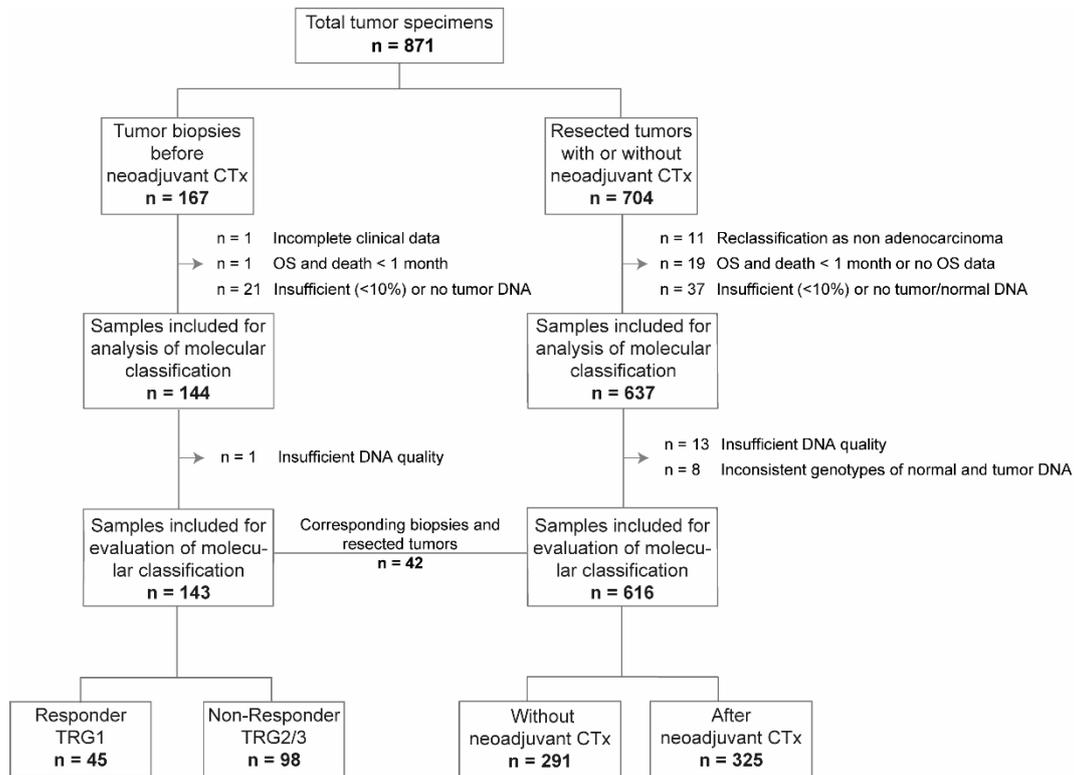
The patient cohorts were described recently in Kohlruss et al. (2019) and results are summarized below.

#### **3.1.1. Study enrolment**

The study population consisted of a total of 871 tumor specimens and was subdivided into two different tumor cohorts. The biopsy cohort consisted of patients with tumor biopsies before neoadjuvant CTx including responding patients with TRG1 and non-responding patients with TRG2 or TRG3. 167 tumor biopsies were initially included in the study and were analyzed for AI and CIN within the scope of a master thesis (Krenauer 2018). 24 tumor biopsies had to be excluded for the evaluation of the molecular classification due to incomplete clinical data ( $n = 2$ ), tumor cell contents  $< 10\%$  ( $n = 21$ ) or insufficient tumor DNA quality ( $n = 1$ ). The final dataset of the biopsy cohort comprised 143 tumor biopsies before CTx encompassing 45 patients with TRG1, 34 with TRG2 and 64 with TRG3.

The resected tumor cohort consisted of patients with tumors treated with surgery alone (non-CTx group) and with tumors after neoadjuvant CTx (CTx group). 704 patients were initially included in the study and 88 had to be excluded for analysis. Thereof, 11 tumors were reclassified as non-adenocarcinoma, 19 tumors showed an OS  $< 1$  month or no OS data were available, 37 tumors had insufficient tumor or normal tissue material, 13 tumors had insufficient DNA quality and eight patients revealed inconsistent genotypes of normal and tumor DNA. Of the analyzed 616 resected tumors, 291 samples were from patients treated with surgery alone and 325 from patients after treatment with neoadjuvant CTx. Corresponding tumor biopsies before and resected tumors after neoadjuvant CTx of 42 patients were included in this study.

An overview of the study enrolment with the respective exclusion criteria of the patients of the different cohorts is shown in Figure 16.



**Figure 16: Flow chart diagram of patient's inclusion for the analysis and evaluation of the molecular classification**

### 3.1.2. Patient characteristics

The clinical-pathological characteristics of the patients with 616 analyzed resected tumors and 143 analyzed tumor biopsies before CTx are summarized in Table 19.

The median survival of the patients with tumor biopsies was 48.1 months and of the patients treated with surgery alone 85.0 months compared to 32.2 months of the patients who received preoperative CTx. The median follow-up time for patients with tumor biopsies was 69.6 months and 57.9 months for patients with resected tumors.

Overall, the patients had more frequently tumors with proximal tumor localization (69.9% in the biopsy cohort and 48.7% in the resected tumor cohort). Differences between the patients treated with surgery alone and treated with neoadjuvant CTx were primarily found regarding the tumor localization. Of the 291 primary resected patients, 97 (33.3%) had tumors with proximal and 92 (31.6%) with distal tumor localization compared to 203 (62.5%) tumors with proximal and 39 (12%) with distal tumor localization of the 325 patients treated with neoadjuvant CTx. In the biopsy cohort, the distribution of intestinal (50.3%) and non-intestinal (49.7%) type tumors was balanced. In the resected cohort, 347 (56.3%) tumors were classified as intestinal and 269 (43.7%) as non-intestinal. As preoperative CTx treatment is recommended for patients with advanced tumor stages (cT3/cT4), the cohorts encompassed more tumors with cT3/cT4 especially in the CTx patient group (95.1%).

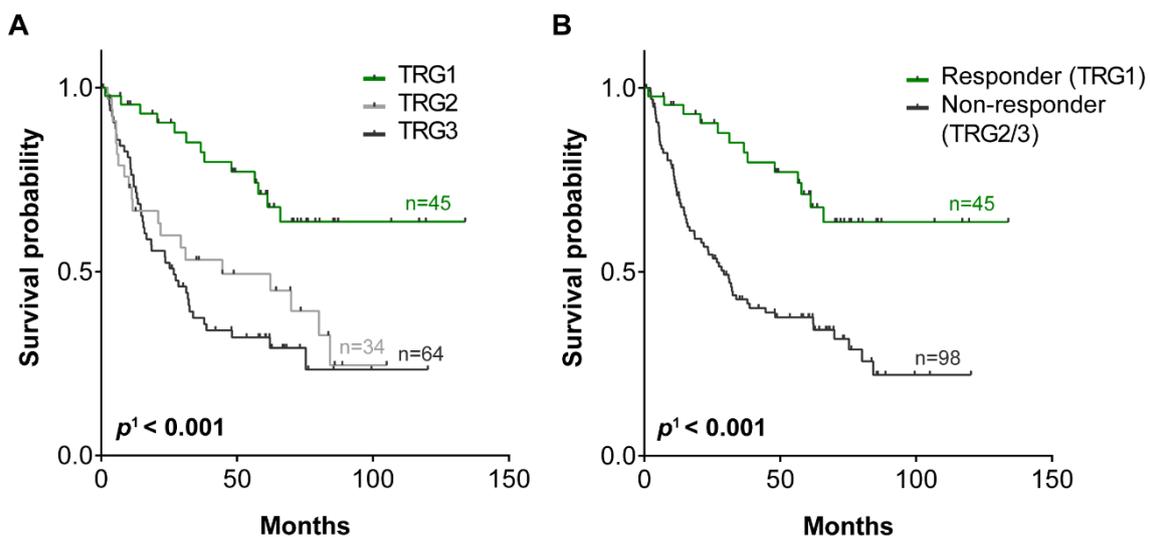
**Table 19: Clinical-pathological characteristics of the patient cohorts**

Category	Value	Tumor biopsies before neoadjuvant CTx		Resected specimens					
		n	%	All		Without neoadjuvant CTx		After neoadjuvant CTx	
		n	%	n	%	n	%	n	%
<b>Cases</b>	Total	<b>143</b>	100	<b>616</b>	100	<b>291</b>	100	<b>325</b>	100
<b>Age [yrs]</b>	Median	61.1		64.6		68.1		61.3	
	Range	23.1-78.0		28.3-90.9		32.1-90.9		28.3-81.2	
<b>Follow-up period [mo]</b>	Median	69.6		57.9		58.8		56.7	
	95% CI	61.6-77.6		53.2-62.6		50.7-67.0		47.5-65.9	
<b>Overall survival [mo]</b>	Median	48.1 <sup>1</sup>		44.6		85.0		32.2	
	95% CI	26.2-70.0		30.2-59.0		51.7-118.3		24.1-40.3	
<b>Gender</b>	Male	109	76.2	452	73.4	193	66.3	259	79.7
	Female	34	23.8	164	26.6	98	33.7	66	20.3
<b>Tumor localization</b>	Proximal	100	69.9	300	48.7	97	33.3	203	62.5
	Middle	23	16.1	153	24.8	84	28.9	69	21.2
	Distal	14	9.80	131	21.3	92	31.6	39	12.0
	Total/limitis	6	4.20	28	4.5	14	4.8	14	4.3
	n/a	-	-	4	<1	4	1.4	-	-
<b>Laurén classification</b>	Intestinal	72	50.3	347	56.3	155	53.3	192	59.1
	Non intestinal	71	49.7	269	43.7	136	46.7	133	40.9
<b>Tumor grade</b>	G1/2	33	23.1	125	20.3	80	27.5	45	13.8
	G3/4	110	76.9	399	64.8	210	72.2	189	58.2
	n/a	-	-	92	14.9	1	<1	91	28.0
<b>Clinical tumor stage</b>	cT2	8	5.6	144	23.4	129	44.3	15	4.6
	cT3/cT4	131	91.6	470	76.3	161	55.3	309	95.1
	n/a	4	2.8	2	<1	1	<1	1	<1
<b>(y)pT<sup>2</sup></b>	(y)pT0	9	6.3	-	-	-	-	-	-
	(y)pT1	12	8.4	56	9.1	42	14.4	14	4.3
	(y)pT2	20	14.0	79	12.8	47	16.2	32	9.8
	(y)pT3	81	56.6	328	53.3	139	47.8	189	58.2
	(y)pT4	19	13.3	153	24.8	63	21.6	90	27.7
	n/a	2	1.4	-	-	-	-	-	-
<b>(y)pN<sup>2</sup></b>	Negative	61	42.7	189	30.7	104	35.7	85	26.2
	Positive	80	55.9	427	69.3	187	64.3	240	73.8
	n/a	2	1.4	-	-	-	-	-	-
<b>Metastasis status</b>	No	97	67.8	533	86.5	272	93.5	261	80.3
	Yes	44	30.8	83	13.5	19	6.5	64	19.7
	n/a	2	1.4	-	-	-	-	-	-
<b>Resection category</b>	R0	117	81.8	469	76.1	235	80.8	234	72.0
	R1	24	16.8	147	23.9	56	19.2	91	28.0
	n/a	2	1.4	-	-	-	-	-	-
<b>Tumor regression grade</b>	TRG1	45	31.4	-	-	-	-	-	-
	TRG2	34	23.8	153 <sup>4</sup>	47.1 <sup>4</sup>	-	-	153	47.1
	TRG3	64 <sup>3</sup>	44.8	172 <sup>4</sup>	52.9 <sup>4</sup>	-	-	172	52.9
<b>Response to neoadjuvant CTx</b>	Responder (TRG1)	45	31.5	0	0	-	-	0	0
	Non-responder (TRG2/3)	98 <sup>3</sup>	68.5	325	100	-	-	325	100

<sup>1</sup>OS was defined as time between the date of operation and death by any cause. For two patients who were not operated, the date of start of CTx was used. <sup>2</sup>TNM classification of malignant tumors according to 7<sup>th</sup> Edition UICC. <sup>3</sup>Two patients with tumor progression during CTx were not operated and classified as TRG3 and as non-responders respectively. <sup>4</sup>Tumor regression grade corresponded only to patients with tumors treated with neoadjuvant CTx.

### 3.1.3. Response to neoadjuvant CTx

The association of response to neoadjuvant CTx and survival was analyzed in the 143 tumor biopsies before CTx. The tumor regression grade was significantly associated with survival ( $p < 0.001$ , Figure 17A). Within the patients with TRG2 and TRG3 only little differences in OS were observed and therefore both groups were classified as non-responders. Survival analysis of TRG1 (responder) in comparison to TRG2/3 showed a statistically significant difference ( $p < 0.001$ , Figure 17B). The median survival of the responding patients was not reached and the 5-years OS rate was 71% compared to a median survival of 29.3 months (CI, 21.2-37.4) and a 5-years OS rate of 38% for the non-responding patients (Supplementary Table 5).



**Figure 17: Response to neoadjuvant CTx and association with overall survival in the pre-therapeutic tumor biopsy cohort**

<sup>1</sup> $p$ -values (overall) of log rank test

### 3.2. Microsatellite analysis and CIN classification

AI and CIN were determined using the microsatellite based multiplex PCR assays established previously during my master thesis (Kohlruss 2016). Within the scope of this study, the multiplex assays were further developed to determine CIN by comparing it with the genome-wide OncoScan method and to define a reliable CIN classification according to TCGA. The multiplex PCR represents an adequate and cost efficient diagnostic tool for the analysis of large GC cohorts suitable in routine diagnostic settings. The performance of the microsatellite based multiplex PCR assays was initially tested on a subset of 100 tumors treated with surgery alone. In a preselection step, these tumors were analyzed for MSI whereby 10 of the 100 tumors showed MSI and only the 90 microsatellite stable tumors were further analyzed for AI and CIN. The results determined with the multiplex PCR assays were then compared to a genome-wide analysis using the OncoScan assay of chromosomal alterations in a subset of these tumors to define a classification of CIN.

Additionally, the multiplex PCR assays were analyzed in relation to tumor heterogeneity and the limit of detection was assessed for reliable CIN detection in the context of the tumor cell content (Figure 8, Material and Methods).

Results of the determination of the microsatellite based CIN classification were recently published in Scientific Reports (Kohlruss et al. 2018) and presented in the following paragraphs including Figures 18, 19, 20 and 22.

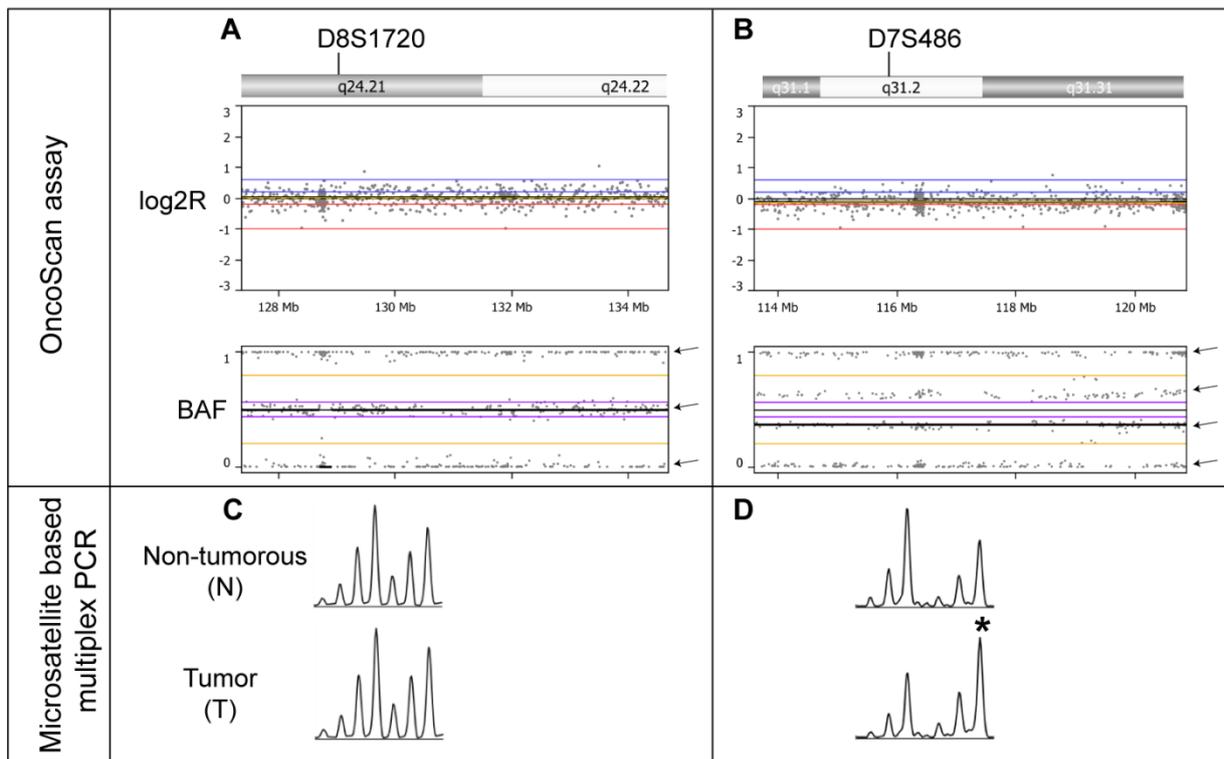
### **3.2.1. Comparison of multiplex PCR assays with genome-wide method**

#### **3.2.1.1. AI at single microsatellite loci**

The OncoScan assay enabled a genome-wide analysis of copy number gains or losses and additionally indicated regions of AI and loss of heterozygosity (LOH) including copy neutral LOH. A subset of 30 tumors, of the initially selected 100 tumors treated with surgery alone, was analyzed with the OncoScan assay and the occurrence of chromosomal alterations at a single microsatellite locus was determined.

In the 30 tumors, 190 measurements of AI of overall 527 informative measurements at the single microsatellite loci were detected using the multiplex PCR assays. The occurrence of copy number gains, losses and calls of AI determined by the OncoScan analysis were compared with the results for AI determined by the multiplex PCR assays and overall 442 (84%) concordant measurements were observed. The best concordance of both methods was found at microsatellite marker D12S1631 (12p12) and D7S486 (7q31) with 96% respectively and the worst at marker D17S250 (17q21) with 52%. More information about the concordance of both methods of every single microsatellite locus is given in Supplementary Table 6.

Examples of AI detected with the OncoScan assay compared to AI detected with the multiplex PCR assays are shown in Figure 18. The log<sub>2</sub>R and BAF graphs generated by the Nexus Express software are displayed for every chromosomal arm. One of the tumors shows zero values in the log<sub>2</sub>R graph and a normal three-band pattern in the BAF plot, which indicates no AI, copy number gain or loss at the region of the microsatellite marker D8S1720 (Figure 18A). In addition, the relation of the allele intensity of the shorter and longer allele of this respective microsatellite locus is not shifted (Figure 18C). In contrast, the other tumor was positive for AI in the OncoScan and multiplex PCR analysis. The BAF plots show a clear four-band pattern at the region of the microsatellite marker D7S486 (Figure 18B) and a clear shift is shown in the allele intensities in the tumor in comparison to the non-tumorous tissue (Figure 18D).



**Figure 18: Examples of tumors which are negative or positive for AI in the OncoScan and multiplex PCR assays**

In the log<sub>2</sub>R plot each dot represents copy number values which are calculated from signal intensities of a tumor sample compared to a normal reference and the B-allele frequency (BAF) generates allelic information at each SNP position. At the region of the microsatellite marker D8S1720 (**A**) the tumor has zero values in the log<sub>2</sub>R and a normal three-band pattern in the BAF plot (arrows), which indicates no AI. At the region of the microsatellite marker D7S486 (**B**) the BAF plots show a clear four band pattern (arrows), which indicates AI. The relation of the allele intensity of the first and second allele of the respective microsatellite locus is not shifted at the microsatellite marker D8S1720, which indicates no AI (**C**). At the marker D7S486 (**D**) a clear shift is shown in the allele intensities in the tumor (asterisk) in comparison to the non-tumorous tissue (N), which indicates AI (modified according to Kohlruss et al. 2018).

The 85 (16%) discrepant measurements of the two methods were considered more closely and categorized in different types. The most frequent type of discrepancy was the occurrence of balanced gains and copy neutral LOH (CN-LOH) detected only by the OncoScan method. Furthermore, 21 (25%) measurements of chromosomal alterations were clearly detected by the OncoScan method and 6 (7%) measurements of AI only by the multiplex PCR assays. Additionally, some of the discrepancies were related to AI-values very close to the cut-off borderlines of the respective microsatellite marker or to chromosomal positions of a respective marker nearby or between alteration events in the OncoScan analysis. The different types of discrepancies are summarized in Table 20.

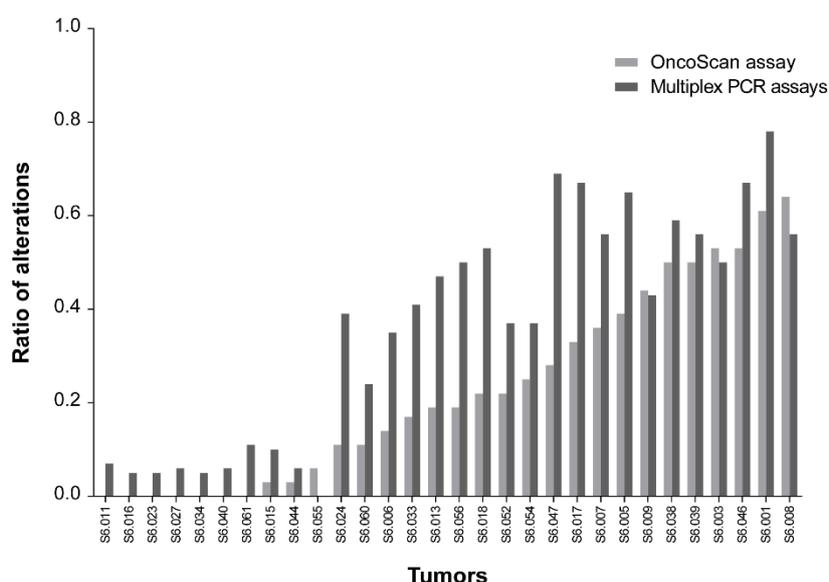
**Table 20: Types of discrepancy between the multiplex PCR assays and OncoScan analysis**

Type of discrepancy	Number of discrepant results	Frequency of discrepancy (%)	Detection by multiplex PCR assays	Detection by OncoScan method
Balanced gain / CN-LOH	32	38		x
Cut-off borderline for AI in OncoScan analysis	8	9		x
Clear detection with OncoScan	21	25		x
Clear detection with multiplex PCR assays	6	7	x	
Cut-off borderline for AI in multiplex PCR analysis	14	17	x	
Chromosomal position borderline of microsatellite marker in OncoScan analysis	4	5	x	
Total number of discrepant results	85 (16%)			
Total number of concordant results	442 (84%)			

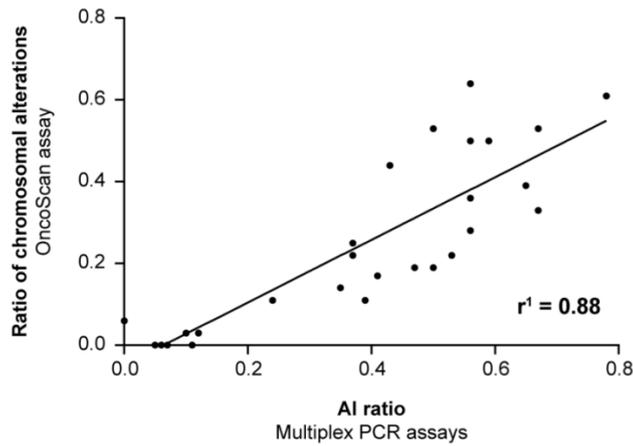
### 3.2.1.2. Genome wide alterations of chromosomal arms

Next, the genome wide extent of alterations affecting all chromosomal arms was determined for the 30 tumors analyzed with the OncoScan method and compared to the AI ratios determined by the multiplex PCR assays. A chromosomal arm was considered to be altered if at least 80% of one arm was lost, gained or demonstrated AI, referring to the definition of TCGA (2017). Figure 19 shows the ratios of chromosomal alterations detected by the OncoScan assay compared to the AI ratios determined by the multiplex PCR assays per tumor. AI ratios were measured in the range from 0 to 0.78 and ratios of chromosomal alterations detected by OncoScan in the range from 0 to 0.64. Detailed information about the ratios of alterations of the 30 analyzed tumors is shown in Supplementary Table 7.

A strong correlation was found for the ratios of alterations detected by both methods (Pearson correlation coefficient  $r = 0.88$ , Figure 20).



**Figure 19: Overview of the ratios of alterations in the 30 tumors detected with the OncoScan assay compared to the multiplex PCR assays**



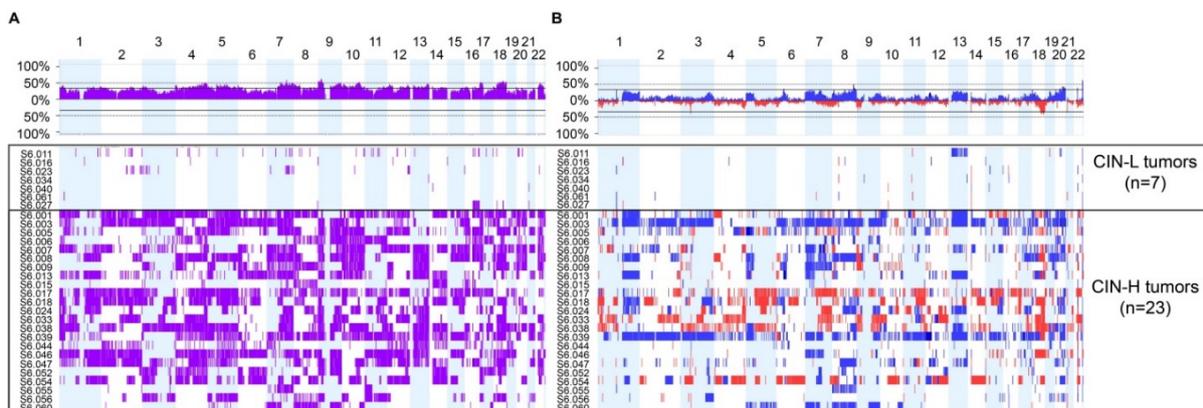
**Figure 20: Correlation of the ratios of chromosomal alterations detected with the OncoScan assay compared to the AI ratios determined with the multiplex PCR assays**

<sup>1</sup>Pearson correlation coefficient

### 3.2.2. Determination of CIN according to definition of TCGA based on OncoScan and microsatellite based multiplex PCR analysis

The next step was to establish a classification of CIN based on both methods and according to the definition of TCGA (2017). The 30 tumors were classified as having CIN according to their number of altered chromosomal arms determined by the OncoScan assay and resulted in 23 (77%) tumors with CIN-H and 7 (23%) with CIN-L as indicated in Table 21.

A genome-wide overview about the extent of copy number alterations and AI per tumor and chromosome is shown in the heat maps in Figure 21. The tumors were arranged according to their CIN-status and it is shown, that in the CIN-L tumors no chromosomal arm was altered for at least for 80%.



**Figure 21: Genome-wide OncoScan analysis of 30 tumors and OncoScan based CIN classification**

Summarized frequencies of AI per chromosome (A) and copy number alterations (B) are shown. The heat maps represent the occurrence of AI (purple), copy number gains (blue) and losses (red) per chromosome and tumor sample.

This OncoScan based CIN classification was set as a standard and a corresponding threshold value of the AI ratios determined with the multiplex PCR assays was determined. A high concordance of 90% for the CIN classification of both methods was observed for two possible cut-off values. Accordingly, seven tumors with AI ratios  $\leq 0.11$  were correctly classified as CIN-L and 20 tumors with AI ratios  $\geq 0.24$  were classified as CIN-H. As shown in Table 21, three tumors showed alterations at one or two chromosomal arms detected by the OncoScan assay and were therefore classified as CIN-H. By contrast, the AI ratios of these three tumors determined by the multiplex PCR assays were in the range of 0 to 0.10 and according to the cut-off value of an AI ratio  $\leq 0.11$  they would be classified as CIN-L. This shows that, a cut-off  $< 0.11$  resulted in the false positive classification of 10% of the cases as CIN-H, whereas a cut-off of  $\leq 0.24$  resulted in the false positive classification of 10% of the cases as CIN-L compared to the OncoScan analysis (Table 21).

**Table 21: Classification of CIN according to the definition of TCGA based on results of the OncoScan and microsatellite based multiplex PCR analysis**

Tumors	OncoScan analysis		Microsatellite based multiplex PCR analysis	
	Number of altered chromosomal arms <sup>1</sup>	CIN-status	AI ratio	CIN-status <sup>2</sup>
S6.008	23		0.56	
S6.001	22		0.78	
S6.046	19		0.67	
S6.003	19		0.50	
S6.038	18		0.59	
S6.039	18		0.56	
S6.009	16		0.43	
S6.005	14		0.65	
S6.007	13		0.56	
S6.017	12		0.67	
S6.047	10		0.69	
S6.054	9		0.37	
S6.052	8		0.37	
S6.018	8		0.53	
S6.013	7	CIN-H	0.47	CIN-H
S6.056	7	(n=23)	0.50	(n=20)
S6.033	6		0.41	
S6.006	5		0.35	
S6.024	4		0.39	
S6.060	4		0.24	
S6.055	2		<b>0.00</b>	
S6.044	1		<b>0.06</b>	<b>false-positive</b>
S6.015	1		<b>0.10</b>	<b>(n=3)</b>
S6.011	0		0.07	
S6.016	0		0.05	
S6.023	0		0.05	
S6.027	0		0.06	
S6.034	0	CIN-L	0.05	CIN-L
S6.040	0	(n=7)	0.06	(n=7)
S6.061	0		0.11	

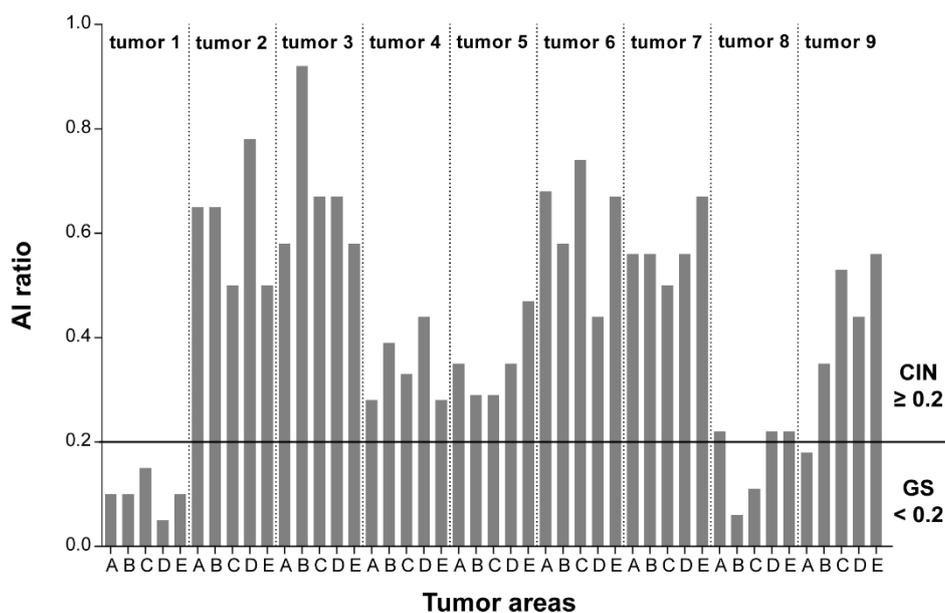
<sup>1</sup>Number of chromosomal arms which possessed at least 80% a copy number gain, loss or AI; <sup>2</sup>CIN-status according to OncoScan based classification

The multiplex PCR assays were applied on a cohort of 90 tumors and AI-ratio values in the range from 0 to 0.78 were observed. 13 tumors demonstrated AI ratios  $\leq 0.11$ , 62 tumors AI ratios  $\geq 0.24$  and 15 remaining tumors demonstrated AI ratios between 0.12 and 0.22. The median AI ratio of these tumors was 0.20 and a final cut-off value of  $\geq 0.2$  for the classification of tumors as CIN-H and  $< 0.2$  as CIN-L using the microsatellite based multiplex PCR assays was proposed. Essentially in line with the definition of the molecular subgroups of TCGA (2014), tumors which showed AI ratios  $\geq 0.2$  were classified as CIN and  $< 0.2$  as genomic stable (GS) in further analysis.

### 3.2.3. Microsatellite based CIN classification and tumor heterogeneity

To analyze the performance of the established microsatellite based CIN categorization in relation to tumor heterogeneity, we analyzed 45 tumor areas from nine patients for a variability of the CIN classification between the single areas. Although some variation in the occurrence of AI at the single markers in the various tumor areas was observed, a classification in CIN or GS respectively was concordant in 42 of the overall 45 (93%) tumor areas. Seven of the nine analyzed tumors demonstrated a concordant classification in all five areas, one tumor in four areas and one tumor in three areas (Figure 22). Detailed information about the occurrence of AI at a respective marker in the tumor areas is shown in Supplementary Figure 1.

The probability that tumors would be allocated to a different CIN classification due to intra-tumor variability of the AI ratios was assessed and for a single measurement ( $r=1$ ) the crossing probability was 10.3%. Assuming an increased number of analyzed tumor areas per patient, the resulting crossing probability showed that an average AI ratio reduces the crossing probability to 7.4% and 6.1% for the analysis of two and three tumor areas as described previously (Kohlruss et al. 2018).



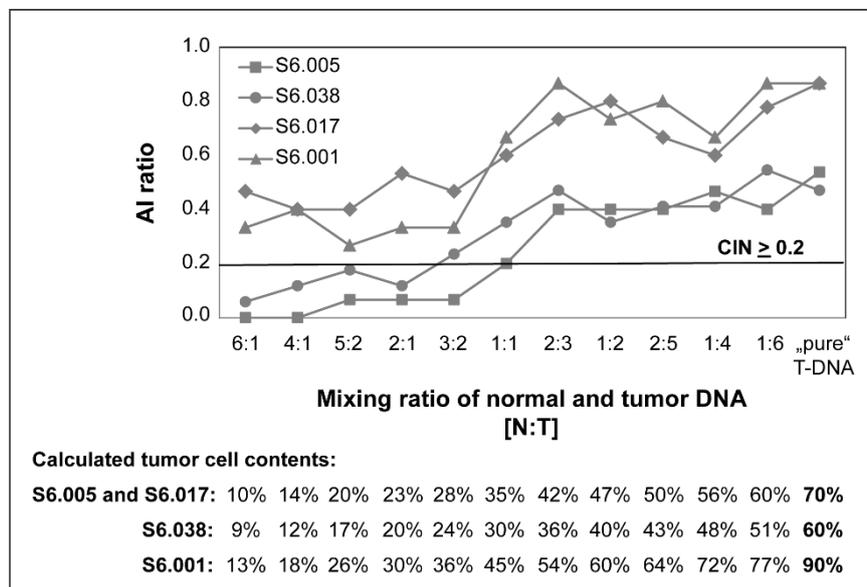
**Figure 22: Microsatellite based CIN classification and tumor heterogeneity**

AI ratios were calculated for nine tumors each with five tumor areas (A-E).

### 3.2.4. Limit of detection of microsatellite based multiplex PCR assays

Due to the fact that tumors are usually a mixture of normal and cancer cells, which could affect the detection of AI, the sensitivity of the established multiplex PCR assays were analyzed by dilution experiments. DNA from four exemplarily selected tumors with intermediate and high AI ratios was mixed with corresponding normal DNA in 11 dilution steps. The AI ratios of every chromosomal region per mixing ratio of normal and tumor DNA were measured.

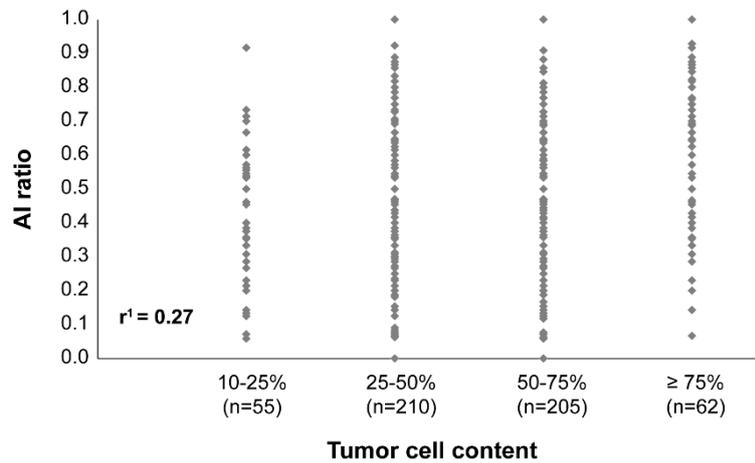
The initial tumor cell contents determined by a pathologist prior to the dilution experiments were 60% (S6.038), 70% (S6.005 and S6.017) and 90% (S6.001). Figure 23 shows that a stable classification in CIN was possible when the tumor cell content was between 24% and 35%. All determined AI ratios per mixing ratio of each tumor and the respective CIN classification are summarized in detail in Supplementary Table 8.



**Figure 23: Results of dilution experiments to determine the detection limit of the multiplex PCR assays**

AI ratios per mixing ratio of normal and tumorous DNA of each patient and the resulting tumor cell contents are shown. The initial tumor cell contents determined by a pathologist are shown in bold.

As a stable classification in CIN-H was given at tumor cell contents between 24%-35% the correlation of AI ratios and tumor cell contents was determined for all tumor specimens. Of the 532 resected specimens included for CIN analysis, 55 samples had tumor cell contents between 10% and 20%. As for these tumors AI ratios between 0.06 and 0.92 were measured and overall there was no linear correlation between the tumor cell content and the AI ratios ( $r = 0.27$ ), these tumor samples were included for CIN analysis (Figure 24). Therefore the limit of detection was set to a minimal tumor cell content of 10%. Detailed results of AI and CIN are summarized in the following paragraph 3.3.



**Figure 24: Correlation of tumor cell contents and AI ratios of all resected tumors analyzed for AI and CIN**  
<sup>1</sup>Pearson coefficient of correlation

### 3.3. Analysis of EBV, MSI and AI

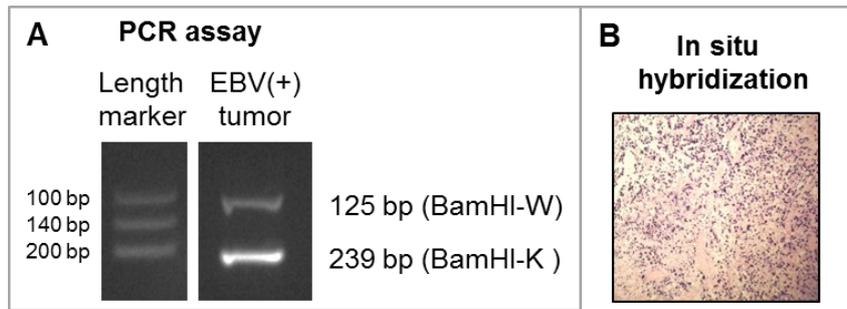
The results of EBV and MSI, especially with a focus on the MSI-L phenotype, have already been published in the Journal of Pathology, Clinical Research (Kohlruss, Grosser et al. 2019) and are summarized in brief in the following paragraphs. These results provided a basis for the determination of AI and CIN and finally for the definition of molecular subgroups in gastric carcinoma. Data collection of EBV and MSI of the resected tumor specimens were jointly collected by the working group and the analysis of EBV, MSI and AI of the tumor biopsies before CTx were mainly performed during a master thesis (Krenauer 2018).

#### 3.3.1. Results of EBV and MSI analysis

##### 3.3.1.1. Detection and frequency of EBV and MSI

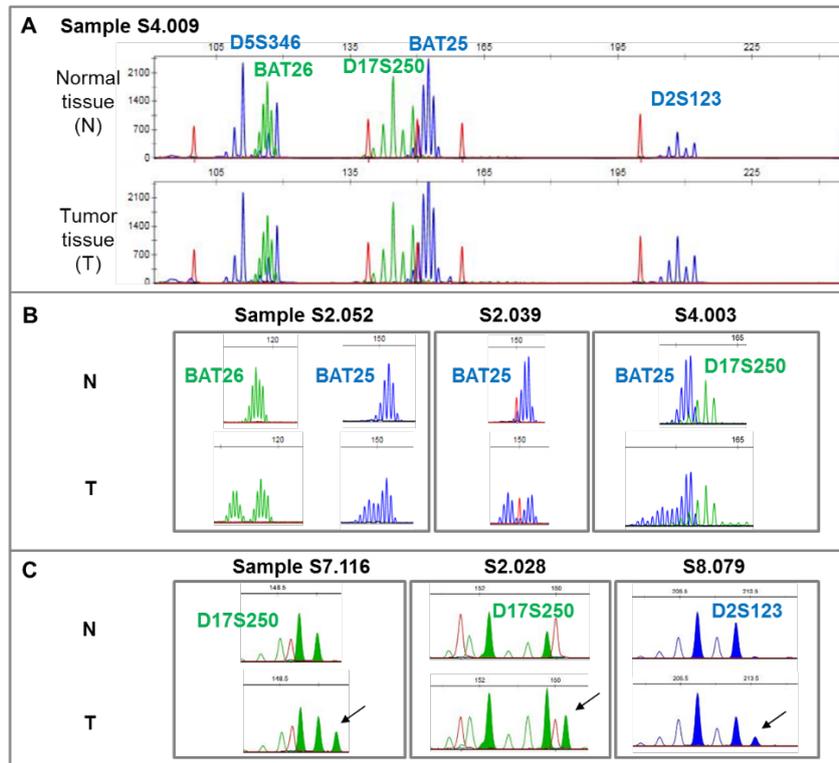
143 tumor biopsies before CTx and 616 resected tumors were analyzed for EBV and MSI and classified in the following four molecular subgroups: EBV(+), MSI-H, MSI-L and MSS/EBV(-).

Tumors with positive PCR signals (Figure 25A) and positive staining in the nuclei of the tumor cells after in situ hybridization were scored as EBV(+) (Figure 25B). Results of the fragment analysis representing different microsatellite patterns of MSS, MSI-H and MSI-L tumors are shown exemplarily in Figure 26. MSS tumors showed no additional peaks at the microsatellite markers compared to the corresponding normal tissue (Figure 26A). In contrast, tumors with MSI-H revealed additional alleles and distinct shifts of alleles at specific microsatellite markers (Figure 26B) and tumors with MSI-L revealed characteristically one additional peak at one of the respective markers (Figure 26C).



**Figure 25: Detection of an EBV-associated tumor using standardized methods**

PCR based assay using primers of EBV specific DNA in BamHI-W and BamHI-K regions (A) and in situ hybridization (B).

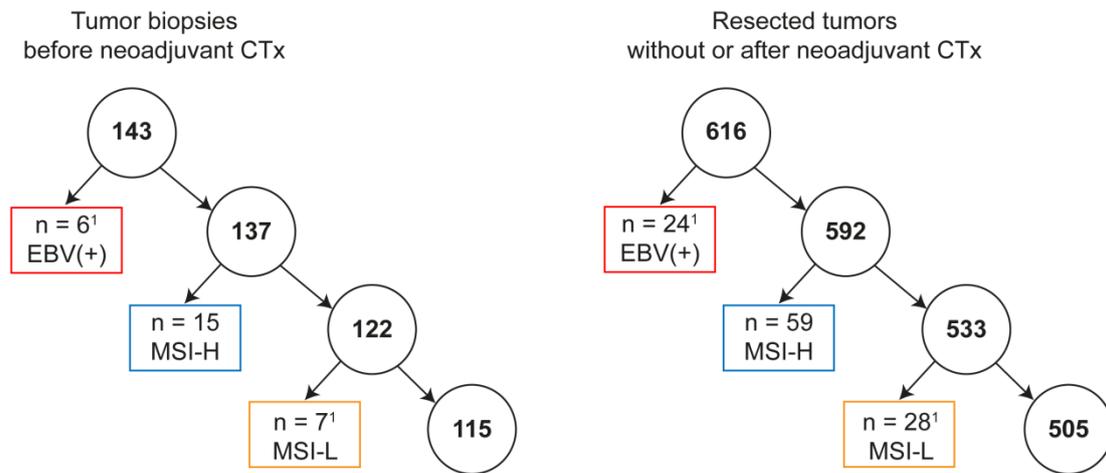


**Figure 26: Examples of microsatellite patterns using a multiplex PCR assay with two mononucleotide and three dinucleotide markers**

The five microsatellite markers were included in the Bethesda panel (Boland et al. 1998). A microsatellite stable tumor is shown which revealed no additional peaks at none of the five markers compared to the normal tissue (A). Three different MSI-H tumors with multiple additional alleles and distinct shifts at a specific marker are shown (B). Three different MSI-L tumors with one additional peak (arrows) at a respective marker are shown (C).

The classification algorithm of EBV and MSI including the total number of EBV(+), MSI-H and MSI-L tumors are shown in Figure 27. Results of EBV and MSI analysis revealed that 6 (4%) of the 143 tumors were classified as EBV(+), 15 (11%) as MSI-H, 7 (5%) as MSI-L and the remaining 115 (80%) tumors as MSS/EBV(-). In the resected tumor cohort, 24 (4%) of the total 616 tumors were classified as EBV(+), 59 (10%) as MSI-H, 28 (5%) as MSI-L and the remaining 505 (81%) tumors as MSS/EBV(-). All MSI-H tumors were negative for EBV.

One tumor biopsy and one resected tumor were positive for both, MSI-L and EBV. The EBV- and MSI-status of the 42 paired tumor biopsies before CTx and resected tumors after CTx was the same in all cases.

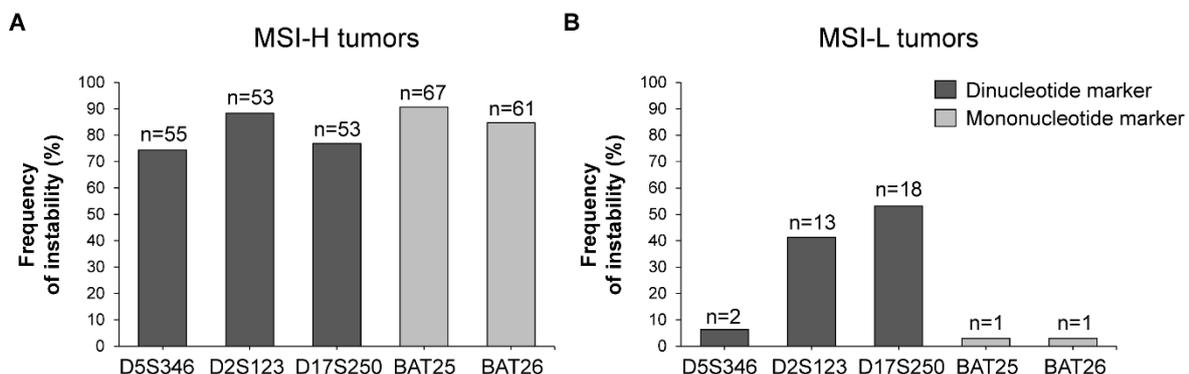


**Figure 27: EBV and MSI classification of patients with tumor biopsies before CTx and resected tumors without or after CTx**

<sup>1</sup>One tumor biopsy and one resected tumor without CTx were positive for MSI-L and EBV.

### 3.3.1.2. Types of instability in MSI-H and MSI-L tumors

Frequencies of unstable markers which were used for MSI testing and types of instability were characterized in MSI-H and MSI-L tumors. In the 74 tumors with MSI-H, the most frequent instability occurred with 91% and 85% at the mononucleotide repeats BAT25 and BAT26 and at the dinucleotide repeat D2S123 with 88% (Figure 28A). These repeats are included in the marker panel used for MSI analysis (Boland et al. 1998). In comparison to that, 33 (94%) of the 35 tumors with MSI-L showed instability at one of the three dinucleotide repeats, the most frequent one at D17S250 with 53% (Figure 28B). When considering the type of instability at the respective dinucleotide marker, it was found that 30 (91%) of the 33 tumors with MSI-L showed an insertion with two base pairs. For two MSI-L tumors, an instability was detected at one of the two mononucleotide repeats BAT25 and BAT26 and were found to be deletions (Figure 28B).



**Figure 28: Frequencies of instability at five microsatellite markers included for MSI analysis.**

Frequencies of instability are shown for patients with MSI-H (n=74, A) and MSI-L tumors (n=35, B).

### **3.3.2. Results of AI analysis**

#### **3.3.2.1. Patient characteristics for AI and CIN analysis**

In analogy to the classification system of the TCGA study, EBV(+) and MSI-H were treated as own molecular subgroups and only patients with MSS/EBV(-) and MSI-L tumors were included in the analysis of AI and CIN. The microsatellite based multiplex PCR assays were applied on the remaining 122 tumor biopsies before CTx and 533 resected tumors for the analysis of AI to determine the molecular CIN-subgroup (Figure 27).

One resected patient with a MSS/EBV(-) tumor, which was initially included for EBV and MSI analysis, was subsequently excluded from the analysis of AI and CIN due to insufficient DNA quality for data evaluation. The final resected tumor cohort consisted of 532 patients encompassing 248 patients treated with surgery alone and 284 treated with neoadjuvant CTx.

The patient characteristics for the analysis of AI and CIN of the analyzed tumor cohorts are summarized in Table 22.

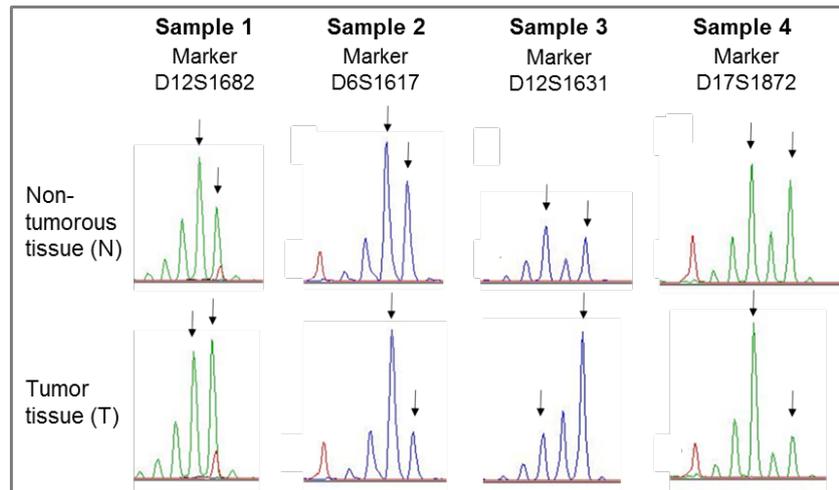
**Table 22: Patient characteristics for AI and CIN analysis**

Category	Value	Resected specimens							
		Tumor biopsies before neoadjuvant CTx		All		Without neoadjuvant CTx		After neoadjuvant CTx	
		n	%	n	%	n	%	n	%
<b>Cases</b>	Total	122	100	532	100	248	100	284	100
<b>Age [yrs]</b>	Median	60.6		64.3		66.9		61.2	
	Range	23.1 - 78.0		28.3 - 90.9		32.1 - 90.9		28.3 - 81.2	
<b>Follow-up period [mo]</b>	Median	70.8		57.9		56.4		60.7	
	95% CI	64.6 - 77.0		53.2 - 62.6		50.7 - 62.1		51.9 - 69.5	
<b>Overall survival [mo]</b>	Median	44.6 <sup>1</sup>		39.0		61.1		30.3	
	95% CI	18.5 - 70.8		29.3 - 48.8		27.5 - 94.7		25.2 - 35.4	
<b>Gender</b>	Male	92	75.4	391	73.5	165	66.5	226	79.6
	Female	30	24.6	141	26.5	83	33.5	58	20.4
<b>Tumor localization</b>	Proximal	88	72.1	268	50.4	81	32.7	187	65.8
	Middle	17	13.9	126	23.7	74	29.8	52	18.3
	Distal	12	9.8	107	20.1	76	30.6	31	10.9
	Total/limitis	5	4.1	27	5.1	13	5.2	14	4.9
<b>Laurén classification</b>	n/a	0	0	4	<1	4	1.6	0	0
	Intestinal	62	50.8	292	54.9	127	51.2	165	58.1
<b>Tumor grade</b>	Non-intestinal	60	49.2	240	45.1	121	49.8	119	41.9
	G1/2	30	24.6	113	21.1	71	28.6	42	14.8
<b>Clinical tumor stage (cT)</b>	G3/4	92	75.4	343	64.5	176	71.0	167	58.8
	n/a	0	0	76	14.3	1	<1	75	26.4
<b>(y)pT<sup>2</sup></b>	cT2	7	5.7	126	23.7	112	45.2	14	4.9
	cT3/4	112	91.8	404	75.9	135	54.4	269	94.8
	n/a	3	2.5	2	<1	1	<1	1	<1
<b>(y)pT<sup>2</sup></b>	(y)pT0	8	6.6	0	0	0	0	0	0
	(y)pT1	10	8.2	50	9.4	38	15.3	12	4.2
	(y)pT2	17	13.9	68	12.8	42	16.9	26	9.2
	(y)pT3	69	56.6	279	52.4	113	45.6	166	58.5
	(y)pT4	16	13.1	135	25.4	55	22.2	80	28.2
<b>(y)pN<sup>2</sup></b>	n/a	2	1.6	0	0	0	0	0	0
	Negative	53	43.4	158	29.7	87	35.1	71	25
	Positive	67	54.9	374	70.3	161	64.9	213	75
<b>Metastasis status</b>	n/a	2	1.6	0	0	0	0	0	0
	No	82	67.2	453	85.2	229	92.3	224	78.9
	Yes	38	31.1	79	14.8	19	7.7	60	21.1
<b>Resection category</b>	n/a	2	1.6	0	0	0	0	0	0
	R0	100	82	398	74.8	198	79.8	200	70.4
	R1	20	16.4	134	25.2	50	20.2	84	29.6
<b>Tumor regression grade</b>	n/a	2	1.6	0	0	0	0	0	0
	TRG1	38	31.2	-	-	-	-	-	-
	TRG2	33	27.0	141 <sup>3</sup>	49.6 <sup>3</sup>	-	-	141	49.6
<b>MSI and EBV status<sup>4</sup></b>	TRG3	51	41.8	143 <sup>3</sup>	50.4 <sup>3</sup>	-	-	143	50.4
	MSI-L	6	4.9	26	4.9	14	5.6	12	4.2
	MSS/EBV(-)	116	95.1	506	95.1	234	94.4	272	95.8

<sup>1</sup>OS was defined as time between the date of operation and death by any cause. For two patients who were not operated, the date of start of CTx was used. <sup>2</sup>Classification according to 7th Edition UICC; <sup>3</sup>Tumor regression grade refers only to patients with tumors after neoadjuvant CTx. <sup>4</sup>Patients with MSI-H and EBV(+) tumors were excluded for AI and CIN analysis.

### 3.3.2.2. Detection and frequency of AI

AI was analyzed for every single chromosomal region according to the individually determined specific cut-off values of each microsatellite marker. The five regions 8q24, 9p21, 12p12, 17q12 and 18q21 were covered with two microsatellite markers and AI was counted when at least one of the both markers detected AI. Results of fragment analysis of tumors with AI at a respective marker are shown exemplarily in Figure 29. The microsatellite patterns revealed shifts of the peak heights of the alleles or a reduction in the peak height of the longer allele.



**Figure 29: Examples of microsatellite patterns of tumors with AI at a respective microsatellite locus compared to the corresponding non-tumorous tissues**

The respective shorter and longer alleles are marked by arrows.

The frequency of AI at 17 chromosomal regions is shown for the resected tumor and the biopsy cohort in Figure 30. The bars represent the percentage of tumors with AI per number of informative markers. The tumor biopsies revealed somewhat higher frequencies of AI compared to the resected tumors. AI at 9p21 and 12p12 represented the most frequent alterations in both cohorts with 80% and 61% in the resected tumors and with 79% and 69% in the tumor biopsies (Table 23). AI at 18q21, 17p13, 2p21 and 8q24 was found in more than 50% of all resected tumors and 2p21, 18q21, 17p13, 8q24, 17q12, 5q21 and 6p25 in more than 50% of all tumor biopsies (Figure 30, Table 23).

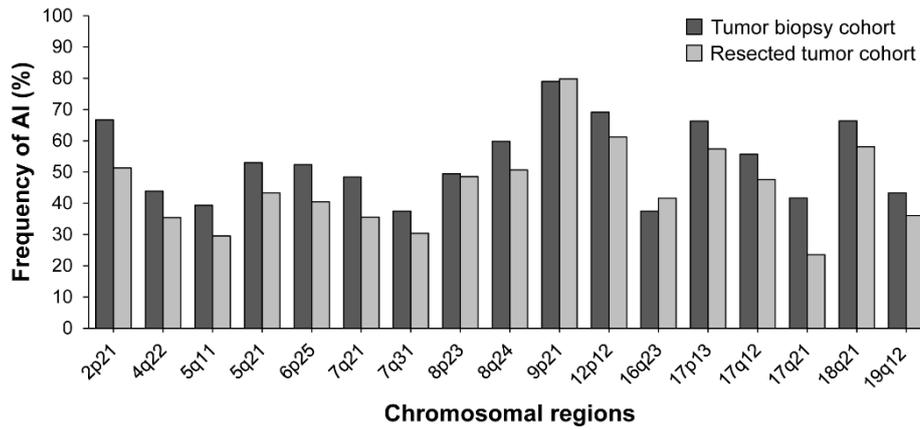


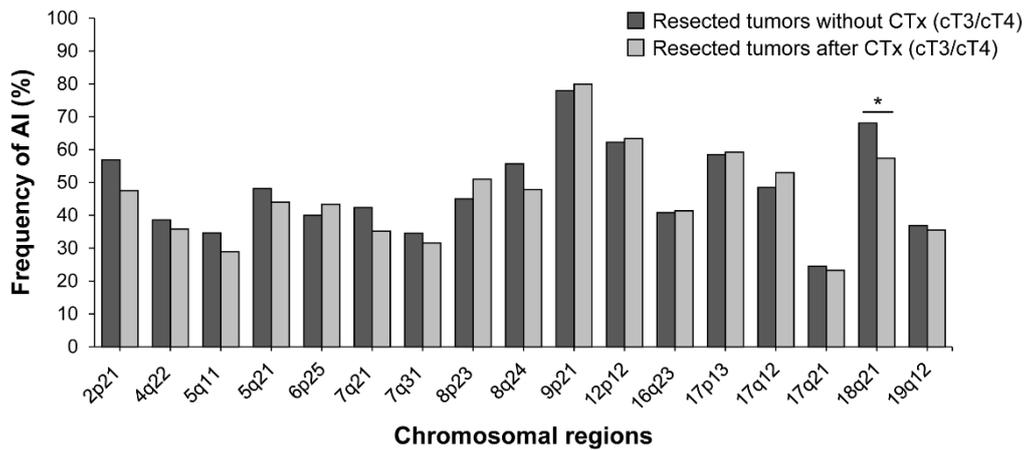
Figure 30: Frequency of AI in the tumor biopsy and resected tumor cohort

Table 23: Frequency of AI in the tumor biopsy and resected tumor cohort

Chromosomal regions	Tumor biopsy cohort (n=122)			Resected tumor cohort (n=532)		
	Number of tumors with AI	Number of informative tumors	Frequency of AI (%) <sup>1</sup>	Number of tumors with AI	Number of informative tumors	Frequency of AI (%) <sup>1</sup>
2p21	44	66	67	152	296	51
4q22	43	98	44	169	477	35
5q11	37	94	39	122	413	30
5q21	53	100	53	197	455	43
6p25	55	105	52	181	447	40
7q21	46	95	48	134	377	36
7q31	33	88	38	120	395	30
8p23	45	91	49	170	350	49
8q24	70	117	60	260	513	51
9p21	94	119	79	407	510	80
12p12	83	120	69	300	490	61
16q23	30	80	38	129	310	42
17p13	65	98	66	217	378	57
17q12	68	122	56	252	530	48
17q21	40	96	42	103	437	24
18q21	75	113	66	269	463	58
19q12	39	90	43	150	416	36

<sup>1</sup>Number of tumors with AI divided by number of informative tumors

Additionally, the frequencies of AI at the 17 chromosomal regions were considered in the resected tumors stratified according to the treatment of CTx and results are shown in Figure 31. For a better comparison of the two patient cohorts, only those tumors with an advanced clinical tumor stage (cT3/cT4) were included for this analysis. A statistically significant difference was observed in 18q21 ( $p = 0.053$ ). The most obvious, but not significant differences were shown in 2p21 ( $p = 0.221$ ) and 8q24 ( $p = 0.143$ ). Tumors without CTx showed higher frequencies at these chromosomal regions compared to tumors treated with CTx. Results of the comparison of the frequencies of AI in the resected non-CTx and CTx tumors with cT3/cT4 are summarized in Supplementary Table 9.



**Figure 31: Frequency of AI in the resected tumors without and after neoadjuvant CTx**

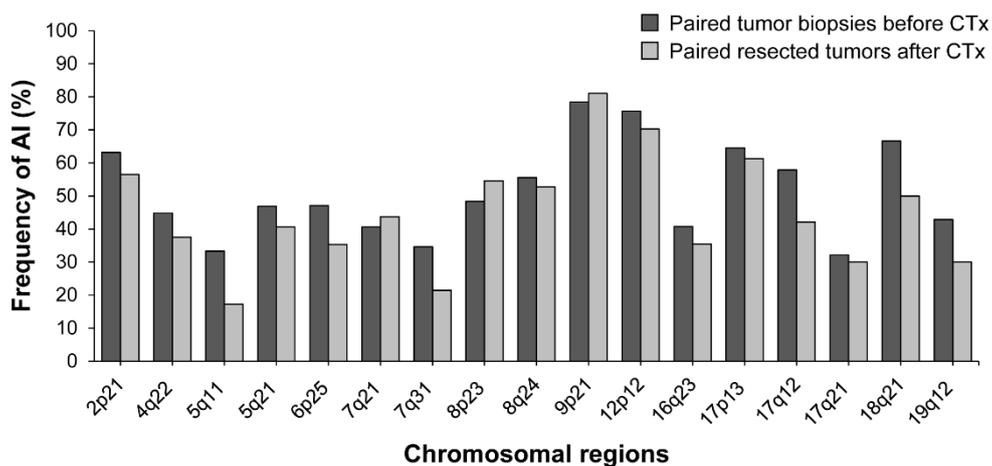
Only tumors with a clinical tumor stage cT3/cT4 were included in this analysis. \*Significant  $p$ -values of Chi-square test shown in Supplementary Table 9.

### 3.3.2.3. Frequency of AI in corresponding tumor biopsies and resected tumors

To further investigate the differences of the frequencies of AI at a respective chromosomal region between the tumor biopsy and resected tumor cohort, 38 corresponding tumor biopsies before and resected tumors after CTx were compared.

Figure 32 shows, that in the tumor biopsies generally higher frequencies of AI occurred at the most chromosomal regions except of 9p21, 8p23 and 7q21 compared to the corresponding resected tumors after CTx. The most obvious, but not statistically significant differences were found at 18q21 ( $p = 0.161$ ), 17q12 ( $p = 0.169$ ) and 5q11 ( $p = 0.156$ ). All frequencies of AI of the corresponding patients and the respective  $p$ -values are listed in Table 24.

The distribution of AI ratios of 38 corresponding tumor biopsies before CTx and resected tumors after CTx were compared by the Wilcoxon signed-rank test and revealed no significant difference for all paired tumors ( $p = 0.128$ ).



**Figure 32: Frequency of AI in paired tumor biopsies before and resected tumors after CTx**

A total number of 38 corresponding tumors were analyzed.

**Table 24: Comparison of the frequencies of AI in 38 paired tumor biopsies before and resected tumors after CTx**

Chromosomal regions	Corresponding tumor biopsies before CTx (TRG2/3)			Corresponding resected tumors after CTx (TRG2/3)			p-value <sup>1</sup>
	Number of tumors with AI	Number of informative tumors	Frequency of AI (%)	Number of tumors with AI	Number of informative tumors	Frequency of AI (%)	
2p21	12	19	63	13	23	57	0.663
4q22	13	29	45	12	32	38	0.561
5q11	10	30	33	5	29	17	0.156
5q21	15	32	47	13	32	41	0.614
6p25	16	34	47	12	34	35	0.324
7q21	13	32	41	14	32	44	1.0
7q31	9	26	35	6	28	21	0.280
8p23	15	31	48	18	33	55	0.622
8q24	20	36	56	19	36	53	0.813
9p21	29	37	78	30	37	81	0.772
12p12	28	37	76	26	37	70	0.601
16q23	11	27	41	11	31	35	0.684
17p13	20	31	65	19	31	61	1.0
17q12	22	38	58	16	38	42	0.169
17q21	9	28	32	9	30	30	0.887
18q21	22	33	67	18	36	50	0.161
19q12	12	28	43	9	30	30	0.254

<sup>1</sup>p-values of Chi-square test or Fisher's exact test

### 3.3.2.4. AI and association with patient characteristics

Association with clinical characteristics was analyzed for the 532 patients with resected tumors and results are summarized in Table 25.

AI at 17p13 was associated with male sex ( $p = 0.026$ ); AI at 4q22 ( $p = 0.013$ ), 5q11 ( $p = 0.014$ ), 17q12 ( $p = 0.010$ ) and 18q21 ( $p = 0.001$ ) with poorly differentiated tumors; AI at 7q31 ( $p = 0.045$ ), 17q12 ( $p = 0.002$ ) and 18q21 ( $p = 0.020$ ) with clinical tumor stage cT3/4 and AI at 12p12 with presence of lymph node metastasis ( $p = 0.021$ ).

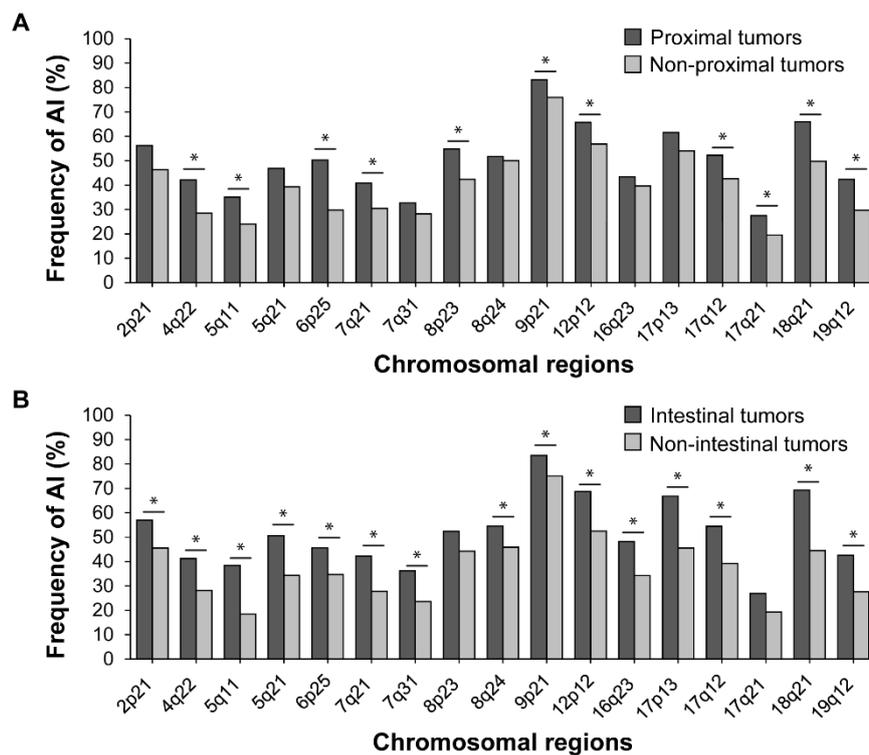
AI at 11 chromosomal regions (4q22, 5q11, 6p25, 7q21, 8p23, 9p21, 12p12, 17q12, 17q21, 18q21, 19q12) was significantly associated with proximal tumor localization. AI was significantly associated with the Laurén histological subtype at all chromosomal regions, except of 8p23 and 17q21, and AI was found more frequent in intestinal type tumors (Table 25).

Due to the significant differences observed in terms of AI, subgroup analysis for the 532 resected patients stratified according to tumor localization and Laurén histological subtypes were performed. The frequencies of AI of the respective patient group are shown in Figure 33 and results are summarized in Supplementary Table 10 and Supplementary Table 11. The resected tumors with proximal tumor localization or intestinal subtype showed generally higher frequencies of AI in comparison to the tumors with non-proximal tumor localization or non-intestinal subtype.

**Table 25: AI at a single chromosomal locus and association with patient characteristics of the resected tumor cohort**

	Sex		Localization		Laurén histological subtype		Tumor grade		cT		pT		pN		Metastasis stage		Resection status		Neoadjuvant CTx	
	Male	Female	Proximal	Non-proximal	Intestinal	Non-intestinal	G1/2	G3/4	cT2	cT3/4	pT1/2	pT3/4	Neg.	Pos.	Yes	No	R0	R1	No	Yes
2p21		0.807		0.089		<b>0.049</b>		0.434		0.485		0.126		0.751		0.960		0.900		0.136
4q22		0.780		<b>0.002</b>		<b>0.003</b>		<b>0.013</b>		0.261		0.401		0.298		0.638		0.609		0.751
5q11		0.338		<b>0.014</b>		<0.001		<b>0.014</b>		0.340		0.636		0.357		0.629		0.768		0.482
5q21		0.347		0.107		<0.001		0.139		0.156		0.552		0.531		0.616		0.505		0.943
6p25		0.867		<0.001		<b>0.018</b>		0.565		0.119		0.935		0.214		0.246		0.417		0.334
7q21		0.273		<b>0.036</b>		<b>0.004</b>		0.352		0.119		0.153		0.223		0.313		0.670		0.748
7q31		0.187		0.339		<b>0.007</b>		0.533		<b>0.045</b>		0.125		0.259		0.872		0.510		0.508
8p23		0.746		<b>0.020</b>		0.126		0.135		0.455		0.138		0.644		0.595		0.069		0.193
8q24		0.493		0.695		<b>0.049</b>		0.319		0.985		0.795		0.099		0.908		0.211		0.164
9p21		0.775		<b>0.044</b>		<b>0.017</b>		0.255		0.654		0.563		0.456		0.170		0.585		0.713
12p12		0.766		<b>0.045</b>		<0.001		0.343		0.114		0.270		<b>0.021</b>		0.191		0.657		0.149
16q23		0.572		0.509		<b>0.013</b>		0.205		0.908		0.581		0.183		0.442		0.881		0.784
17p13		<b>0.026</b>		0.140		<0.001		0.526		0.221		0.789		0.818		0.727		0.279		0.607
17q12		0.426		<b>0.027</b>		<0.001		<b>0.010</b>		<b>0.002</b>		0.110		0.119		0.552		0.511		0.069
17q21		0.883		<b>0.049</b>		0.061		0.353		0.980		0.808		0.567		0.948		0.380		0.525
18q21		0.895		<0.001		<0.001		<b>0.001</b>		<b>0.020</b>		0.982		0.459		0.915		0.219		0.626
19q12		0.140		<b>0.007</b>		<b>0.002</b>		0.111		0.983		0.175		0.155		0.686		0.822		0.871

In total, 532 patients with resected tumors without and after preoperative CTx were analyzed; *p*-values of Chi-square test or Fisher's exact test are shown and *p*-values < 0.05 are highlighted



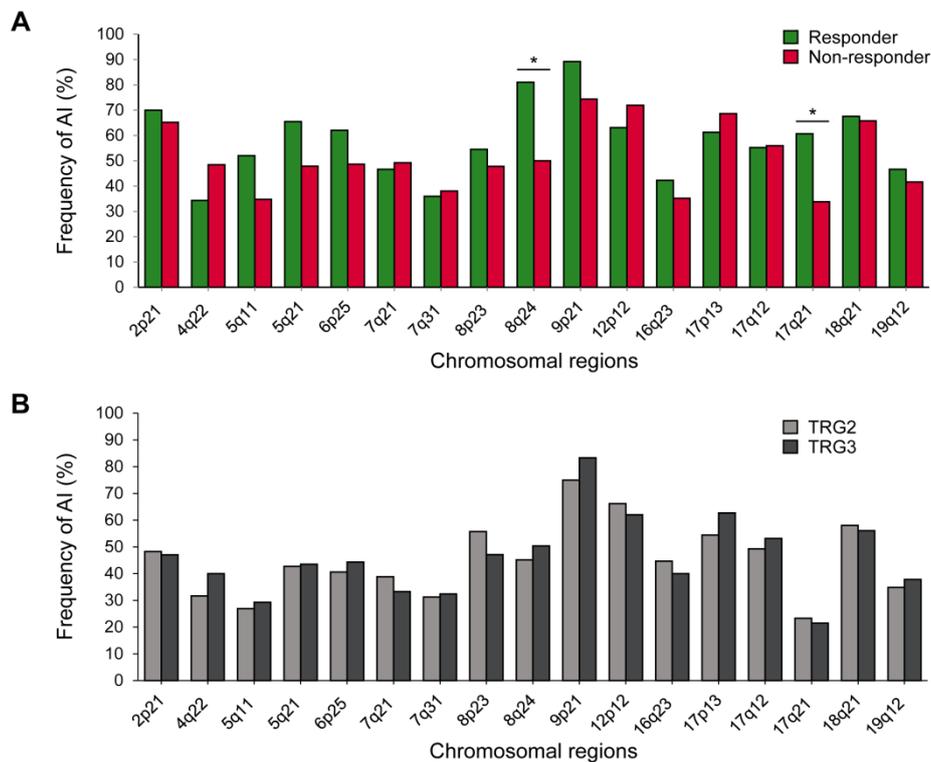
**Figure 33: Frequency of AI in the resected tumor cohort and association with tumor localization and Laurén histological subtypes**

Frequencies of AI are shown for patients with proximal and non-proximal tumors (A) and for patients with intestinal and non-intestinal tumors (B). \*Significant *p*-values of Chi-square test shown in Table 25.

### 3.3.2.5. AI and association with response to neoadjuvant CTx

Association of AI at a single microsatellite locus with response was analyzed for the 122 patients with tumor biopsies before CTx (Figure 34A). AI was significantly associated with response to neoadjuvant CTx at 8q24 ( $p = 0.001$ ) and 17q21 ( $p = 0.015$ ). An obvious but not statistically significant difference was observed between the responding and non-responding patients at 9p21 ( $p = 0.067$ ). In these regions, AI was more frequently found in tumors of responding patients.

In addition, the frequencies of AI at a single chromosomal region were observed for the 284 resected tumors after CTx stratified according to their tumor regression grade (Figure 34B). No significant association was found between AI at a single microsatellite locus and the tumor regression grade, merely AI at 9p21 showed a slight difference ( $p = 0.089$ ). AI at this chromosomal region was more frequently found in tumors with TRG3 (83%) compared to those with TRG2 (75%). All results including the total number of tumors with detected AI are summarized in Supplementary Table 12 and Supplementary Table 13.



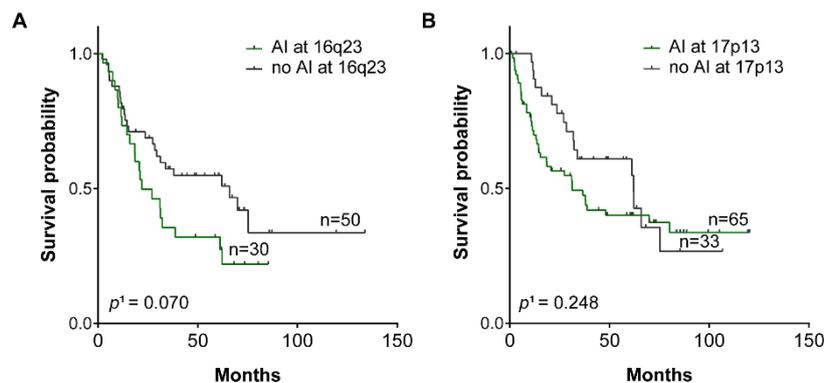
**Figure 34: Frequency of AI and association with response to neoadjuvant CTx**

Frequencies of AI are shown for responding and non-responding patients with tumor biopsies before CTx (A) and of resected tumors after CTx stratified according to their tumor regression grade (B). \*Significant  $p$ -values of Chi-square test shown in Supplementary Table 12 and 13.

### 3.3.2.6. AI and association with survival

AI at a respective chromosomal region was analyzed for survival in the tumor biopsies before CTx and in the resected tumors. In the following, survival curves with significant or obvious differences ( $p$ -values  $< 0.1$ ) are highlighted and shown in Figures 35-37. Additionally, survival analysis regarding the AI-status at chromosomal region 17p13, which covers the TP53 gene locus, is shown. All survival data in association with the AI-status including 1, 3 and 5 years survival rates are shown in detail in Supplementary Table 14 for the tumor biopsy cohort and Supplementary Table 15 for the resected tumor cohort.

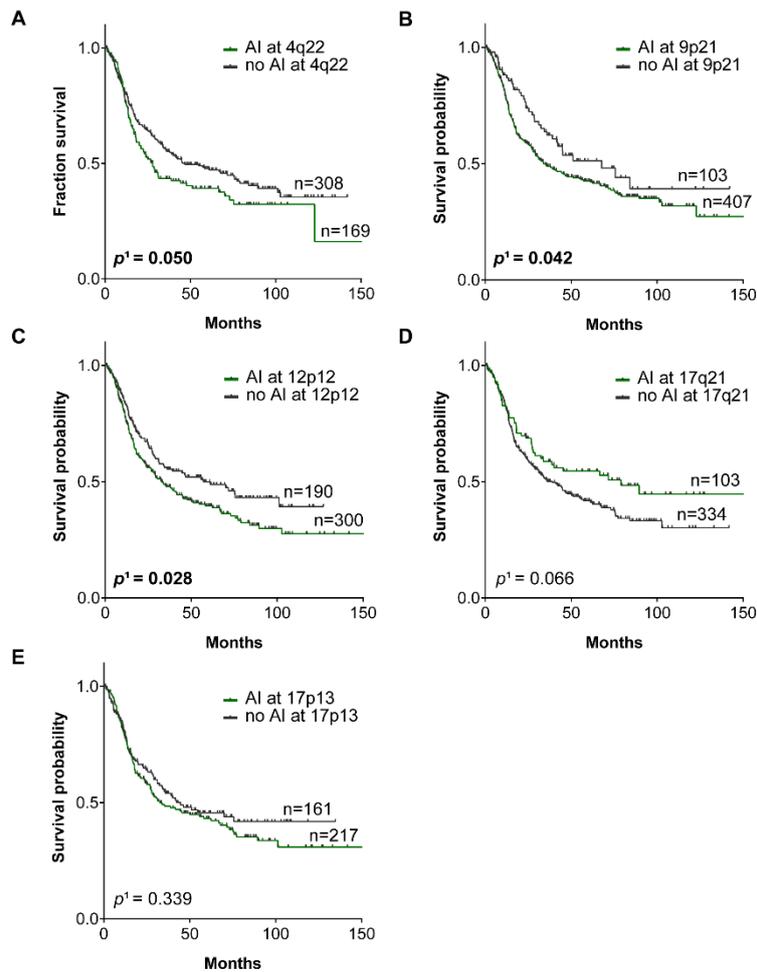
In the tumor biopsies before CTx no statistically significant associations of AI at a specific chromosomal region were found. An obvious difference was found at 16q23 in which patients with tumor biopsies having no AI had a better OS than tumor biopsies with AI at this locus ( $p = 0.070$ ) and AI at 17p13 revealed no significant difference ( $p = 0.248$ ; Figure 35).



**Figure 35: Survival and AI-status of the tumor biopsy cohort at respective chromosomal regions**

Kaplan-Meier curves are shown for patients with tumor biopsies before preoperative CTx having AI or no AI at 16q23 (A) and 17p13 (B). <sup>1</sup> $p$ -values of log rank test

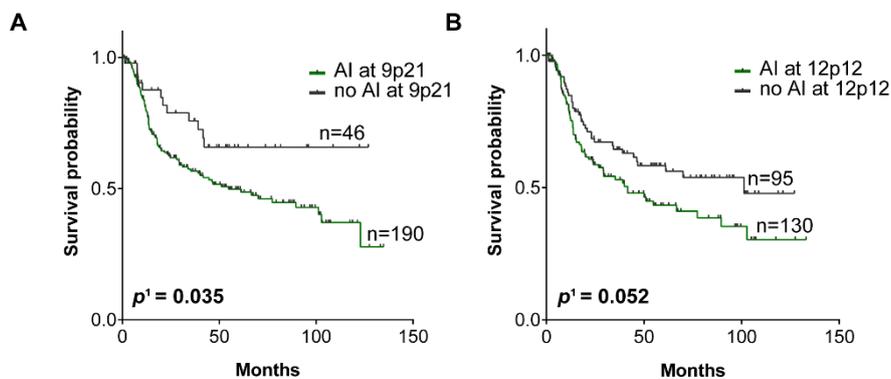
In the resected tumors, AI at chromosomal regions 4q22 ( $p = 0.050$ ), 9p21 ( $p = 0.042$ ) and 12p12 ( $p = 0.028$ ) showed a significant association with survival (Figure 36A-C). Patients with tumors having no AI at the respective chromosomal loci demonstrated a better survival with a median survival of 46.7 months compared to 27.9 months for 4q22, 67.7 months compared to 33.9 months for 9p21 and 57.1 months compared to 32.7 months for 12p12 (Supplementary Table 15). An obvious difference, but not statistically significant, was shown at chromosomal region 17q21 ( $p = 0.066$ , Figure 36D). Patients with tumors having AI at this locus showed a better survival than tumors without AI with a median survival of 79.0 months compared to 39.1 months (Supplementary Table 15). AI at 17p13 revealed no significant difference ( $p = 0.339$ ; Figure 36E).



**Figure 36: Survival and AI-status of the resected tumor cohort at respective chromosomal regions**

Kaplan-Meier curves are shown for patients with resected tumors without and after preoperative CTx having AI or no AI at 4q22 (A), 9p21 (B), 12p12 (C), 17q21 (D) and 17p13 (E). <sup>1</sup>*p*-values of log rank test, *p*-values < 0.05 in bold

Survival analysis in association with AI at a respective chromosomal region was separately performed in the non-CTx and CTx resected patient group and revealed that AI at the chromosomal regions 9p21 (*p* = 0.035) and 12p12 (*p* = 0.052) showed a significant association with survival only in the non-CTx group with a better OS in tumors having no AI at the respective chromosomal loci (Figure 37).



**Figure 37: Survival and AI-status of the non-CTx tumor cohort at respective chromosomal regions**

Kaplan-Meier curves are shown for patients with tumors treated with surgery alone having AI or no AI at 9p21 (A) and 12p12 (B). <sup>1</sup>*p*-values of log rank test, *p*-values < 0.05 in bold

Multivariable analysis was performed for the non-CTx patients (n = 248) with pre-therapeutically factors (sex, age, Laurén histological subtypes, tumor localization, clinical tumor stage cT) and AI-status at 4q22, 5q21, 9p21 and 12p12. Only chromosomal regions which showed significant or obvious differences ( $p$ -value < 0.1) were included in the multivariable analysis. Analyzing the AI-status and the pre-therapeutically available clinical factors revealed the clinical parameters age ( $p = 0.001$ ) and cT ( $p < 0.001$ ) as independent pre-therapeutic prognostic factors as shown in Table 26.

Including the post-therapeutically available factors (sex, age, Laurén histological subtypes, tumor localization, pT, nodal status, resection status, metastasis status) revealed age ( $p < 0.001$ ), pT ( $p = 0.004$ ) and pN ( $p < 0.001$ ) as independent post-therapeutic prognostic factors (Table 26).

**Table 26: Significant pre-and post-therapeutically available factors for survival in association with the AI-status in multivariable analysis in the non-CTx patient group**

	HR	95% CI	$p$ -value <sup>1</sup>
<b>Pre-therapeutic factors</b>			
Age	1.03	1.01-1.06	0.001
cT			
cT2			
cT3/4	2.61	1.75-3.89	<0.001
<b>Post-therapeutic factors</b>			
Age	1.03	1.01-1.05	<0.001
pT <sup>2</sup>			
pT1/2			
pT3/4	1.44	1.12-1.85	0.004
pN			
pN0			
pN1	3.32	1.95-5.67	<0.001

<sup>1</sup> $p$ -values of Cox's regression (forward likelihood ratio); <sup>2</sup>Classification according to 7<sup>th</sup> Edition UICC

### 3.4. Classification of CIN

First, the patients with tumor biopsies and resected tumors were categorized in genomic stable (GS) and CIN according to the classification system based on TCGA and the OncoScan data as shown previously in chapter 3.2.2. In a second step, a modified classification system of CIN in four subgroups according to the calculation of the quartiles of the AI ratios was proposed. Additionally, the performance of the classification of CIN in four subgroups was analyzed with a restricted number of chromosomal regions covering only common fragile sites or tumor suppressor genes to consider the respective AI marker related to functional aspects.

For a better characterization and understanding of the clinical relevance of the molecular CIN-subgroups, association with clinical characteristics, response to neoadjuvant CTx and survival analysis were performed for both CIN classification systems in the respective tumor cohorts.

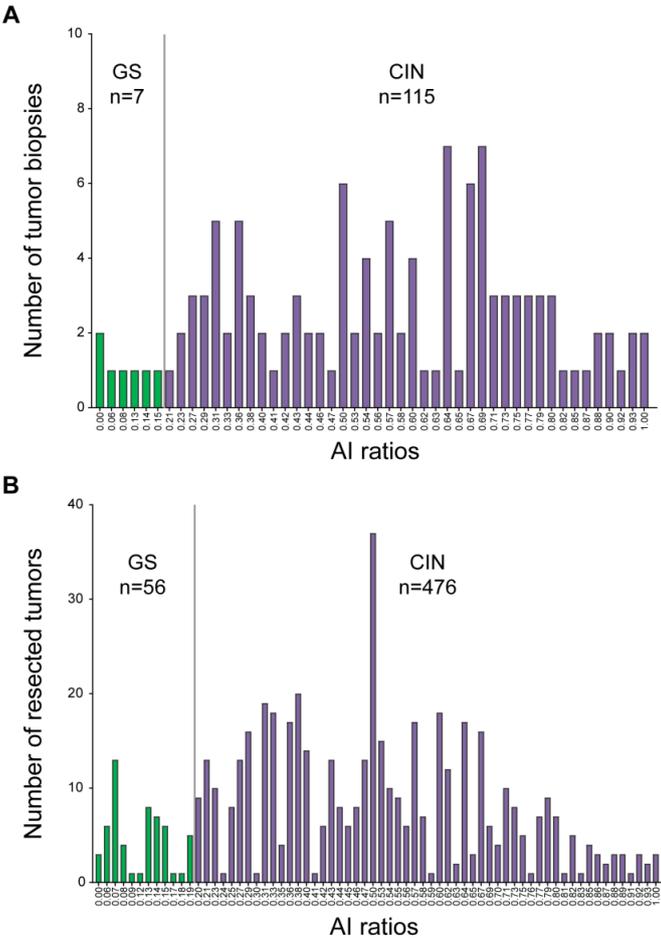
### 3.4.1. CIN classification according to TCGA

#### 3.4.1.1. Distribution and frequency of CIN

The distributions of the AI ratios in the tumor biopsies before CTx and the resected tumors are shown in Figure 38. Each bar represents the number of patients with a specific AI ratio measured in the range of 0 to 1. AI ratios in the range of 0 to 0.5 were measured in 41% of all tumor biopsies and AI ratios in the range of 0.5 to 1 in 59%. In the resected tumors, AI ratios in the range of 0 to 0.5 were measured in 58.5% and AI ratios from 0.5 to 1 in 41.5%. Additionally, the distribution of the AI ratios was also considered separately in the non-CTx and CTx patient groups and is shown in Supplementary Figure 2.

The tumors were classified as genomic stable (GS) when the AI ratios were < 0.2 and as CIN when the AI ratios were ≥ 0.2 according to the definition of TCGA as reported previously. In the tumor biopsy cohort, 7 (6%) of the 122 biopsies were classified as GS and 115 (94%) as CIN (Figure 38A). In the resected tumor cohort, 56 (11%) of the 532 tumors were GS and 476 (89%) were CIN (Figure 38B).

In the non-CTx group, 26 (10%) of 248 tumors were classified as GS and 222 (90%) as CIN. In the CTx-group a similar distribution of the AI ratios were observed and 30 (11%) of the 284 tumors were GS and 254 (89%) were CIN (Supplementary Figure 2).



**Figure 38: Distribution of AI ratios and CIN classification according to TCGA**

The total numbers of tumor biopsies before CTx (A) and of all resected tumors (B) are shown for each AI ratio. The GS subgroup is colored in green and the CIN subgroup in purple.

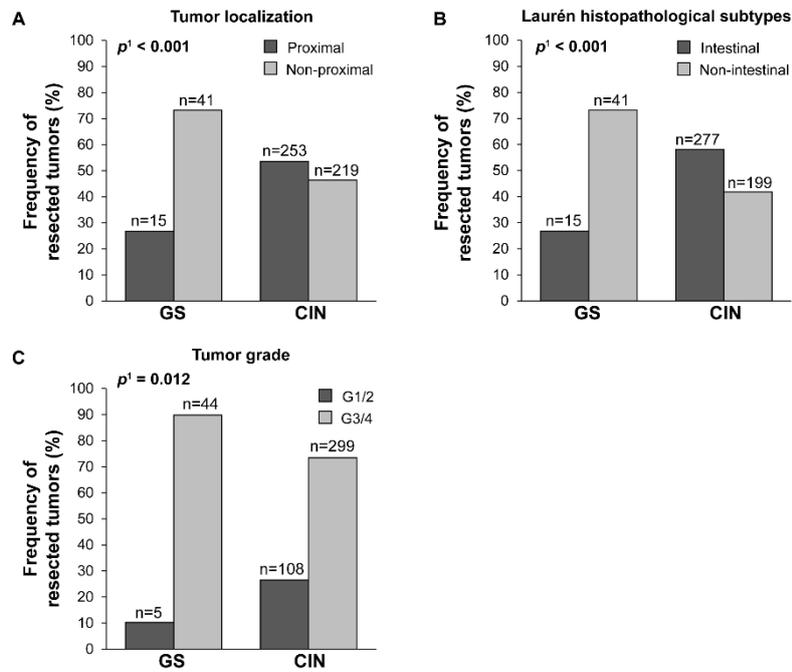
### 3.4.1.2. CIN according to TCGA and association with patient characteristics

Association with clinical characteristics was analyzed for the 532 patients with resected tumors without and after preoperative CTx and results are summarized in Table 27. CIN was found more frequently in proximal tumors ( $p < 0.001$ ), in intestinal type tumors ( $p < 0.001$ ) and was associated with poor differentiation ( $p = 0.012$ ). The significant differences are graphically shown in Figure 39.

**Table 27: CIN according to TCGA and association with patient characteristics**

Category	Value	All resected tumors ( $n=532$ )		$p$ -value <sup>1</sup>
		GS $n$	CIN $n$	
<b>Cases</b>	Total	56	476	
<b>Age [yrs]</b>	Median	64.9	64.2	
	Range	33.4 - 85.2	28.3 - 90.9	
<b>Age Median [yrs]</b>	< 64.3	27	238	0.800
	≥ 64.3	29	238	
<b>Gender</b>	Male	40	351	0.711
	Female	16	125	
<b>Tumor localization</b>	Proximal	15	253	<0.001
	Non-proximal	41	219	
	n/a	0	4	
<b>Laurén classification</b>	Intestinal	15	277	<0.001
	Non-intestinal	41	199	
<b>Tumor grade</b>	G1/2	5	108	0.012
	G3/4	44	299	
	n/a	7	69	
<b>Clinical tumor stage</b>	cT2	15	111	0.576
	cT3/4	41	163	
	n/a	2	0	
<b>(y)pT<sup>2</sup></b>	(y)pT1/2	14	104	0.591
	(y)pT3/4	42	372	
<b>(y)pN<sup>2</sup></b>	Negative	18	140	0.672
	Positive	38	336	
<b>Metastasis status</b>	No	45	408	0.286
	Yes	11	68	
<b>Resection category</b>	R0	47	351	0.097
	R1	9	125	
<b>Neoadjuvant CTx</b>	No	26	222	0.976
	Yes	30	254	

<sup>1</sup> $p$ -values of Chi-square test;  $p$ -values < 0.05 in bold; <sup>2</sup>Classification according to 7<sup>th</sup> Edition UICC

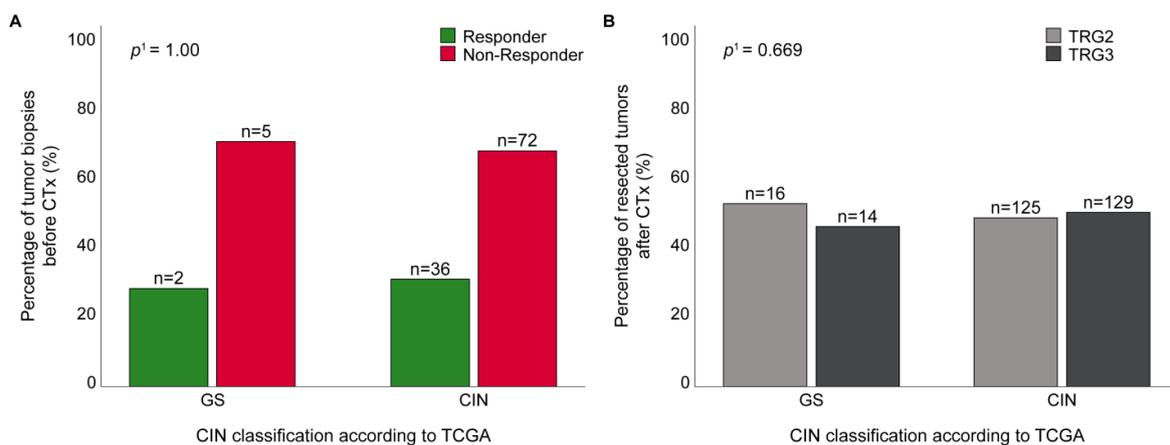


**Figure 39: CIN according to TCGA and significant associations with clinical parameters**

Frequencies of resected tumors without and after CTx categorized into GS or CIN are shown in relation to tumor localization (A), Laurén histological subtype (B) and tumor grade (C). <sup>1</sup>p-values of Chi-square test

### 3.4.1.3. CIN according to TCGA and response to neoadjuvant CTx

The association of CIN with response to neoadjuvant CTx was analyzed in the tumor biopsies before CTx. The two molecular subgroups, GS and CIN, were not associated with response ( $p = 1.00$ , Figure 40A). Additionally, the association of CIN with the tumor regression grade was analyzed for the resected tumors after neoadjuvant CTx and revealed also no significant difference ( $p = 0.669$ , Figure 40B).



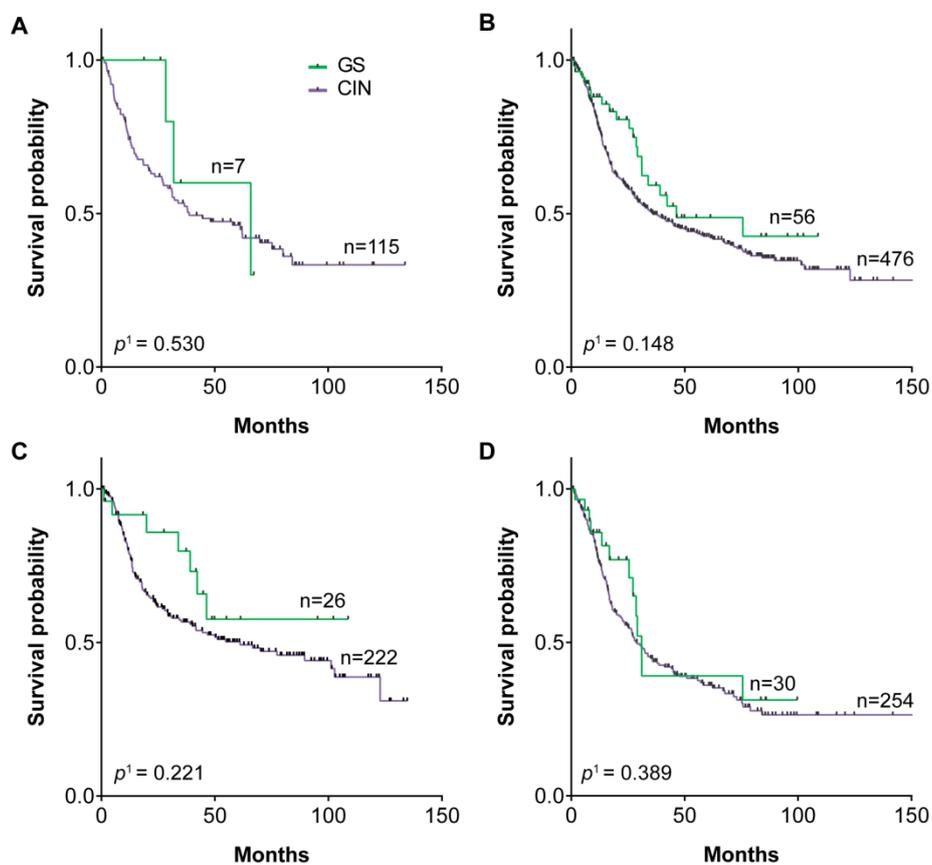
**Figure 40: CIN according to TCGA and association with response to neoadjuvant CTx**

The response of patients with tumor biopsies before CTx (A) and tumor regression of patients with resected tumors after CTx (B) are shown in relation to the CIN classification according to TCGA. <sup>1</sup>p-values of Chi-square test

### 3.4.1.4. CIN classification according to TCGA and association with survival

Survival analysis regarding the two subgroups of CIN according to TCGA was performed in the tumor biopsy and resected tumor cohort. Additionally, survival analysis of the resected tumors was separately performed stratified according to CTx treatment (yes/no). The estimated survival curves of the patient cohorts are shown in Figure 41 and all survival data are summarized in Supplementary Table 16.

The comparison of OS of patients with pre-therapeutic tumor biopsies regarding the two molecular subgroups of CIN showed no statistically significant difference ( $p = 0.530$ ; Figure 41A). Also in the resected tumor cohort no statistically significant differences were observed ( $p = 0.148$ ; Figure 41B), not in the non-CTx group ( $p = 0.221$ ; Figure 41C) either in the CTx group ( $p = 0.389$ ; Figure 41D). Overall, patients with GS tumors showed a slightly better survival than patients with CIN. In the tumor biopsy cohort, the median survival of patients with GS tumors was 65.9 months compared to 38.7 months in tumors with CIN and in the resected cohort 46.3 months compared to 35.9 months.



**Figure 41: CIN classification according to TCGA and survival of the patients**

Kaplan-Meier curves are shown for patients with tumor biopsies before neoadjuvant CTx (A) and for patients with resected tumors: all patients (B), patients treated without CTx (C), patients treated with neoadjuvant CTx (D). <sup>1</sup>p-values of log rank test

Subgroup analysis within all resected specimens stratified according to tumor localization and Laurén histopathological subtypes was performed. The most obvious differences were emphasized in the following. Survival curves and data are shown in detail in Supplementary Figure 3 and Supplementary Table 16.

An evident but not statistically significant difference was observed in the non-proximal group ( $p = 0.091$ ; Supplementary Figure 3B). Patients with GS tumors had a better survival than patients with CIN tumors (HR, 0.59; 95% CI, 0.32-1.10). In the proximal group no statistically significant difference was found but patients with GS tumors had a worse survival compared to patients with CIN tumors (HR, 1.25; 95% CI, 0.64-2.46,  $p = 0.517$ ). Subgroup analysis within all resected specimens stratified according to Laurén histological subtypes revealed a statistically significant difference in the non-intestinal group ( $p = 0.013$ ; Supplementary Figure 3D). Patients with GS tumors had a significant better survival than patients with CIN tumors (HR, 0.45; 95% CI, 0.24-0.86).

### 3.4.2. Modified CIN classification according to quartiles of AI ratios

#### 3.4.2.1. Distribution and frequency of modified CIN-subgroups

In the next step, a modified CIN classification system independently of the TCGA and OncoScan data was proposed. Therefore, the tumors were classified in four subgroups of CIN according to the calculation of quartiles of the AI ratios. The resulting cut-off values for the classification of CIN in one of the four subgroups and the total number of tumor biopsies and resected tumors per CIN subgroup are shown in Table 28.

The subgroups were termed as follows: CIN-low (CIN-L), low CIN-medium (low CIN-M), high CIN-medium (high CIN-M) and CIN-high (CIN-H). The CIN-L subgroup (AI ratio  $\leq 0.25$ ) corresponded most likely to the genomic stable (GS) subgroup according to TCGA as shown previously.

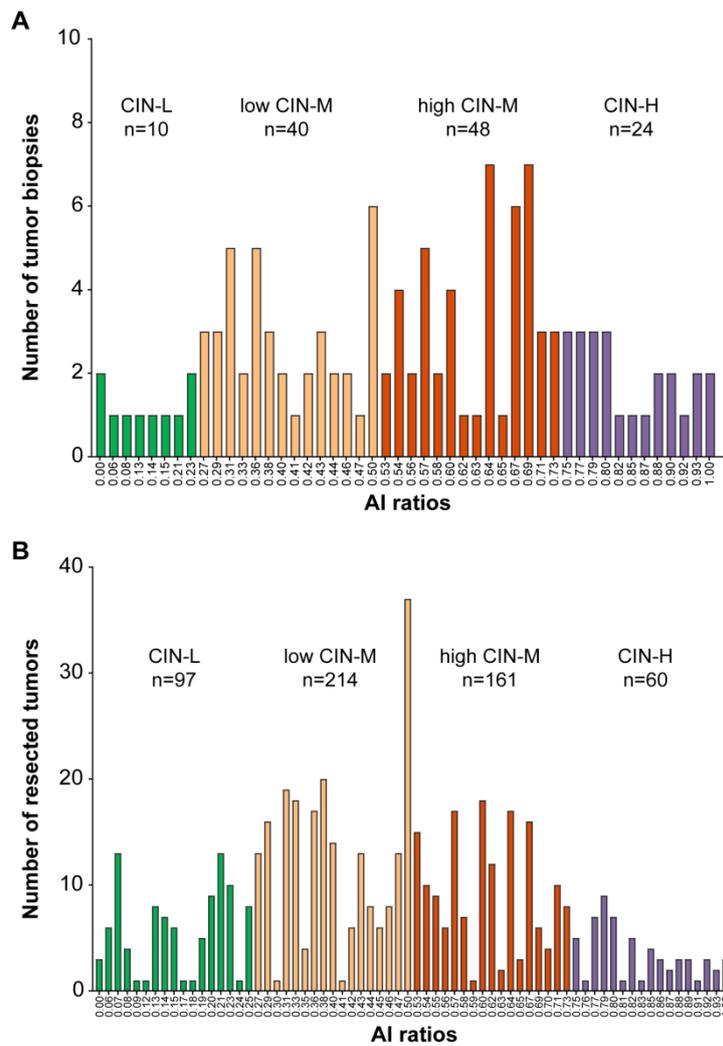
The distribution of AI ratios and frequencies of CIN-subgroups in the tumor biopsy and resected tumor cohort are shown in Figure 42.

**Table 28: Cut-off values for CIN classification in four subgroups according to the quartiles of AI ratios**

	<b>CIN-L<sup>1</sup></b>	<b>low CIN-M</b>	<b>high CIN-M</b>	<b>CIN-H</b>
Quartiles	25%	25-50%	50-75%	75%
Cut-off values of AI ratios	$\leq 0.25$	$> 0.25$ and $\leq 0.50$	$> 0.50$ and $< 0.75$	$\geq 0.75$
Number of tumor biopsies (%)	10 (8.2)	40 (32.8)	48 (39.4)	24 (19.6)
Number of resected tumors (%)	97 (18.2)	214 (40.3)	161 (30.3)	60 (11.2)

The total number of tumor biopsies before CTx and all resected tumors is shown for each CIN-subgroup.

<sup>1</sup>CIN-L corresponded most likely to the genomic stable (GS) subgroup according to the definition of TCGA (2014)



**Figure 42: Distribution of AI ratios and modified CIN classification according to the quartiles of AI ratios**  
 The total numbers of tumor biopsies before CTx (**A**) and of all resected tumors (**B**) are shown for each AI ratio. The CIN-L subgroup is colored in green, low CIN-M in light orange, high CIN-M in dark orange and CIN-H in purple.

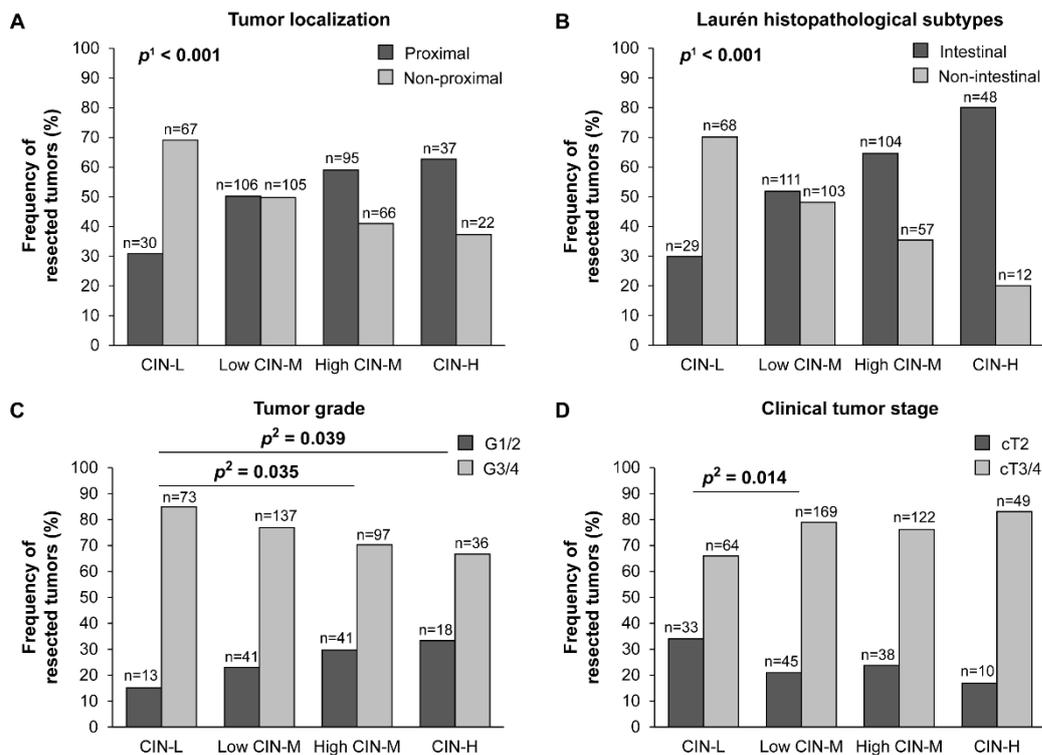
### 3.4.2.2. Modified CIN classification and association with patient characteristics

Association with clinical characteristics was analyzed for 532 patients with resected tumors and results are shown in Table 29. CIN-L was more frequent found in non-proximal ( $p < 0.001$ , Figure 43A) and non-intestinal tumors ( $p < 0.001$ , Figure 43B). Additionally, tumors with CIN-L were associated with poor differentiation ( $p = 0.036$ , Figure 43C) and with lower clinical tumor stage cT2 ( $p = 0.045$ , Figure 43D).

**Table 29: Modified CIN classification and association with patient characteristics**

Category	Value	Resected tumors ( $n=532$ )				$p$ -value <sup>1</sup>
		CIN-L $n$	Low CIN-M $n$	High CIN-M $n$	CIN-H $n$	
<b>Cases</b>	Total	97	214	161	60	
<b>Age [yrs]</b>	Median	64.4	63.4	64.4	66.8	
	Range	31.7-85.2	32.4-85.1	28.3-90.9	32.1-88.3	
<b>Age Median [yrs]</b>	< 64.3	48	113	79	25	0.494
	≥ 64.3	49	101	82	35	
<b>Gender</b>	Male	68	152	124	47	0.393
	Female	29	62	37	13	
<b>Tumor localization</b>	Proximal	30	106	95	37	<0.001
	Non-proximal	67	105	66	22	
	n/a	0	0	4	0	
<b>Laurén classification</b>	Intestinal	29	111	104	48	<0.001
	Non-intestinal	68	103	57	12	
<b>Tumor grade</b>	G1/2	13	41	41	18	0.036
	G3/4	73	137	97	36	
	n/a	11	36	23	6	
<b>Clinical tumor stage</b>	cT2	33	45	38	10	0.045
	cT3/4	64	169	122	49	
	n/a	0	0	1	1	
<b>(y)pT<sup>2</sup></b>	(y)pT1/2	25	46	33	14	0.778
	(y)pT3/4	72	168	128	46	
<b>(y)pN<sup>2</sup></b>	Negative	33	71	38	16	0.155
	Positive	64	143	123	44	
<b>Metastasis status</b>	No	84	183	137	49	0.859
	Yes	13	31	24	11	
<b>Resection category</b>	R0	80	148	123	47	0.064
	R1	17	66	38	13	
<b>Neoadjuvant CTx</b>	No	46	93	80	29	0.667
	Yes	51	121	81	31	

<sup>1</sup> $p$ -values (overall) of Chi-square test or Fisher's exact test;  $p$ -values < 0.05 in bold; <sup>2</sup>Classification according to 7<sup>th</sup> Edition UICC



**Figure 43: Modified CIN classification and significant associations with clinical parameters**

The frequency of resected tumors without and after CTx categorized into the respective modified CIN-subgroup is shown in relation to tumor localization (A), Laurén histological subtype (B), tumor grade (C) and clinical tumor stage (D). <sup>1</sup>Overall *p*-values of Chi-square test; <sup>2</sup>*p*-values of Chi-square test compared to CIN-L (ref.)

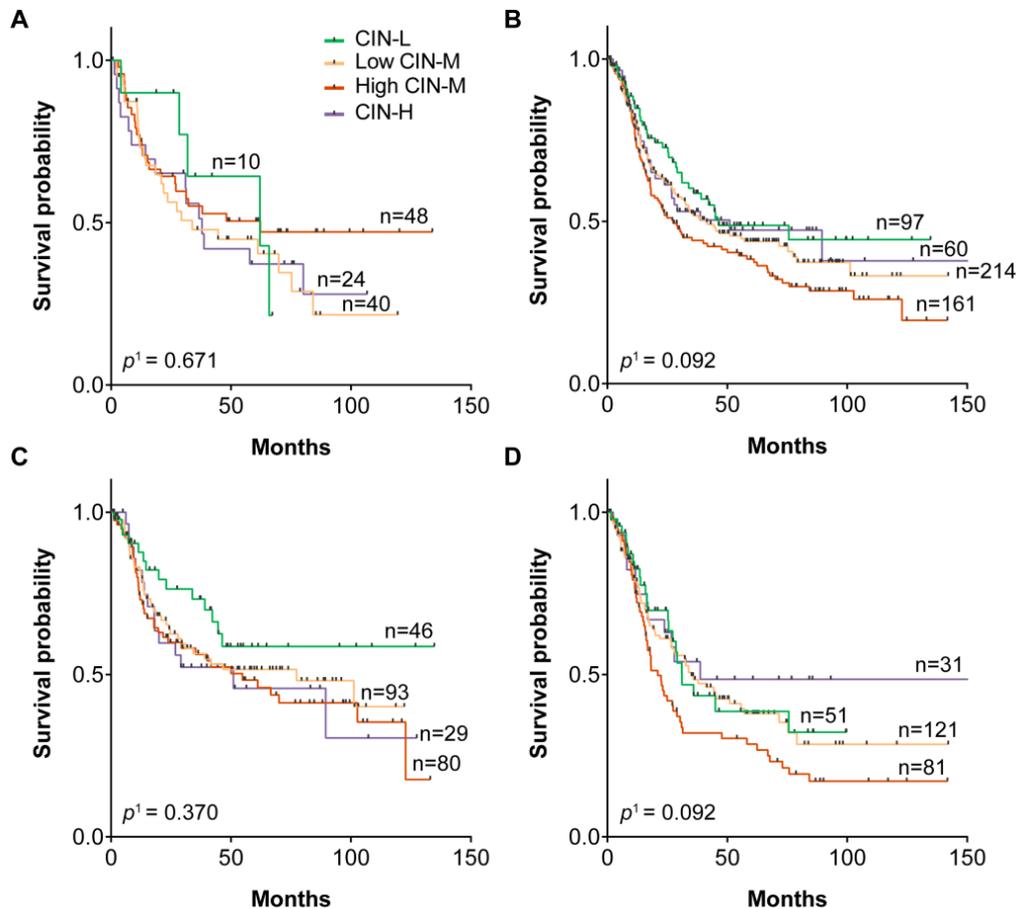
### 3.4.2.3. Modified CIN classification and association with survival

OS was compared for the four subgroups CIN-L, low CIN-M, high CIN-M and CIN-H in the tumor biopsy and resected tumor cohort taken the first as reference group. The subgroups were determined according to the quartiles of the AI ratios. The resected tumor cohort was additionally stratified according to CTx treatment (yes/no). Kaplan Meier curves are shown in Figure 44 and all survival data of the patient cohorts and subgroups are summarized in Supplementary Table 17.

Comparison of OS of patients with tumor biopsies before CTx regarding the four subgroups of CIN revealed no statistically significant difference (overall log rank *p* = 0.671; Figure 44A). Patients with CIN-L and high CIN-M tumors showed the best OS with a median survival of 62 months compared to the other two remaining subgroups. Within the patients with CIN-H and low CIN-M tumors only little differences in OS were observed.

In the resected tumors no statistically significant difference was observed overall four subgroups (*p* = 0.092; Figure 44B). Patients with CIN-L tumors followed by CIN-H revealed the best survival and patients with high CIN-M tumors showed significantly the worst survival (HR, 1.57; 95% CI, 1.07-2.29; *p* = 0.021) with a median survival of 27.4 months. In the non-CTx group no statistically significant difference was observed regarding the four CIN-subgroups (*p* = 0.370; Figure 44C). The CIN-L group showed the best survival compared to the others with a 5 years OS rate of 58.7%.

In the CTx group an evident but not statistically significant difference was shown over all four subgroups ( $p = 0.092$ ; Figure 44D). Patients with high CIN-M revealed the worst survival (HR, 1.53; 0.94-2.49;  $p = 0.087$ ) with a median survival of 21 months compared to 35.1 months of patients with low CIN-M tumors and 38.7 months with CIN-H tumors.



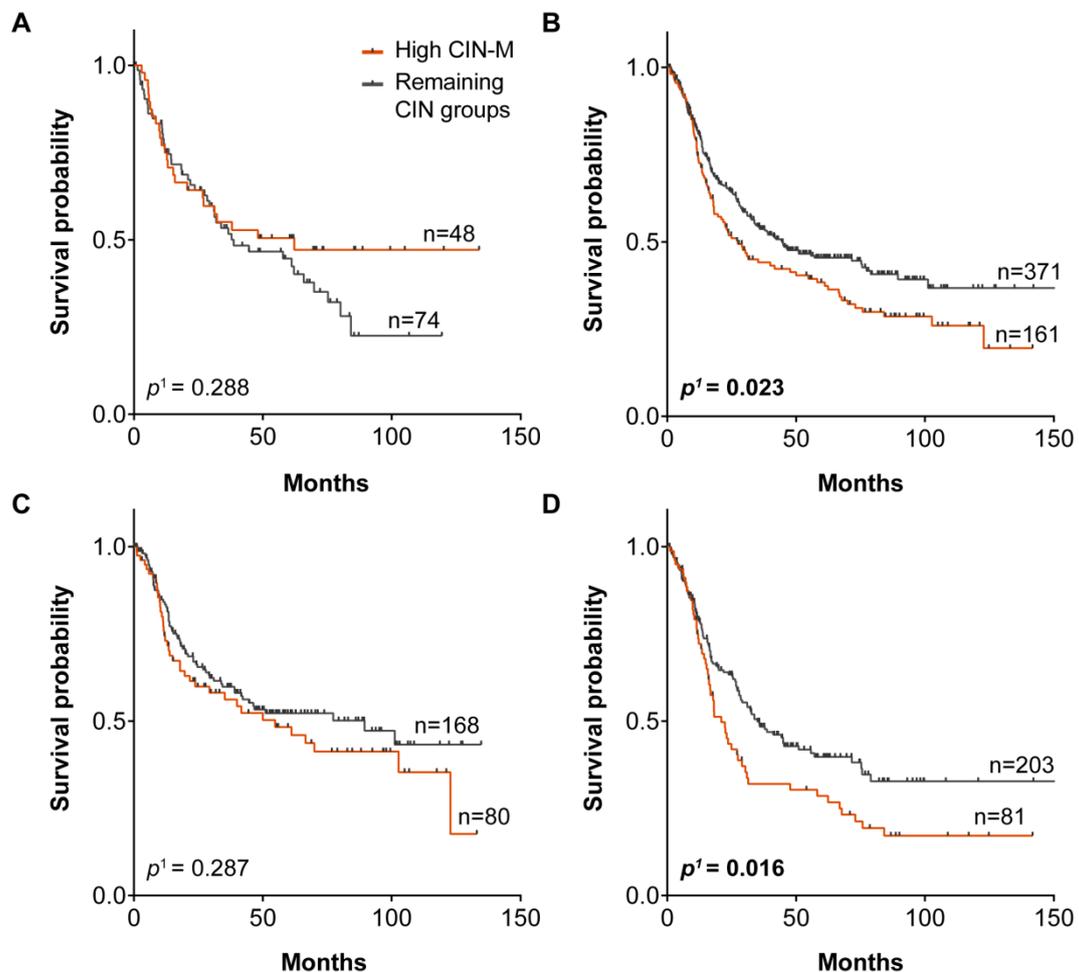
**Figure 44: Modified CIN classification according to the quartiles of AI ratios and survival of the patients**

Kaplan-Meier curves are shown for patients with tumor biopsies before neoadjuvant CTx (A) and for patients with resected tumors: all patients (B), patients treated without CTx (C), patients treated with neoadjuvant CTx (D).  $^1p$ -values (overall) of log rank test

Due to the fact that patients with high CIN-M tumors had the worst survival in the resected tumor cohort and patients with low CIN-M tumors had more or less survival rates similar to CIN-H and -L patients, the three groups CIN-L, low CIN-M and CIN-H were combined in one group. Patients with tumors having these subtypes were termed as remaining CIN-groups and compared to high CIN-M. All survival data are summarized in Table 30 and survival curves are shown in Figure 45.

Comparison of OS of patients with tumor biopsies before CTx regarding the two subgroups of CIN revealed no statistically significant difference ( $p = 0.288$ ; Figure 45A). Patients with high CIN-M tumors showed a better survival compared to the remaining CIN-groups (HR, 0.76; 95% CI, 0.46-1.26) with a median survival of 62.2 months compared to 38.7 months.

In the resected tumor cohort a statistically significant difference was observed ( $p = 0.023$ ; Figure 45B). Especially in the CTx group, patients with high CIN-M tumors showed significantly the worst survival (HR, 1.49; 95% CI, 1.07-2.08;  $p = 0.016$ ; Figure 45D). In the non-CTx group the difference between the two subgroups was not statistically significant. However, the high CIN-M group revealed a worse survival compared to the remaining CIN-groups (HR, 1.24; 95% CI, 0.84-1.83;  $p = 0.287$ ; Figure 45C). Additionally, subgroup analysis within all resected specimens stratified according to tumor localization and Laurén histopathological subtypes was performed. Patients with high CIN-M tumors and proximal tumor localization showed the worst survival compared to non-proximal located tumors, although the differences were not statistically significant ( $p = 0.083$ ). Comparison of OS of the patients stratified according to Laurén histopathological subtypes revealed no statistically significant difference in both groups ( $p = 0.117$  for intestinal type tumors and  $p = 0.062$  for non-intestinal type tumors). All survival data and survival curves of the patient subgroups are shown in Supplementary Table 18 and Supplementary Figure 4.



**Figure 45: CIN-status in two subgroups according to the quartiles of AI ratios and survival of the patients**  
Kaplan-Meier curves are shown for patients with tumor biopsies before neoadjuvant CTx (A) and for patients with resected tumors: all patients (B), patients treated without CTx (C), patients treated with neoadjuvant CTx (D).  
<sup>1</sup> $p$ -values of log rank test;  $p$ -values < 0.05 in bold

**Table 30: Survival data of the patient cohorts and subgroups in association with high CIN-M versus the remaining CIN-groups**

	CIN classification according to quartiles	No.	Events	Survival probability [%]			Median survival [months] (95% CI)	HR (95% CI)	p-value <sup>1</sup>
				1 yr	3 yrs	5 yrs			
<b>Tumor biopsies before CTx</b>	High CIN-M	48	24	75.0	55.1	50.5	62.2	0.76 (0.46-1.26)	0.288
	Remaining CIN-groups	74	44	76.0	53.4	44.6	38.7 (12.0-65.4)	1 ref.	
	Total	122	68	75.6	54.1	47.1	44.6 (18.5-70.8)	-	
<b>All resected specimens</b>	High CIN-M	161	95	74.0	44.1	38.4	27.4 (17.6-37.2)	1.34 (1.04-1.72)	<b>0.023</b>
	Remaining CIN-groups	371	169	81.0	53.7	45.5	44.6 (24.1-56.1)	1 ref.	
	Total	532	264	78.9	50.7	43.2	39.0 (29.3-48.8)	-	
<b>Resected tumors without CTx</b>	High CIN-M	80	40	74.4	56.2	48.3	54.8 (23.7-85.9)	1.24 (0.84-1.83)	0.287
	Remaining CIN-groups	168	68	83.4	59.9	52.3	89.5 (40.6-138.4)	1 ref.	
	Total	248	108	80.5	58.7	50.8	61.1 (27.5-94.7)	-	
<b>Resected tumors after CTx</b>	High CIN-M	81	55	73.6	32.0	28.6	21.0 (15.5-26.5)	1.49 (1.07-2.08)	<b>0.016</b>
	Remaining CIN-groups	203	101	79.1	48.4	39.8	35.1 (23.9-46.3)	1 ref.	
	Total	284	156	77.5	43.7	36.6	30.3 (25.2-35.4)	-	

<sup>1</sup>p-values of log rank test; p-values < 0.05 in bold

### 3.4.2.4. High CIN-M subgroup and association with patient characteristics

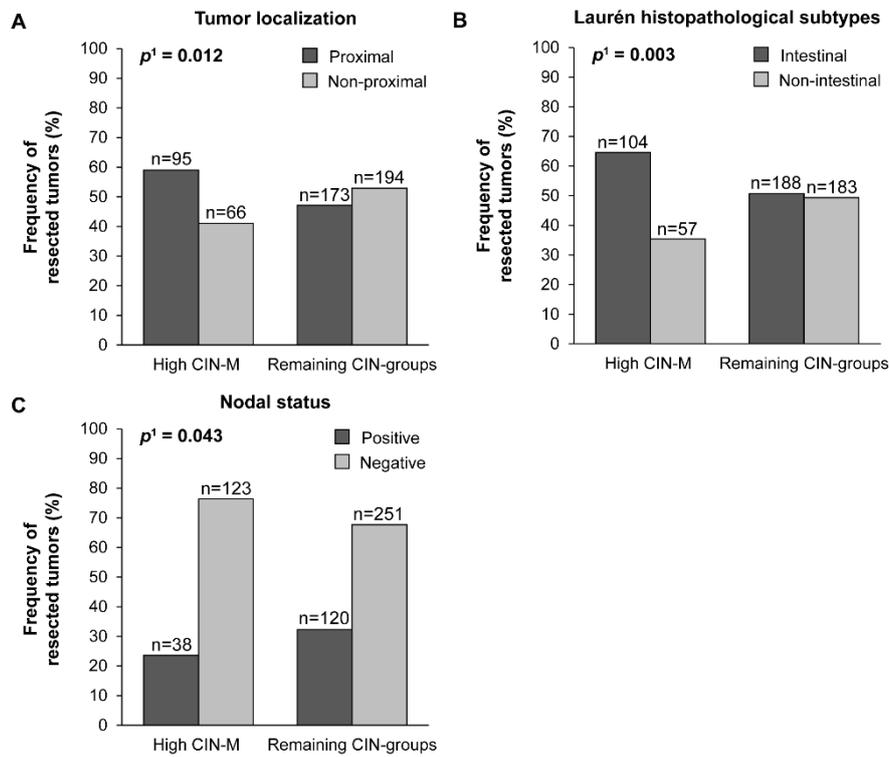
Due to the association of the high CIN-M tumors with the worst OS, high CIN-M was considered as a distinct subgroup of CIN and association with patient characteristics was analyzed for the 532 resected tumors. Results are summarized in Table 31.

High CIN-M was associated with proximal tumor localization ( $p = 0.012$ ), intestinal subtype ( $p = 0.003$ ) and positive lymph node status ( $p = 0.043$ ). Figure 46 shows, that high CIN-M was found in 95 (59%) of the 532 cases in proximal located and in 66 (41%) in non-proximal tumors compared to 173 (47%) and 194 (53%) in the remaining CIN-groups, respectively. The high CIN-M group was found in 104 (65%) of the 532 cases in intestinal and in 57 (35%) in non-intestinal type tumors compared to 188 (51%) and 183 (49%) in the remaining CIN-groups, respectively. High CIN-M was more frequent found in tumors with positive nodal status (76%) compared to the remaining CIN-groups (68%).

**Table 31: High CIN-M subgroup and association with patient characteristics**

Category	Value	Resected tumors ( $n=532$ )		$p$ -value <sup>2</sup>
		Remaining CIN-groups <sup>1</sup> $n$	High CIN-M $n$	
<b>Cases</b>	Total	371	161	
<b>Age [yrs]</b>	Median	63.9	64.4	
	Range	31.7 - 88.3	28.3 - 90.9	
<b>Age Median [yrs]</b>	< 64.3	186	79	0.821
	≥ 64.3	185	82	
<b>Gender</b>	Male	267	124	0.225
	Female	104	37	
<b>Tumor localization</b>	Proximal	173	95	<b>0.012</b>
	Non-proximal	194	66	
	n/a	4	0	
<b>Laurén classification</b>	Intestinal	188	104	<b>0.003</b>
	Non intestinal	183	57	
<b>Tumor grade</b>	G1/2	72	41	0.108
	G3/4	246	97	
	n/a	53	23	
<b>Clinical tumor stage</b>	cT2	88	38	0.993
	cT3/4	282	122	
	n/a	1	1	
<b>(y)pT<sup>3</sup></b>	(y)pT1/2	85	33	0.538
	(y)pT3/4	286	128	
<b>(y)pN<sup>3</sup></b>	Negative	120	38	<b>0.043</b>
	Positive	251	123	
<b>Metastasis status</b>	No	316	137	0.980
	Yes	55	24	
<b>Resection category</b>	R0	275	123	0.579
	R1	96	38	
<b>Neoadjuvant CTx</b>	No	168	80	0.349
	Yes	203	81	

<sup>1</sup>The three subgroups CIN-L, low CIN-M and CIN-H were summarized as one group and termed as remaining CIN-groups; <sup>2</sup> $p$ -values of Chi-square test or Fisher's exact test;  $p$ -values < 0.05 in bold; <sup>3</sup>Classification according to 7<sup>th</sup> Edition UICC

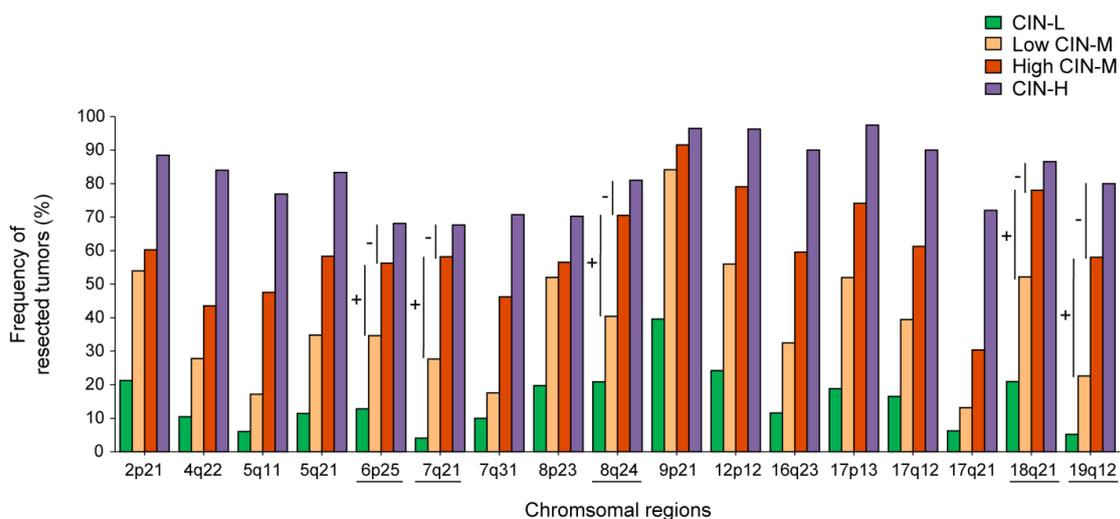


**Figure 46: High CIN-M subgroup and significant associations with clinical parameters**

The frequency of resected tumors without and after CTx categorized into the respective modified CIN-subgroup is shown in relation to tumor localization (A), Laurén histological subtype (B) and nodal status (C).  $^1p$ -values of Chi-square test

### 3.4.2.5. Modified CIN-subgroups and frequency of AI at single microsatellite loci

The frequencies of AI at a single microsatellite locus were analyzed for the modified CIN-subgroups determined according to the quartiles of the AI ratios. The increase of the frequencies from the low CIN-M subgroup to the high CIN-M was compared with the increase of the frequencies from the high CIN-M group to CIN-H per chromosomal region. At 6p25, 7q21, 8q24, 18q21 and 19q12 the increase from high CIN-M to CIN-H is lower and thus marked by a negative sign in Figure 47. At these five chromosomal loci, the increase of the frequencies from low to high CIN-M was 2- to 3-fold higher in comparison to those from high CIN-M to CIN-H. Results comparing the frequencies of AI between the CIN-subgroups are summarized in Supplementary Table 19.



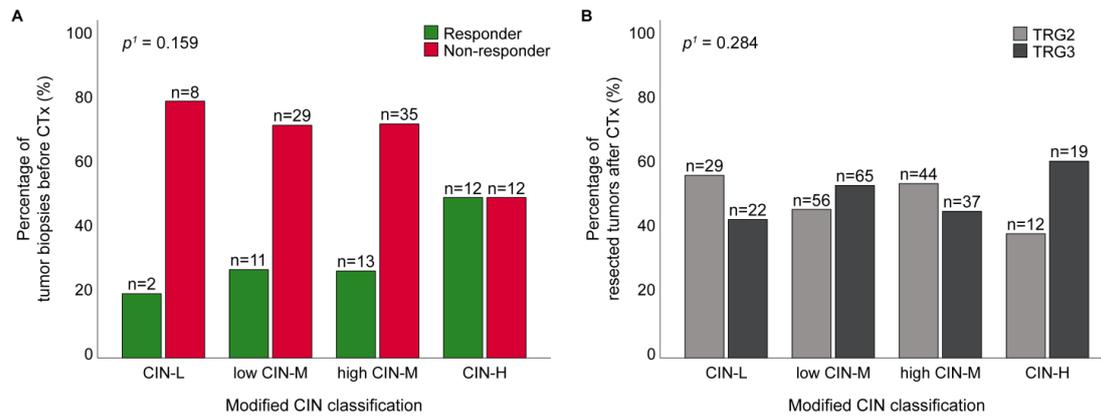
**Figure 47: Modified CIN-subgroups and frequency of AI at single microsatellite loci**

Frequencies of all resected tumors having AI at a respective chromosomal region and classification into the four modified CIN-subgroups according to the quartiles of the AI ratios. The increase of the frequencies from the low CIN-M subgroup to the high CIN-M is higher (positive sign) compared to the increase of the frequencies from high CIN-M to CIN-H which is lower (negative sign) at respective chromosomal regions (underlined).

### 3.4.2.6. Modified CIN classification and response to neoadjuvant CTx

The association of the four CIN-groups with response to neoadjuvant CTx was analyzed in 122 tumor biopsies before CTx. Results of the different constellations of the CIN classifications are summarized in Table 32.

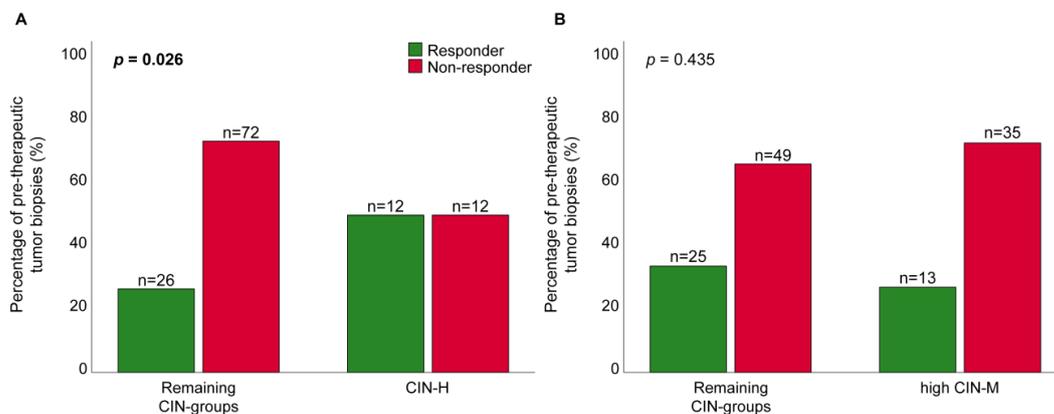
The four molecular subgroups were not significantly associated with response to neoadjuvant CTx ( $p = 0.159$ , Figure 48A). However, 12 (50%) of the 24 tumor biopsies with CIN-H were of responding and 12 (50%) of non-responding patients (Table 32). The association of the four CIN-groups based on the quartiles of AI ratios of the modified classification with the tumor regression grade was additionally analyzed on 284 resected tumors after neoadjuvant CTx and revealed no significant difference ( $p = 0.284$ , Figure 48B).



**Figure 48: Modified CIN classification according to the quartiles of AI ratios and association with response to neoadjuvant CTx**

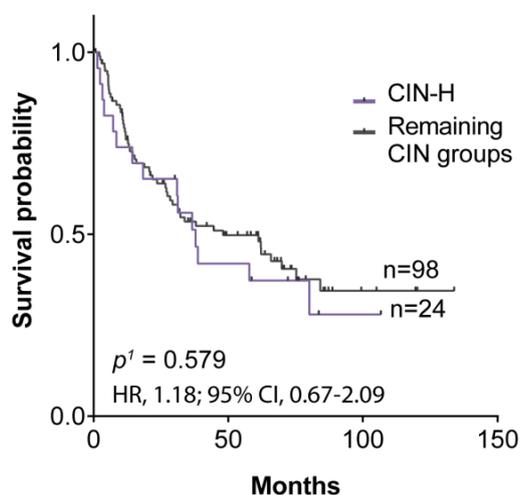
The response of patients with tumor biopsies before CTx (A) and tumor regression of patients with resected tumors after CTx (B) is shown in relation to the modified CIN classification.  $p$ -values (overall) of Chi-square test

Comparing the CIN-H subgroup to the remaining CIN-groups showed that 12 (50%) of the 24 tumors with CIN-H were of responding patients, whereas 26 (26.5%) of the 98 tumors of the remaining CIN-groups were of responding and 72 (73.5%) of non-responding patients (Table 32). The CIN-H subgroup was significantly associated with better response to neoadjuvant CTx compared to the remaining CIN-groups ( $p = 0.026$ ; Figure 49A). In contrast, the high CIN-M group was not associated with better response ( $p = 0.435$ ; Figure 49B). The corresponding survival curves of patients with tumor biopsies of the CIN-H group and remaining CIN-groups are shown in Figure 50.



**Figure 49: Modified CIN-subgroups and association with response to neoadjuvant CTx**

The response of patients with tumor biopsies before CTx is shown in relation to the CIN-H subgroup (A) and high CIN-M subgroup (B).  $p$ -values of Chi-square test



**Figure 50: CIN-H subgroup and survival of patients with tumor biopsies before CTx**

<sup>1</sup>*p*-value of log rank test

**Table 32: Modified CIN classification according to quartiles of the AI ratios and association with response to neoadjuvant CTx**

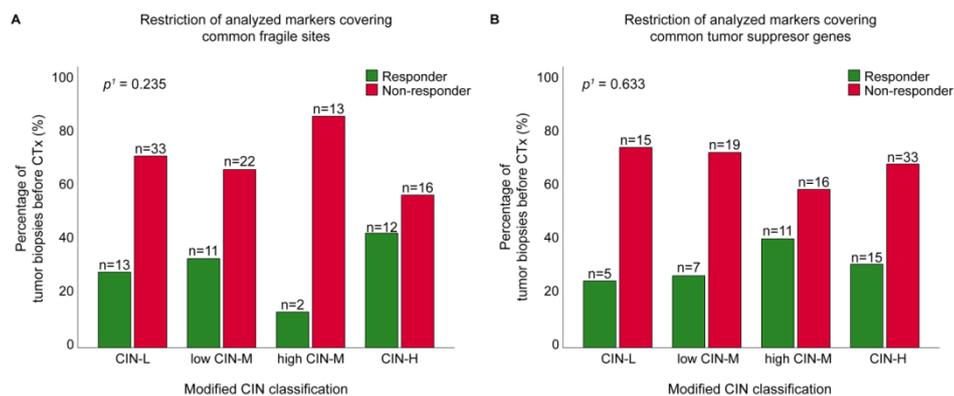
Number of tumor biopsies before CTx (n=122)							
CIN classification in four subgroups							
	CIN-L	Low CIN-M	<i>p</i> -value <sup>1</sup>	High CIN-M	<i>p</i> -value <sup>1</sup>	CIN-H	<i>p</i> -value <sup>1</sup>
Responder (TRG1)	2	11		13		12	
Non-responder (TRG2/3)	8	29	0.629	35	0.642	12	0.105
Total	10	40		48		24	
CIN-H versus remaining CIN-groups							
	Remaining CIN-groups <sup>2</sup>		CIN-H	<i>p</i> -value <sup>3</sup>			
Responder (TRG1)	26		12				
Non-responder (TRG2/3)	72		12	<b>0.026</b>			
Total	98		24				
High CIN-M versus remaining CIN-groups							
	Remaining CIN-groups <sup>4</sup>		High CIN-M	<i>p</i> -value <sup>3</sup>			
Responder (TRG1)	25		13				
Non-responder (TRG2/3)	49		35	0.435			
Total	74		48				

<sup>1</sup>*p*-values of Chi-square test compared to CIN-L (ref.); <sup>2</sup>CIN-L, low CIN-M and high CIN-M were summarized as one CIN-group; <sup>3</sup>*p*-values of Chi-square test; <sup>4</sup>CIN-L, low CIN-M and CIN-H were summarized as one CIN-group; *p*-values < 0.05 in bold

### 3.4.3. Restriction of the number of analyzed microsatellite markers and CIN classification

The number of analyzed microsatellite markers was reduced considering functional aspects and CIN classification based on quartiles of the AI ratios of the tumors was performed. Therefore, only microsatellite loci covering common fragile sites or tumor suppressor genes were selected. According to this, the five chromosomal regions 4q22, 5q11, 6p25, 8p23 and 16q23 covering fragile sites and 5q21, 17p13 and 18q21 covering tumor suppressor genes were chosen. AI and CIN were calculated with the restricted number of microsatellite markers in the tumor biopsy and resected tumor cohort. The association of the four CIN-subgroups defined with a restricted number of microsatellite loci with response to CTx was analyzed in the tumor biopsies before CTx and patient's survival was also compared for the four subgroups.

The four CIN-subgroups defined by a restricted number of microsatellite markers covering common loci of fragile sites ( $p = 0.235$ ) or tumor suppressor genes ( $p = 0.633$ ) were not associated significantly with response to CTx (Figure 51).

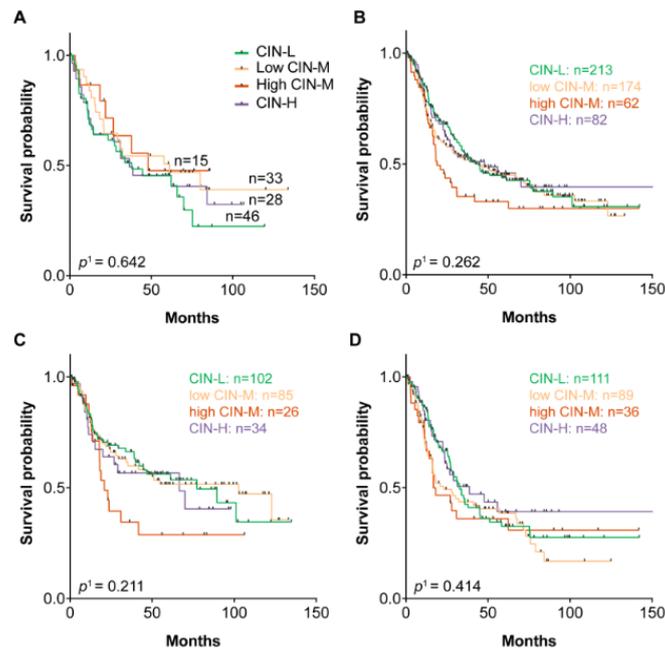


**Figure 51: CIN-subgroups defined by a restricted number of microsatellite markers and association with response to neoadjuvant CTx**

The percentage of tumor biopsies before CTx is shown for the four CIN-groups defined by a restricted number of microsatellite markers covering common loci of fragile sites (A) and tumor suppressor genes (B).  $p$ -values (overall) of Chi-square test

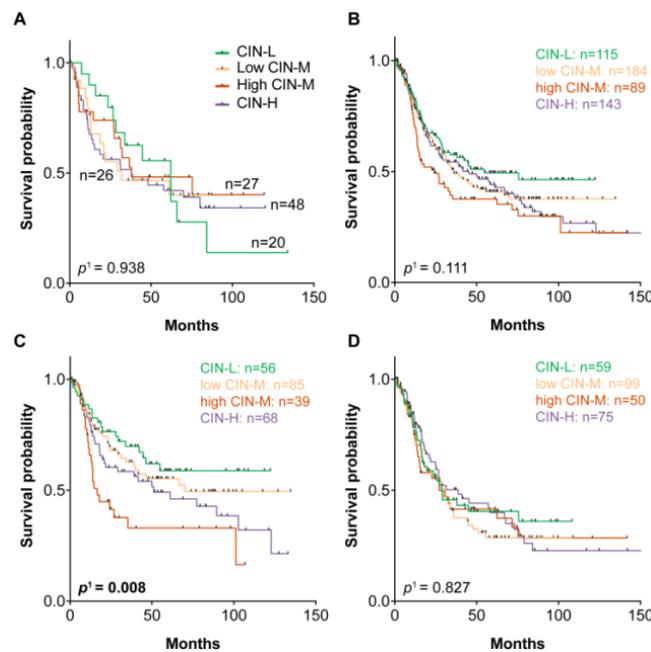
Comparison of survival of patients with tumor biopsies before CTx regarding the four CIN-subgroups defined by a restricted number of microsatellite loci covering common fragile sites ( $p = 0.0.642$ ; Figure 52A) or tumor suppressor genes ( $p = 0.938$ ; Figure 53A) revealed no statistically significant differences. The resected tumors demonstrated a slight difference regarding the four CIN-subgroups with a reduced number of microsatellite loci covering fragile sites ( $p = 0.262$ ; Figure 52B). As shown previously, patients with high CIN-M tumors showed the worst survival even at reduction of the number of microsatellite markers. Especially in the non-CTx patient group, patients with high CIN-M tumors showed a worse OS in both classification systems compared to the remaining CIN-groups ( $p = 0.211$  for fragile sites and  $p = 0.008$  for tumor suppressor genes; Figure 52C and Figure 53C).

In the CTx patient group no significant differences were observed in neither of the two classification systems ( $p = 0.414$  for fragile sites and  $p = 0.827$  for tumor suppressor genes; Figure 52D and Figure 53D).



**Figure 52: Overall survival discriminated by the CIN-status in four subgroups according to chromosomal regions covering common fragile sites**

Kaplan-Meier curves are shown for patients with tumor biopsies before neoadjuvant CTx (A) and for patients with resected tumors: all patients (B), patients treated without CTx (C), patients treated with neoadjuvant CTx (D). <sup>1</sup> $p$ -values (overall) of log rank test



**Figure 53: Overall survival discriminated by the CIN-status in four subgroups according to chromosomal regions covering common tumor suppressor genes**

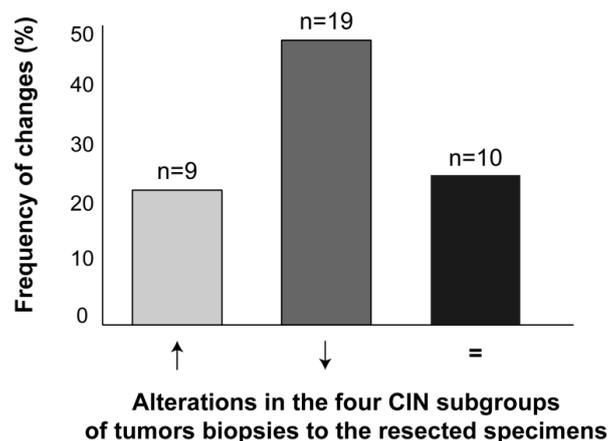
Kaplan-Meier curves are shown for patients with tumor biopsies before neoadjuvant CTx (A) and for patients with resected tumors: all patients (B), patients treated without CTx (C), patients treated with neoadjuvant CTx (D). <sup>1</sup> $p$ -values (overall) of log rank test

### 3.4.4. CIN classification of paired tumor biopsies and resected tumors

To determine tumor cell plasticity in the context of neoadjuvant CTx, the classification of CIN was analyzed and compared in 38 corresponding tumor biopsies before CTx and resected tumors after CTx.

CIN classification according to TCGA in two subgroups showed that 33 (87%) of the 38 corresponding tumor biopsies and resected tumors after CTx had an identical CIN-status, three tumor biopsies before CTx changed from the CIN-L subgroup into CIN-H in the resected tumor and two from CIN-H into CIN-L.

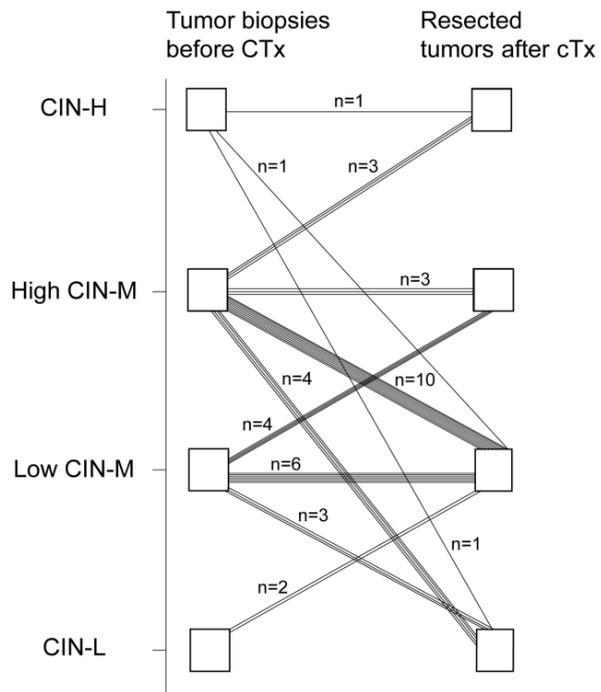
Comparison of the paired tumor biopsies and resected tumors classified according to the modified CIN classification in four subgroups based on quartiles of the AI ratios revealed that 10 (26%) of the 38 tumors were classified in identical CIN-groups (=), 9 (24%) tumor biopsies before CTx changed their CIN-status from a lower group into an higher group (↑) in the resected specimen after CTx and 19 (50%) of the tumor biopsies changed from a higher CIN subgroup into a lower one (↓) (Figure 54).



**Figure 54: Alterations in the four modified CIN-subgroups between corresponding tumor biopsies before CTx and resected tumors after CTx**

Identical CIN classification between tumor biopsies and resected tumors (=); Change from lower CIN-group in the tumor biopsies to higher one in the corresponding resected tumors (↑); Change from higher CIN-group in the tumor biopsies to lower one in the corresponding resected tumors (↓).

Alterations in the four CIN-subgroups between pre-therapeutic tumor biopsies before CTx and the corresponding post-therapeutic resected tumors after CTx are shown in detail in Figure 55. Between the TRG2 and TRG3 tumors no essential differences were observed.

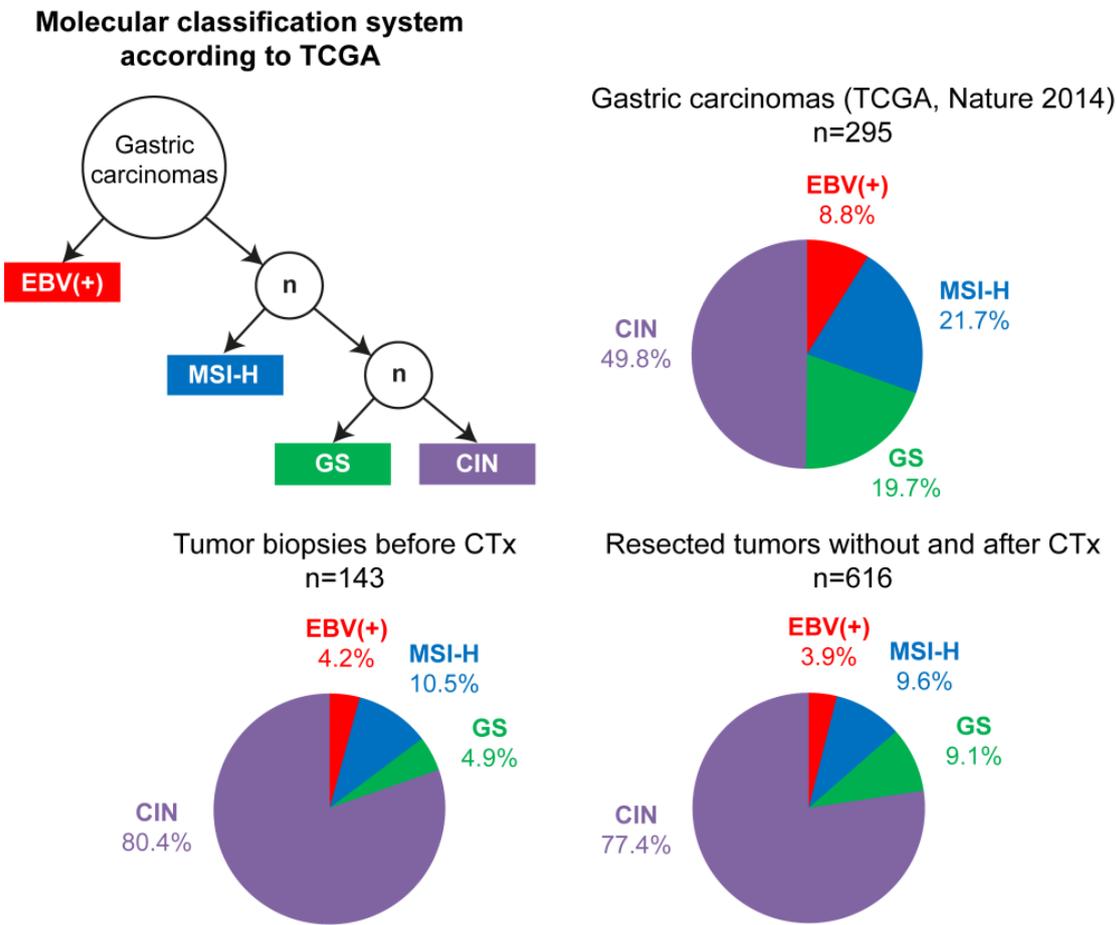


**Figure 55: Alterations in the four modified CIN-subgroups between pre-therapeutic tumor biopsies and corresponding post-therapeutic resected tumors**

### 3.5. Original and modified TCGA classification

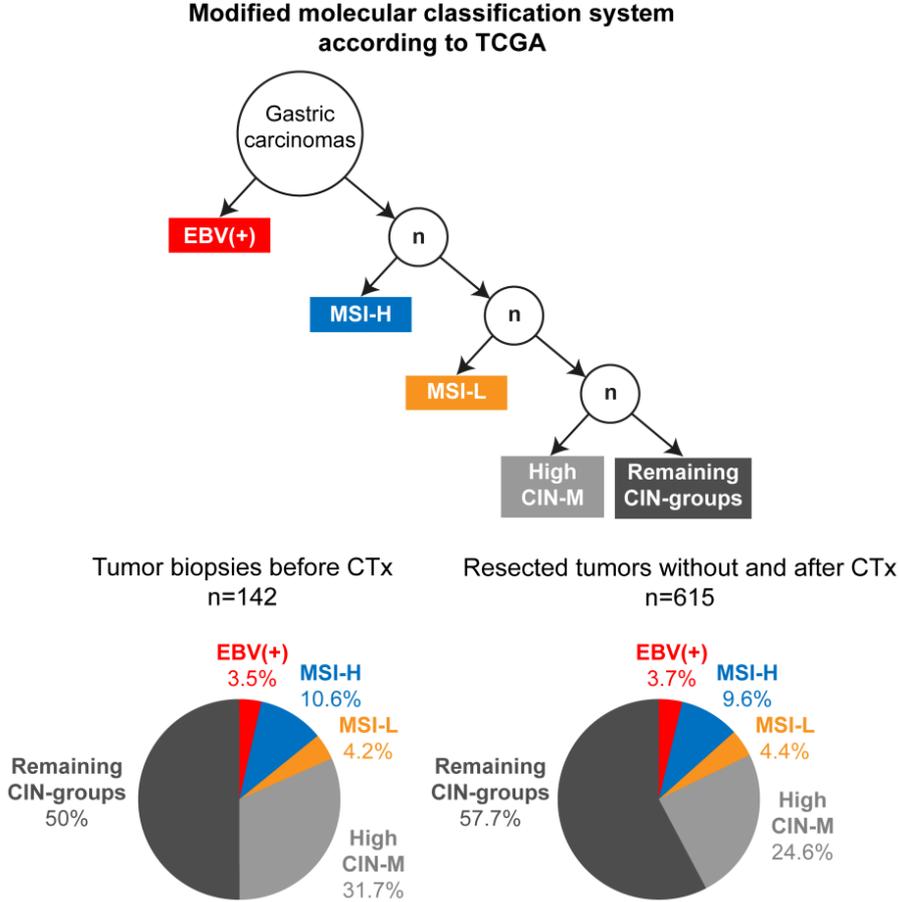
#### 3.5.1. Molecular classification algorithms and frequency of molecular subgroups

The tumors were first classified according to the classification algorithm of the TCGA study as shown in Figure 56. The frequencies of the molecular subtypes are shown for the 295 gastric carcinomas included in the TCGA study and for the 143 tumor biopsies before CTx and 616 resected tumors analyzed in this study. Accordingly, 6 (4.2%) of the 143 analyzed pre-therapeutic tumor biopsies were classified as EBV(+), 15 (10.5%) as MSI-H, 7 (4.9%) as GS and 115 (80.4%) as CIN. In the resected tumor cohort, 24 (3.9%) patients had tumors with EBV(+), 59 (9.6%) with MSI-H, 56 (9.1%) with GS and 477 (77.4%) with CIN.



**Figure 56: Molecular classification algorithm according to TCGA and frequency of molecular subgroups**  
 The tumors were classified into the molecular subtypes EBV positive (red), MSI-H (blue), GS (green) and CIN (purple) according to the shown flowchart. The frequencies of the subtypes are shown for the gastric carcinomas included in the TCGA study and for the tumor biopsies before CTx and resected tumors analyzed in this study.

Based on the recently published EBV and MSI data in the Journal of Pathology Clinical Research of the patient cohorts (Kohlruss, Grosser et al. 2019), a modified molecular classification system in five molecular subgroups was proposed as shown in Figure 57. The exception of this classification system is that tumors were classified as MSI-H and MSI-L and the tumors, which were negative for EBV and MSI, were classified according to their grade of chromosomal instability into two subtypes of CIN. The EBV(+) and MSI-H subgroups were in both classification systems identical. One tumor biopsy and one resected tumor which were positive for MSI-L and EBV were excluded in this molecular classification. The modified classification system revealed 6 (4.2%) of the 142 tumor biopsies classified as MSI-L, 45 (31.7%) as high CIN-M and 71 (50%) tumors were classified into the remaining CIN-groups. In the resected tumor cohort, 27 (4.4%) of the 615 tumors were classified as MSI-L, 151 (24.6%) as high CIN-M and 355 (57.7%) were classified into the remaining CIN-groups (Figure 57).



**Figure 57: Modified molecular classification system and frequency of molecular subgroups**  
 The tumors were classified into the molecular subtypes EBV positive (red), MSI-H (blue), MSI-L (orange), high CIN-M (light gray) and remaining CIN-groups (dark gray) according to the shown flowchart. The frequencies of the subtypes are shown for the tumor biopsies before CTx and resected tumors analyzed in this study.

### 3.5.2. Molecular classification and association with patient characteristics

The association of the molecular subgroups with clinical characteristics was analyzed for all patients with resected tumors for both molecular classification systems.

Overall, the original molecular classification system according to TCGA was statistically significant associated with age ( $p < 0.001$ ), tumor localization ( $p < 0.001$ ), Laurén classification ( $p < 0.001$ ), tumor grade ( $p = 0.006$ ) and resection status ( $p = 0.032$ ). EBV(+) occurred more frequently in male patients, was associated with tumor localization in the middle of the stomach and with intestinal tumor type. MSI-H arose more frequently in patients with older age, was associated with intestinal type tumors and with absence of metastasis. The CIN subgroup was more frequently found in proximal located tumors, was associated with intestinal type tumors and poor differentiation. Results are summarized in Table 33 and  $p$ -values for each molecular subgroup with defining the GS subgroup as reference is shown in detail in Supplementary Table 20.

**Table 33: Molecular classification according to TCGA and association with patient characteristics**

Category	Value	Resected tumors (n=616)				$p$ -value <sup>1</sup>
		GS [n]	EBV(+) [n]	MSI-H [n]	CIN [n]	
<b>Cases</b>	Total	56	24	59	477	
<b>Age [yrs]</b>	Median	64.9	57.4	71.4	64.1	
	Range	33.4-85.2	29.2-80.9	40.4-84.9	28.3-90.9	
<b>Age Median [yrs]</b>	<64.6	27	17	13	249	<b>&lt;0.001</b>
	≥64.6	29	7	46	228	
<b>Gender</b>	Male	40	22	38	352	0.082
	Female	16	2	21	125	
<b>Tumor localization</b>	Proximal	15	9	23	253	<b>&lt;0.001</b>
	Non-proximal	41	14	36	220	
	NA	0	1	0	4	
<b>Laurén classification</b>	Intestinal	15	15	39	278	<b>&lt;0.001</b>
	Non-intestinal	41	9	20	199	
<b>Tumor grade</b>	G1/2	5	1	11	108	<b>0.016</b>
	G3/4	44	18	38	299	
	n/a	7	5	10	70	
<b>Clinical tumor stage</b>	cT2	15	6	12	111	0.873
	cT3/4	41	18	47	364	
	n/a	0	0	0	2	
<b>(y)pT<sup>2</sup></b>	(y)pT1/2	14	6	11	104	0.845
	(y)pT3/4	42	18	48	373	
<b>(y)pN<sup>2</sup></b>	Negative	18	9	21	141	0.675
	Positive	38	15	38	336	
<b>Metastasis status</b>	No	45	23	56	409	0.060
	Yes	11	1	3	68	
<b>Resection category</b>	R0	47	23	47	352	<b>0.032</b>
	R1	9	1	12	125	
<b>Neoadjuvant CTx</b>	No	26	8	35	222	0.143
	Yes	30	16	24	255	

<sup>1</sup> $p$ -values (overall) of Chi-square or Fisher's exact test;  $p$ -values < 0.05 in bold; <sup>2</sup>Classification according to 7<sup>th</sup> Edition UICC

Similar to the findings according to the original TCGA classification, the modified classification system was statistically significant associated with age ( $p < 0.001$ ), gender ( $p = 0.026$ ), Laurén classification ( $p = 0.004$ ) and tumor grade ( $p = 0.052$ ). It was noticeable that MSI-L occurred more frequently in intestinal type tumors and high CIN-M was associated with proximal tumor localization and intestinal subtype compared to the remaining CIN-groups. Results are summarized in Table 34 and  $p$ -values for each molecular subgroup with the remaining CIN-groups defined as reference is shown in detail in Supplementary Table 21.

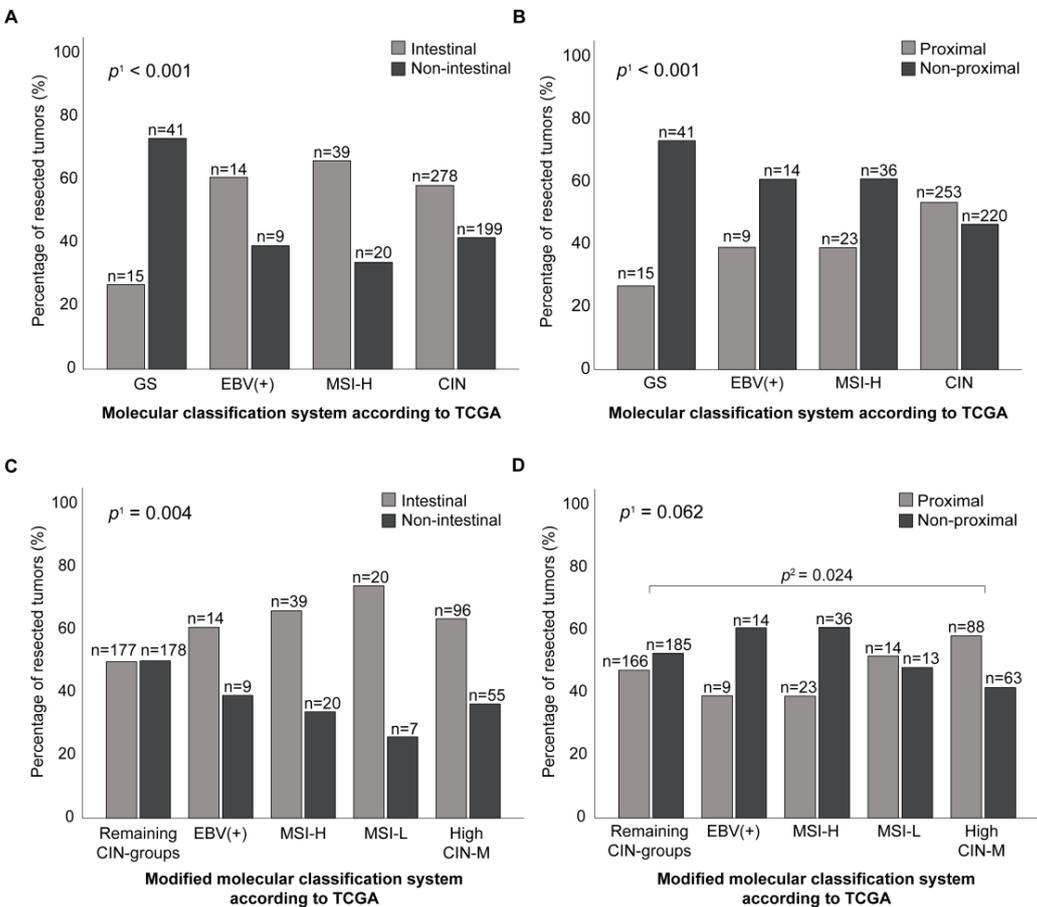
**Table 34: Modified molecular classification based on TCGA and association with patient characteristics**

Category	Value	Resected tumors (n=615 <sup>1</sup> )					$p$ -value <sup>2</sup>
		Remaining CIN-groups [n]	EBV(+) [n]	MSI-H [n]	MSI-L [n]	High CIN-M [n]	
<b>Cases</b>	Total	355	23	59	27	151	
<b>Age [yrs]</b>	Median	63.7	57.6	71.4	66.5	64.4	
	Range	31.7-88.3	29.2-80.9	40.4-84.9	49.3-82.2	28.3-90.9	
<b>Age Median [yrs]</b>	<64.6	189	16	13	11	76	<b>&lt;0.001</b>
	≥64.6	166	7	46	16	75	
<b>Gender</b>	Male	255	22	38	23	114	<b>0.026</b>
	Female	100	1	21	4	37	
<b>Tumor localization</b>	Proximal	166	9	23	14	88	0.062
	Non-proximal	185	14	36	13	63	
	n/a	4	0	0	0	0	
<b>Laurén classification</b>	Intestinal	177	14	39	20	96	<b>0.004</b>
	Non-intestinal	178	9	20	7	55	
<b>Tumor grade</b>	G1/2	70	0	11	4	39	<b>0.052</b>
	G3/4	233	18	38	20	90	
	n/a	52	5	10	3	22	
<b>Clinical tumor stage (y)pT<sup>3</sup></b>	cT2	83	5	12	5	38	0.907
	cT3/4	271	18	47	22	112	
	n/a	1	0	0	0	1	
	(y)pT1/2	80	6	11	6	32	
<b>(y)pN<sup>3</sup></b>	(y)pT3/4	275	17	48	21	119	0.948
	Negative	114	9	21	8	37	
	Positive	241	14	38	19	114	
<b>Metastasis status</b>	No	300	22	56	24	130	0.163
	Yes	55	1	3	3	21	
<b>Resection category</b>	R0	263	22	47	20	116	0.189
	R1	92	1	12	7	35	
<b>Neoadjuvant CTx</b>	No	159	7	35	14	75	0.111
	Yes	196	16	24	13	76	

<sup>1</sup>One resected tumor without CTx was positive for MSI-L and EBV and was excluded in the modified molecular classification system; <sup>2</sup> $p$ -values (overall) of Chi-square or Fisher's exact test;  $p$ -values < 0.05 in bold; <sup>3</sup>Classification according to 7<sup>th</sup> Edition UICC

The distribution of the molecular subgroups of both classifications systems according to the tumor localization and Laurén classification is shown in Figure 58. EBV(+), MSI-H and CIN were more frequently found in intestinal type tumors compared to the GS subtype (Figure 58A). Regarding the association of the molecular subgroups with tumor localization, CIN occurred more frequently in proximal located tumors compared to the other subgroups, which were more often found in non-proximal tumors (Figure 58B).

MSI-L and high CIN-M, the subgroups included in the modified classification system, were more frequently found in intestinal type tumors especially in comparison to the remaining CIN-groups which revealed a balanced occurrence of intestinal and non-intestinal tumors (Figure 58C). Figure 58D shows, that the high CIN-M subgroup was more often found in proximal located tumors compared to the other molecular subgroups ( $p = 0.024$ ).

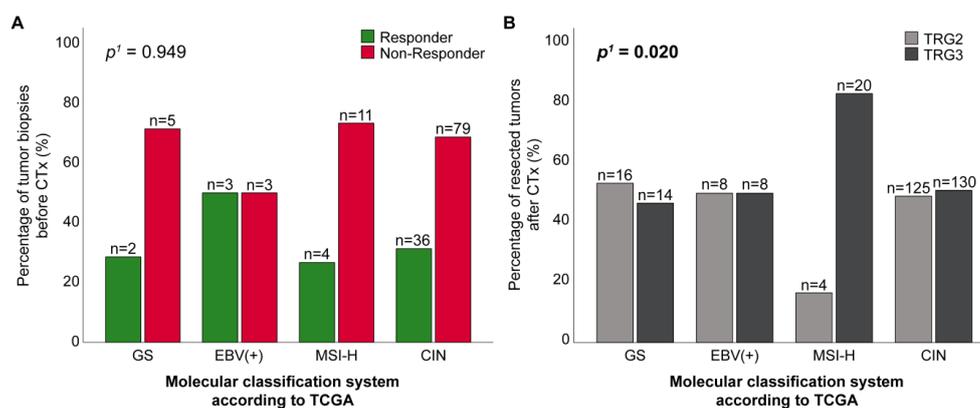


**Figure 58: Molecular classification systems and association with Laurén histological subtypes and tumor localization**

The resected tumors are stratified according to the Laurén classification (A, C) and according to tumor localization (B, D). <sup>1</sup> $p$ -values (overall) of Chi-square test; <sup>2</sup> $p$ -value of Chi-square test compared to remaining CIN-groups (ref.)

### 3.5.3. Molecular classification and association with response to CTx

Overall, the original molecular classification system according to TCGA revealed no significant association with response ( $p = 0.949$ , Figure 59A). In general, the number of non-responding patients was higher in the GS, MSI-H and CIN subgroup and the proportion of responder and non-responder was similar between these subgroups. In the EBV(+) subgroup, three (50%) patients were responder and three (50%) were non-responder (Supplementary Table 22).

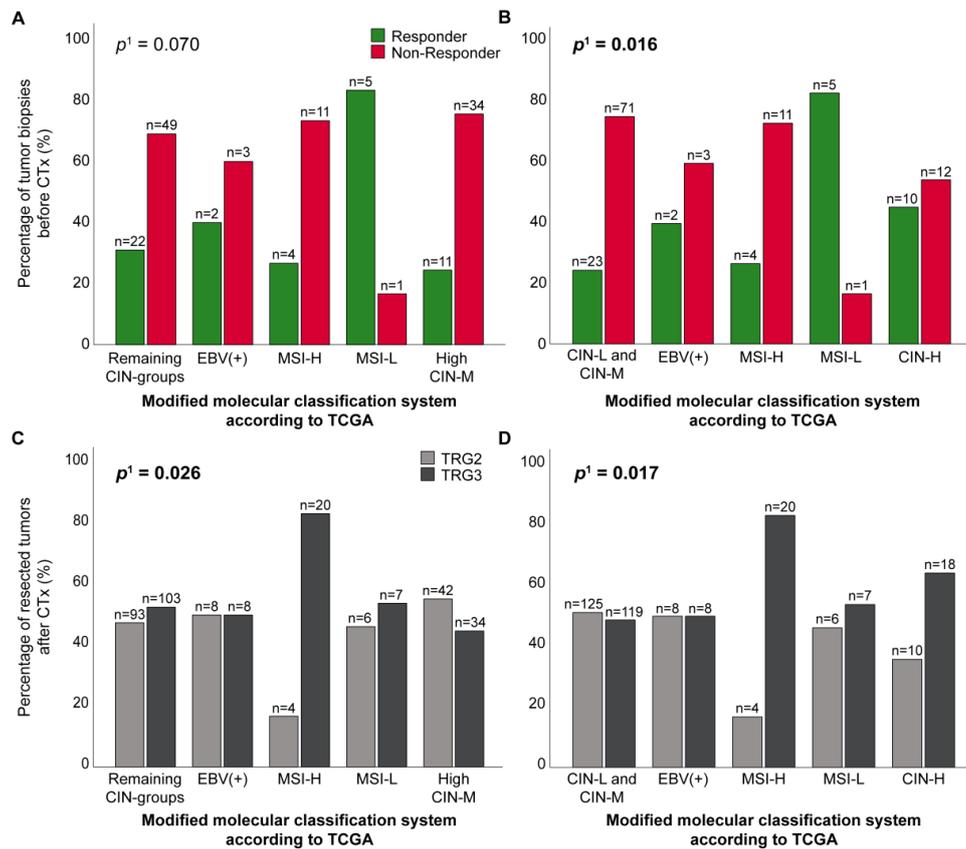


**Figure 59: Molecular classification according to TCGA and association with response to neoadjuvant CTx**  
The response to neoadjuvant CTx is shown for patients with tumor biopsies before CTx (A) and tumor regression for patients with resected tumors after neoadjuvant CTx (B) in relation to the respective molecular subgroup. <sup>1</sup> $p$ -values (overall) of Fisher's exact test;  $p$ -value < 0.05 in bold

The modified molecular classification system with high CIN-M as an own subgroup showed no statistically significant association with response but a tendency is discernible ( $p = 0.070$ , Figure 60A). It is striking that only MSI-L revealed a significant association with better response ( $p = 0.018$ ) compared to the remaining CIN-groups (Supplementary Table 23). Five (83%) of the 6 tumor biopsies with MSI-L were of responding patients with TRG1 compared to 22 (31%) of 71 tumors with remaining CIN-groups (Figure 60A).

As shown previously in chapter 3.4.2.6, the molecular subgroup CIN-H was significantly associated with better response to neoadjuvant CTx compared to the remaining CIN-groups ( $p = 0.026$ ; Figure 49A and Table 32). Ten (45%) of the 22 tumors with CIN-H were of responding patients compared to 23 (25%) of the 94 tumors with remaining CIN-groups (Supplementary Table 24). Thus, the modified classification system with CIN-H as own subgroup revealed an overall statistically significant difference with response ( $p = 0.016$ , Figure 60B).

Additionally, the association of the molecular subgroups of both classification systems with the tumor regression grade was analyzed for the 325 resected tumors after neoadjuvant CTx. In both classification systems, MSI-H was significantly associated with TRG3 tumors ( $p < 0.005$ ). In the TRG3 group, 20 (12%) of the 172 tumors were MSI-H compared to 4 (3%) of 154 in the TRG2 group (Figure 59B and Figure 60C, D). Results of the association of the different molecular classification systems with TRG are summarized in detail in Supplementary Tables 23-25.



**Figure 60: Modified molecular classification system based on TCGA with high CIN-M or CIN-H as own subgroup and association with response to neoadjuvant CTx**

The response to neoadjuvant CTx is shown for patients with tumor biopsies before CTx (**A, B**) and tumor regression for patients with resected tumors after neoadjuvant CTx (**C, D**) in relation to the respective molecular subgroup. <sup>1</sup>*p*-values (overall) of Fisher's exact test; *p*-values < 0.05 in bold

### 3.5.4. Molecular classification and patients survival

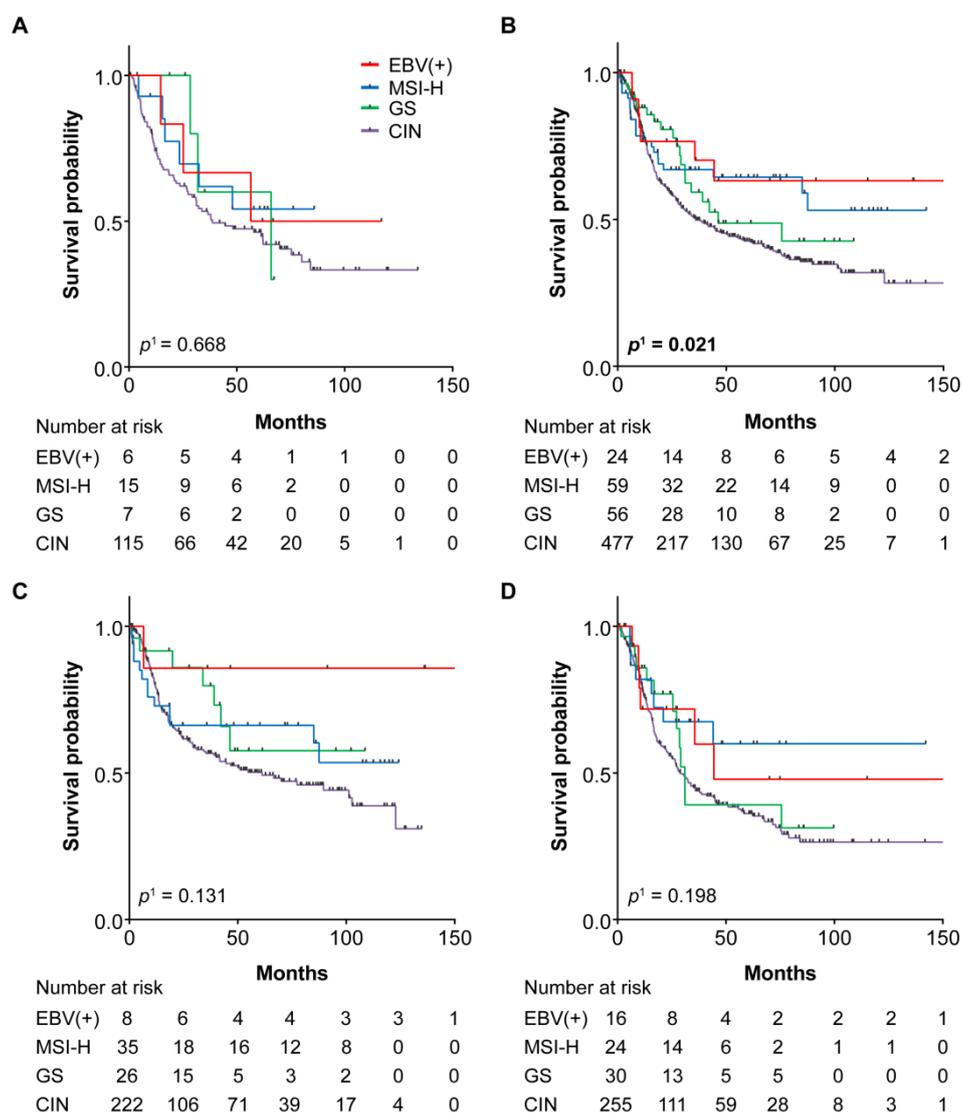
OS was compared for the molecular subgroups in the tumor biopsy and resected tumor cohort. Additionally, analysis of OS of patients in the resected cohort was separately performed in the groups stratified according to their treatment with or without CTx.

#### 3.5.4.1. Molecular classification according to TCGA and association with survival

In the tumor biopsy cohort no statistically significant difference of OS regarding the four molecular subgroups was observed (overall log rank *p* = 0.668, Figure 61A). Patients with EBV(+), MSI-H or genomic stable tumors showed similar survival probabilities with 5-year OS rates of 50%, 54% and 60% respectively (Table 35).

Comparison of OS of all patients with resected tumors regarding the four molecular subgroups revealed a statistically significant difference (overall log rank *p* = 0.021, Figure 61B). The molecular subgroups EBV(+), MSI-H and CIN were compared to the genomic stable tumors (ref). The EBV(+) and MSI-H subgroup demonstrated the best OS compared to the reference group (EBV(+): HR, 0.67, *p* = 0.368; MSI-H: HR, 0.85, *p* = 0.595). Patients with CIN tumors showed the worst survival (HR, 1.40, *p* = 0.151, Table 35).

In the non-CTx and CTx group no statistically significant difference of OS was observed (overall log rank  $p = 0.131$ , Figure 61C and  $p = 0.198$ , Figure 61D). In the non-CTx group, patients with EBV(+) tumors showed the best survival (HR, 0.30;  $p = 0.262$ ) with a 5-years OS rate of 85.7%. In the CTx cohort, patients with EBV(+) and MSI-H tumors showed a better survival compared to GS tumors (EBV(+): HR, 0.81;  $p = 0.672$ ; MSI-H: HR, 0.68;  $p = 0.383$ ). All survival data including the 1, 3 and 5 years survival rates are summarized in Table 35.



**Figure 61: Molecular classification according to TCGA and survival of the patients**

Kaplan-Meier curves are shown for patients with tumor biopsies before neoadjuvant CTx (**A**) and for patients with resected tumors: all patients (**B**), patients treated without CTx (**C**), patients treated with neoadjuvant CTx (**D**). <sup>1</sup> $p$ -values (overall) of log rank test;  $p$ -value < 0.05 in bold

**Table 35: Survival data of the different patient cohorts in association with the molecular classification system according to TCGA**

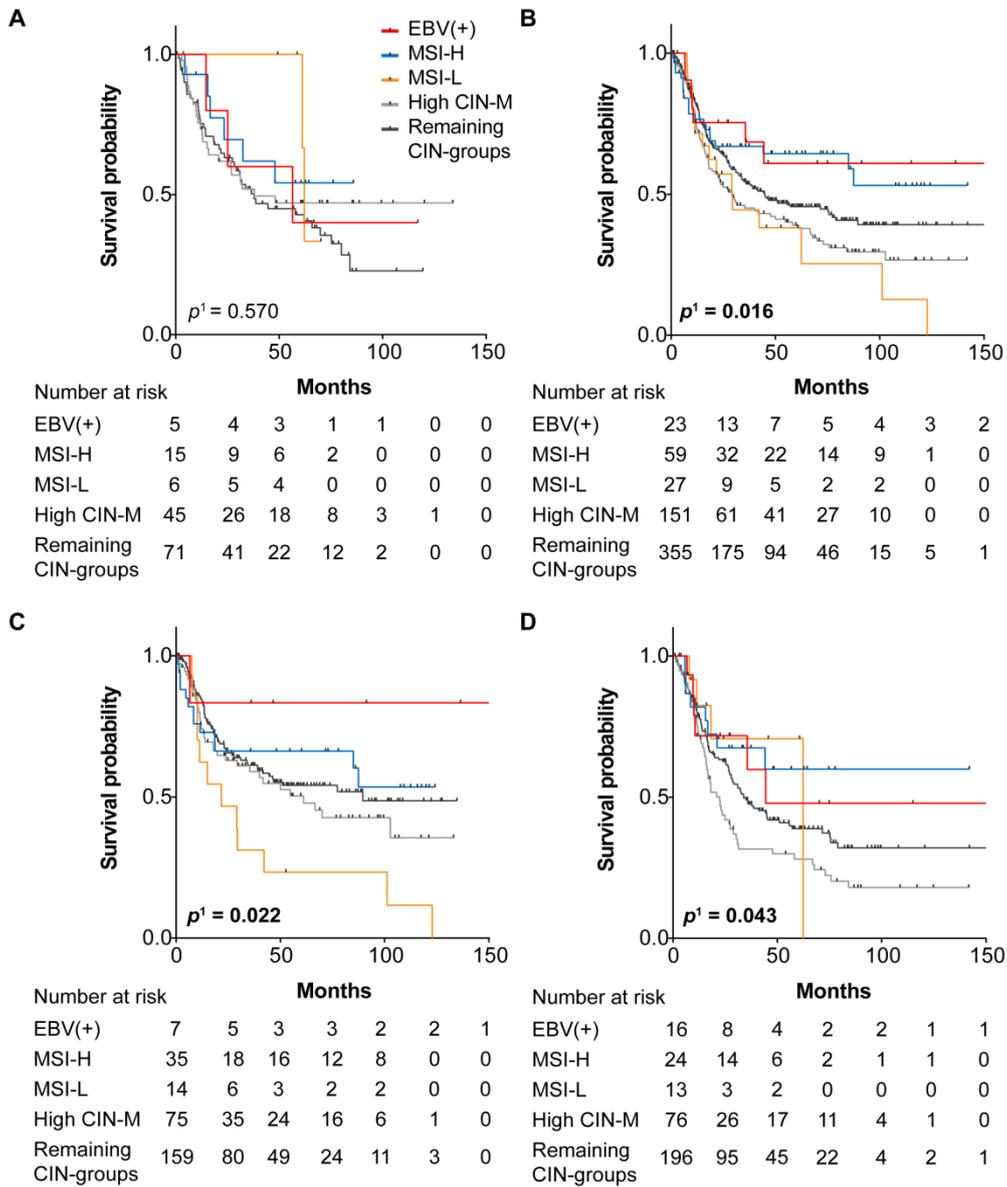
	Molecular classification according to TCGA	No.	Events	Survival probability [%]			Median Survival [months] (95% CI)	HR (95% CI)	p-value <sup>1</sup>
				1 yr	3 yrs	5 yrs			
<b>Tumor biopsies before neoadjuvant CTx</b>	EBV(+)	6	3	100	66.7	50	56.5	1.0 (0.20-4.98)	0.998
	MSI-H	15	6	92.9	61.9	54.2	nr	0.98 (0.25-3.93)	0.979
	CIN	115	65	74.1	53.4	46.2	38.7 (11.8-65.7)	1.47 (0.46-4.67)	0.519
	GS	7	3	100	60	60	65.9 (12.8-119.1)	1 ref	-
	Total	143	77	78.3	55.1	47.3	48.1 (26.2-70)	-	-
<b>All resected tumors</b>	EBV(+)	24	7	76.6	70.2	63.2	nr	0.67 (0.28-1.59)	0.368
	MSI-H	59	21	76.6	66.9	64.4	nr	0.85 (0.46-1.56)	0.595
	CIN	477	244	77.9	49.8	42.7	35.9 (25.3-46.5)	1.40 (0.89-2.20)	0.151
	GS	56	20	88.1	59.3	48.7	46.3 (0-93.0)	1 ref	-
	Total	616	292	78.6	53.1	46.0	44.6 (30.2-59.0)	-	-
<b>Resected tumors without neoadjuvant CTx</b>	EBV(+)	8	1	85.7	85.7	85.7	nr	0.30 (0.04-4.46)	0.262
	MSI-H	35	13	72.9	66.2	66.2	nr	1.17 (0.46-2.93)	0.742
	CIN	222	90	79.5	56.7	50.2	61.1 (27.8-94.4)	1.60 (0.75-3.45)	0.227
	GS	26	6	91.6	79.8	57.6	nr	1 ref	-
	Total	291	122	79.7	60.4	53.8	85.0 (51.7-118.3)	-	-
<b>Resected tumors after neoadjuvant CTx</b>	EBV(+)	16	6	71.8	59.8	47.9	44.0 (na)	0.81 (0.31-2.14)	0.672
	MSI-H	24	8	81.9	67.5	60	nr	0.68 (0.28-1.63)	0.383
	CIN	255	143	76.6	43.9	36.2	29.0 (22.4-35.6)	1.28 (0.72-2.25)	0.399
	GS	30	13	85.5	39.1	39.1	31.0 (28.0-34.0)	1 ref	-
	Total	325	170	77.6	46.4	39.0	32.2 (24.1-40.3)	-	-

<sup>1</sup>p-values of Cox's regression

### 3.5.4.2. Modified molecular classification based on TCGA and association with survival

Similar to the findings in OS regarding the four molecular TCGA subgroups in the tumor biopsy cohort, no statistically significant difference of OS was observed regarding the five modified molecular subgroups (overall log rank  $p = 0.570$ , Figure 62A). The molecular subgroups EBV(+), MSI-H, MSI-L and high CIN-M were compared to the remaining CIN-groups defined as reference. According to the previous published data (Kohlruss, Grosser et al. 2019), MSI-L tumors showed the best survival followed by MSI-H and EBV(+) tumors compared to the reference group (MSI-L: HR, 0.43,  $p = 0.248$ ; MSI-H: HR, 0.61,  $p = 0.250$ ; EBV(+): HR, 0.76,  $p = 0.646$ , Table 36). High CIN-M tumors showed a slightly better survival than the remaining CIN-groups with a 5-years OS rate of 47% compared to 43% (Table 36).

In the resected tumor cohort a statistically significant difference of OS was observed (overall log rank  $p = 0.016$ , Figure 62B). As shown previously for the molecular classification system according to TCGA, patients with EBV(+) and MSI-H tumors showed the best survival compared to the reference group (EBV(+): HR, 0.60,  $p = 0.183$ ; MSI-H: HR, 0.69,  $p = 0.111$ ). Patients with MSI-L tumors showed the worst survival (HR, 1.46,  $p = 0.149$ ) which was especially observed in the non-CTx group (HR, 2.37,  $p = 0.007$ ). The high CIN-M group showed a significant worse OS compared to the remaining CIN-groups (HR, 1.30,  $p = 0.048$ ). This finding was especially evident in the CTx-group (HR, 1.45,  $p = 0.030$ , Table 36). All survival data including the 1, 3 and 5 years survival rates are summarized in Table 36.



**Figure 62: Modified molecular classification based on TCGA and survival of the patients**

Kaplan-Meier curves are shown for patients with tumor biopsies before neoadjuvant CTx (**A**) and for patients with resected tumors: all patients (**B**), patients treated without CTx (**C**), patients treated with neoadjuvant CTx (**D**). <sup>1</sup>*p*-values (overall) of log rank test; *p*-values < 0.05 in bold

**Table 36: Survival data of the different patient cohorts in association with the modified molecular classification system based on TCGA**

	Modified molecular classification system	No.	Events	Survival probability [%]			Median Survival [months] (95% CI)	HR (95% CI)	p-value <sup>1</sup>
				1 yr	3 yrs	5 yrs			
<b>Tumor biopsies before neoadjuvant CTx</b>	EBV(+)	5	3	100	60	40	56.5 (0-123.7)	0.76 (0.24-2.45)	0.646
	MSI-H	15	6	92.9	61.9	54.2	nr	0.61 (0.26-1.42)	0.250
	MSI-L	6	2	100	100	100	62.2 (60.6-63.9)	0.43 (0.11-1.79)	0.248
	High CIN-M	45	23	73.3	52.0	47.0	38.0	0.79 (0.48-1.31)	0.364
	Remaining CIN-groups	71	43	75.3	52.0	42.8	37.9 (22.2-53.6)	1 ref.	-
	Total	142	77	78.3	55.1	47.3	48.1 (26.2-70)	-	-
<b>All resected tumors</b>	EBV(+)	23	7	75.4	68.5	60.9	nr	0.60 (0.28-1.28)	0.183
	MSI-H	59	21	76.6	66.9	64.4	nr	0.69 (0.44-1.09)	0.111
	MSI-L	27	16	71.6	44.4	38.1	29.4 (15.9-42.9)	1.46 (0.87-2.44)	0.149
	High CIN-M	151	87	74.6	45.1	39.0	29.0 (15.6-42.4)	1.30 (1-1.69)	<b>0.048</b>
	Remaining CIN-groups	355	161	81.4	53.7	45.6	44.9 (23.8-66.0)	1 ref.	-
	Total	615	292	78.6	53.1	46.0	44.6 (29.8-59.4)	-	-
<b>Resected tumors without neoadjuvant CTx</b>	EBV(+)	7	1	83.3	83.3	83.3	nr	0.28 (0.04-2.02)	0.205
	MSI-H	35	13	72.9	66.2	66.2	nr	0.86 (0.47-1.57)	0.626
	MSI-L	14	12	62.3	31.2	23.4	21.7 (1.0-42.4)	2.37 (1.27-4.40)	<b>0.007</b>
	High CIN-M	75	35	75.5	59.0	50.4	61.1 (33.2-89.1)	1.22 (0.80-1.84)	0.355
	Remaining CIN-groups	159	61	84.6	61.3	54.1	89.5 (na)	1 ref.	-
	Total	290	122	79.7	60.4	53.8	85.0 (52.1-117.9)	-	-
<b>Resected tumors after neoadjuvant CTx</b>	EBV(+)	16	6	71.8	59.8	47.9	44.0 (na)	0.72 (0.32-1.64)	0.431
	MSI-H	24	8	81.9	67.5	60	nr	0.60 (0.29-1.23)	0.159
	MSI-L	13	4	82.5	70.7	70.7	62.4 (na)	0.72 (0.26-1.95)	0.515
	High CIN-M	76	52	73.6	29.9	28.0	22.4 (16.1-28.8)	1.45 (1.04-2.03)	<b>0.030</b>
	Remaining CIN-groups	196	100	78.9	47.6	38.8	33.8 (22.5-45.1)	1 ref.	-
	Total	325	170	77.6	46.4	39.0	32.2 (24.1-40.3)	-	-

<sup>1</sup>p-values of Cox's regression; p-values < 0.05 in bold

Multivariate analysis was performed separately for the resected non-CTx and CTx group. Analyzing the modified molecular classification system and the pre-therapeutically available factors revealed that cT ( $p < 0.001$ ), age ( $p = 0.001$ ) and the molecular subgroups ( $p = 0.052$ ) were independent prognostic factors in the non-CTx patient group (Table 37). Including the post-therapeutically available factors, only the clinical parameters pN ( $p < 0.001$ ), age ( $p = 0.004$ ), R-category ( $p = 0.020$ ) and tumor localization ( $p = 0.026$ ) were identified as independent prognostic factors in this patient group (Table 37). By contrast, analysis of the subgroup of only completely resected patients (R0) revealed only pN ( $p < 0.001$ ) and age ( $p = 0.006$ ) as independent prognostic factors (Table 37).

In the CTx group, only tumor localization ( $p < 0.001$ ) was identified as independent prognostic factor when analyzing the pre-therapeutically available factors. Including the post-therapeutically available factors revealed the clinical parameters R-category ( $p < 0.001$ ), (y)pN ( $p < 0.001$ ), tumor localization ( $p = 0.002$ ) and M-category ( $p = 0.007$ ) as independent prognostic factors (Table 38).

**Table 37: Multivariable analysis of survival including pre- and post-therapeutically available clinical factors and modified molecular classification system based on TCGA in the resected non-CTx cohort**

	HR	95% CI	<i>p</i> -value <sup>1</sup>
<b>Pre-therapeutic factors<sup>2</sup></b>			
<b>Clinical tumor stage</b>			
cT2	1 ref.	-	
cT3/4	2.73	1.84-4.06	<b>&lt;0.001</b>
<b>Age</b>	1.03	1.01-1.05	<b>0.001</b>
<b>Modified molecular classification according to TCGA</b>			
Remaining CIN-groups	1 ref.	-	-
EBV(+)	0.21	0.03-1.53	0.123
MSI-H	0.57	0.31-1.06	0.074
MSI-L	1.59	0.84-3.01	0.155
High CIN-M	1.10	0.72-1.68	0.648
<b>Post-therapeutic factors<sup>3</sup></b>			
<b>pN<sup>4</sup></b>			
pN0	1 ref.	-	
pN1	3.15	1.95-5.10	<b>&lt;0.001</b>
<b>Age</b>	1.03	1.01-1.04	<b>0.004</b>
<b>Resection status</b>			
R0	1 ref.	-	
R1	1.68	1.08-2.60	<b>0.020</b>
<b>Localization</b>			
Proximal	1 ref.	-	-
Middle	0.67	0.42-1.05	0.079
Distal	0.52	0.33-0.83	0.006
Total	1.09	0.51-2.31	0.830
<b>Post-therapeutic factors (R0 resected, non-CTx cohort)</b>			
<b>pN<sup>4</sup></b>			
pN0	1 ref.	-	
pN1	3.03	1.84-5.0	<b>&lt;0.001</b>
<b>Age</b>	1.03	1.01-1.05	<b>0.006</b>

ref, reference; <sup>1</sup>*p*-values of forward likelihood ratio Cox's regression model; *p*-values < 0.05 in bold; <sup>2</sup>Pre-therapeutic factors included: age, gender, localization, Laurén subtypes, cT, modified molecular classification based on TCGA. <sup>3</sup>Post-therapeutic factors included: age, sex, localization, Laurén subtypes, pT, pN, M-status, R-category, modified molecular classification according to TCGA. <sup>4</sup>Classification according to 7<sup>th</sup> Edition UICC

**Table 38: Multivariable analysis of survival including pre- and post-therapeutically available clinical factors and modified molecular classification system based on TCGA in the resected CTx cohort**

	HR	95% CI	<i>p</i> -value <sup>1</sup>
<b>Pre-therapeutic factors<sup>2</sup></b>			
<b>Localization</b>			<b>&lt;0.001</b>
Proximal	1 ref.	-	-
Middle	0.82	0.55-1.23	0.340
Distal	0.36	0.36-1.05	0.075
Total	1.66	1.66-5.57	<b>&lt;0.000</b>
<b>Post-therapeutic factors<sup>3</sup></b>			
<b>Resection status</b>			
R0	1 ref.	-	<b>&lt;0.001</b>
R1	1.96	1.40-2.76	
<b>(y)pN<sup>4</sup></b>			
(y)pN0	1 ref.		<b>&lt;0.001</b>
(y)pN1	2.74	1.78-4.20	
<b>Localization</b>			<b>0.002</b>
Proximal	1 ref.	-	-
Middle	0.75	0.50-1.14	0.178
Distal	0.66	0.38-1.14	0.135
Total	2.63	1.42-4.86	<b>0.002</b>
<b>Metastasis status</b>			
M0	1 ref.	-	<b>0.007</b>
M1	1.67	1.15-2.43	

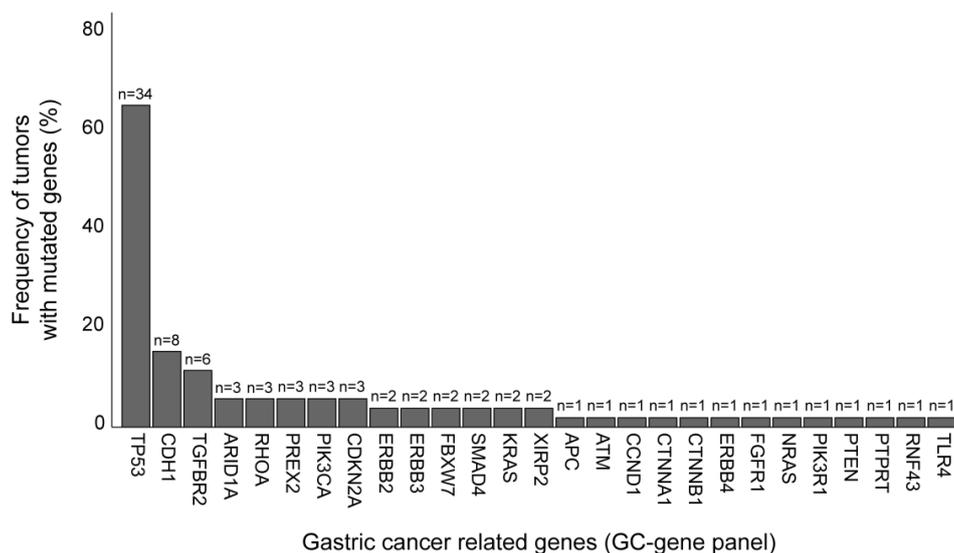
ref, reference; <sup>1</sup>*p*-values of forward likelihood ratio Cox's regression model; *p*-values < 0.05 in bold; <sup>2</sup>Pre-therapeutic factors included: age, gender, localization, Laurén subtypes, cT, modified molecular classification based on TCGA; <sup>3</sup>Post-therapeutic factors included: age, sex, localization, Laurén subtypes, (y)pT, (y)pN, M-status, R-category, modified molecular classification according to TCGA. <sup>4</sup>Classification according to 7<sup>th</sup> Edition UICC

## 3.6. Mutation profiling by next-generation sequencing

### 3.6.1. Frequency of tumors with alterations in GC related genes

Targeted sequencing of 52 GCs was performed using the gastric cancer related gene panel consisting of four primer pools yielding 525 amplicons of 58 genes. Sequence variants were identified at least in one of the genes in 45 (86.5%) of the 52 analyzed GCs. In total, 92 mutations were detected in nearly half of the analyzed genes (27 out of 58 genes). Of the 92 identified mutations, 62 (67.4%) were missense mutations, nine (9.8%) were nonframeshift deletions or substitutions and 21 (22.8%) were truncating mutations.

Figure 63 shows that the most frequent sequence variants among the analyzed tumors were detected in *TP53* (65.4%) followed by *CDH1* (15.4%) and *TGFBR2* (11.5%).



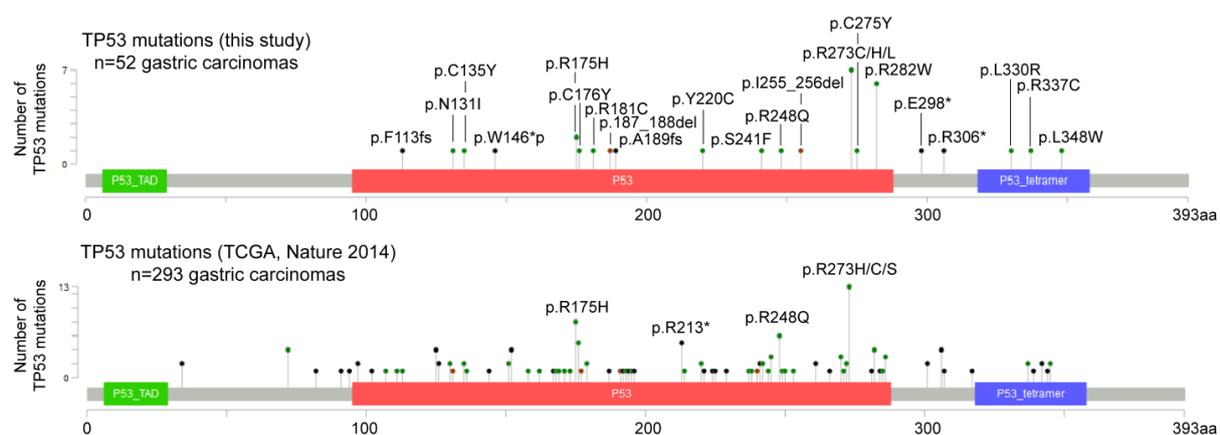
**Figure 63: Frequency of gastric carcinomas with mutations in selected gastric cancer related genes**

Recurrent hotspot mutations in the ERBB family of proteins were found at the amino acid positions V777 (exon 20) and S310 (exon 8) in *ERBB2* and G284 (exon 7) in *ERBB3*. For the RAS family of proteins, missense substitutions in the amino acid position G12 were found for *KRAS* (p.G12S/V) and *NRAS* (p.G12C). One of the two main hotspots for activating mutations was found in the helical domain of *PIK3CA* (p.E542K and p.E545K/A in exon 10). A complete summary of all identified mutations is given in Supplementary Table 25.

### 3.6.2. TP53 mutations

In total, 35 TP53 mutations were identified in 34 tumors whereas one tumor revealed two different TP53 mutations. Of the 35 TP53 mutations, 26 (74.3%) were missense mutations, five (14.3%) were truncating mutations, two (5.7%) were in-frame deletions and two (5.7%) were splice variants. The most frequent TP53 mutations in this study were p.R273C/H/L (n = 7) and p.R282W (n = 6). An overview about the TP53 mutations identified in this study cohort compared to those of the TCGA study is given in Figure 64.

The occurrence of TP53 mutations and Laurén classification revealed no significant correlation ( $p = 0.451$ ) but a significant difference was observed with the tumor localization ( $p = 0.008$ ). 17 (89.5%) of the 19 proximal located tumors showed TP53 mutations compared to 17 (53%) of 32 non-proximal located tumors.



**Figure 64: Identified TP53 mutations in gastric carcinomas**

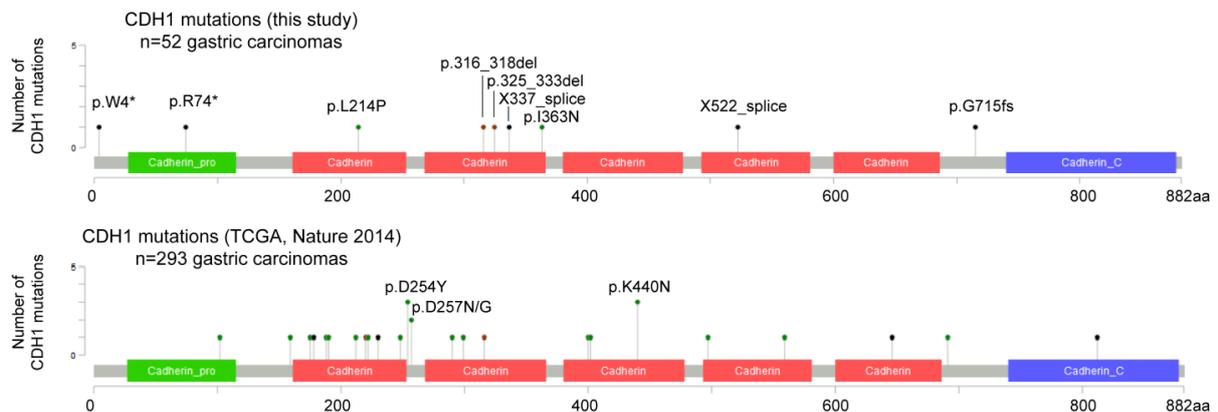
Missense mutations were shown in green, truncating mutations in black and in-frame mutations in brown.

[Designed with Mutation Mapper, cBioPortal: Cerami et al. 2012 and Gao et al. 2014]

### 3.6.3. *CDH1* and *RHOA* mutations

In total, nine *CDH1* mutations were found in seven tumors whereas two tumors showed two different *CDH1* mutations respectively. Of the nine *CDH1* mutations, two (22.2%) were missense mutations, three (33.4%) were truncating mutations, two (22.2%) were in-frame deletions and two (22.2%) were splice variants. All identified *CDH1* mutations in this study occurred only once. An overview about the *CDH1* mutations identified in this study cohort compared to those of the TCGA study is given in Figure 65.

*CDH1* mutations and Laurén classification revealed a significant correlation ( $p = 0.005$ ). Six (86%) of the seven tumors with *CDH1* mutations were classified as non-intestinal subtype. No significant correlation was observed regarding the tumor localization ( $p = 0.236$ ). Two (67%) of the three identified *RHOA* mutations were also associated with the non-intestinal subtype.



**Figure 65: Identified *CDH1* mutations in gastric carcinomas**

Missense mutations were shown in green, truncating mutations in black and in-frame mutations in brown.

[Designed with Mutation Mapper, cBioPortal: Cerami et al. 2012 and Gao et al. 2014]

### 3.6.4. Alterations in GC related genes and association with the MSI-L subtype

The correlation of the occurrence of alterations in gastric cancer related genes with the MSI-L and MSS subtype is shown in Figure 66. All *CDHI* mutations (n = 9) were identified within the MSS subgroup ( $p = 0.084$ ). Regarding the identified *TP53* mutations, it is striking that 10 (71.4%) of the 14 MSI-L tumors showed *TP53* mutations compared to 24 (63.2%) of the 38 MSS tumors. However, the occurrence of *TP53* mutations showed no significant association with the MSI-status ( $p = 0.578$ ).

The average mutation rate per subgroup was calculated by dividing the total number of identified mutations by the total number of tumors per subgroup and revealed a mutation rate of 1.9 for the MSI-L subgroup compared to 1.7 for the MSS subgroup.

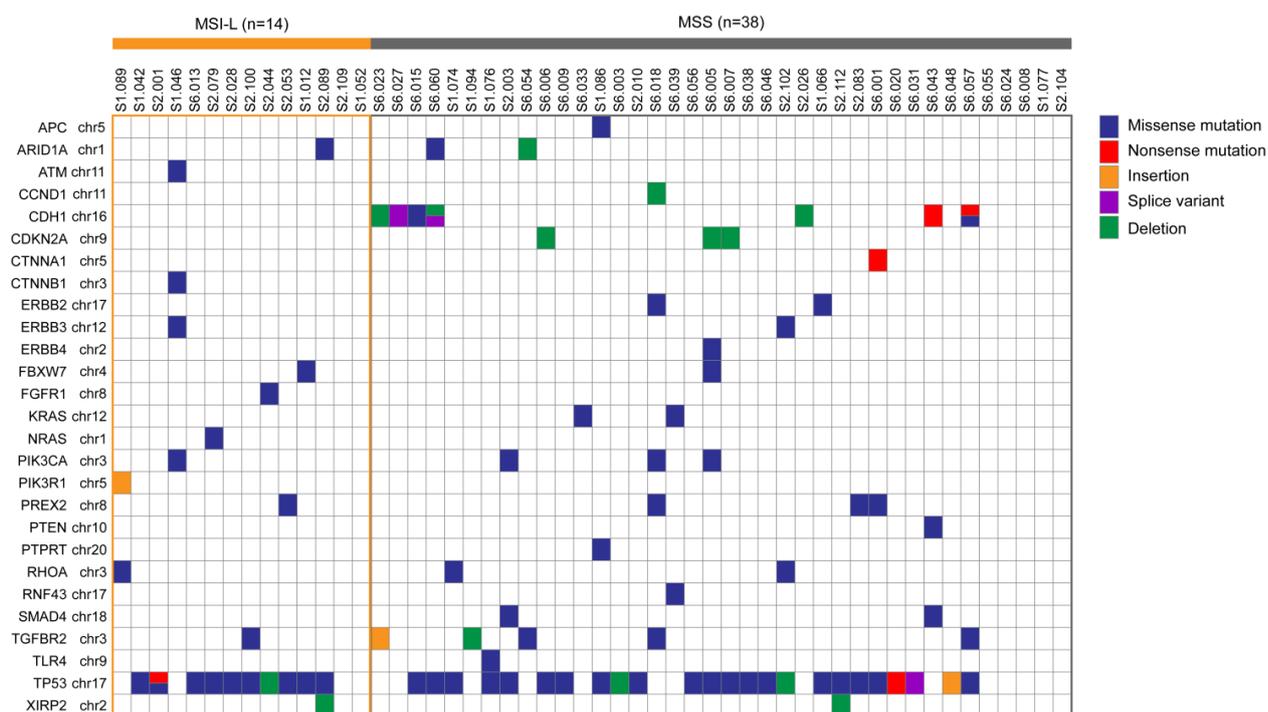


Figure 66: Mutation profile of gastric carcinomas and association with the MSI-L subtype

### 3.6.5. Alterations in GC related genes and association with the CIN-status

The mutation profiles of the 52 tumors stratified according to their CIN classification in four subgroups are shown in Figure 67. It is noticeable that five (71%) of the seven tumors with a *CDH1* mutation were classified as CIN-L compared to 6 (13%) of 45 tumors without a *CDH1* mutation (Figure 67). Thus, the occurrence of *CDH1* mutations was significantly associated with the CIN subtypes ( $p = 0.008$ ). No significant association was observed for *TP53* mutations and the CIN-status. However, *TP53* mutations tended to be more common in tumors with high CIN-M compared to tumors with the remaining CIN-groups (Figure 68). Furthermore, an increase of the frequency of *TP53* mutations was observed from the CIN-L subgroup (28%) to low CIN-M (37.5%) and to high CIN-M (40%) and CIN-H (47%) as shown in Figure 68. The mutation rates of the respective CIN-subgroups were as follows: 1.9 for CIN-L, 1.6 for low CIN-M, 2.0 for high CIN-M and 1.5 for CIN-H.

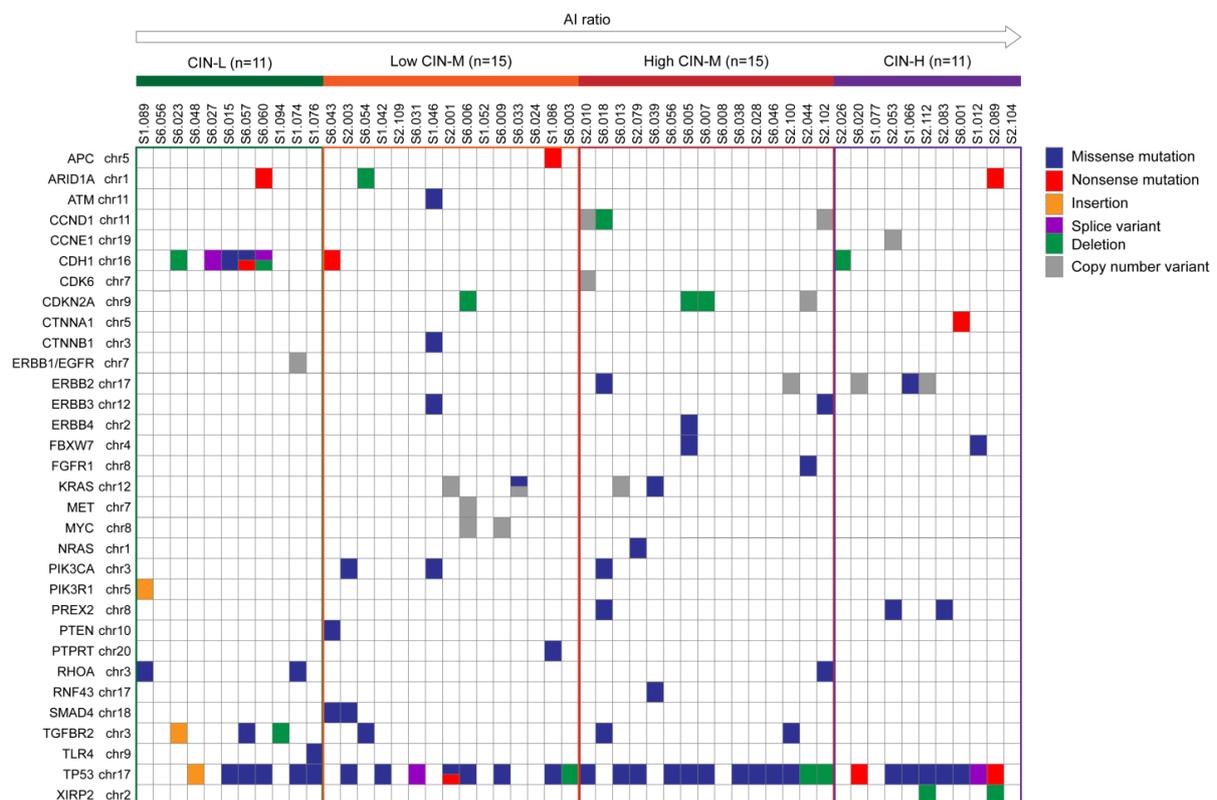
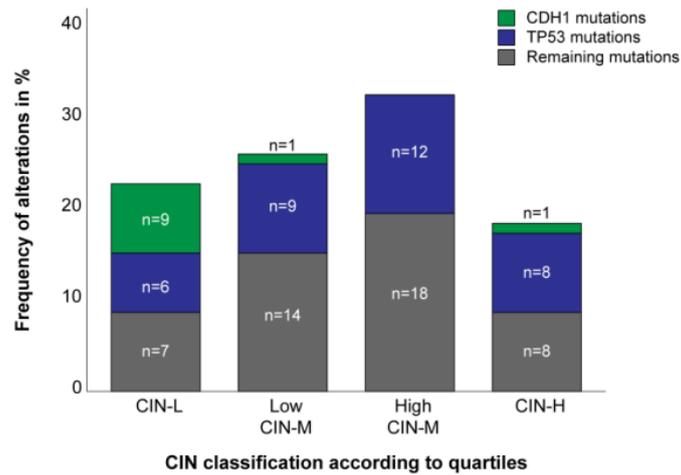
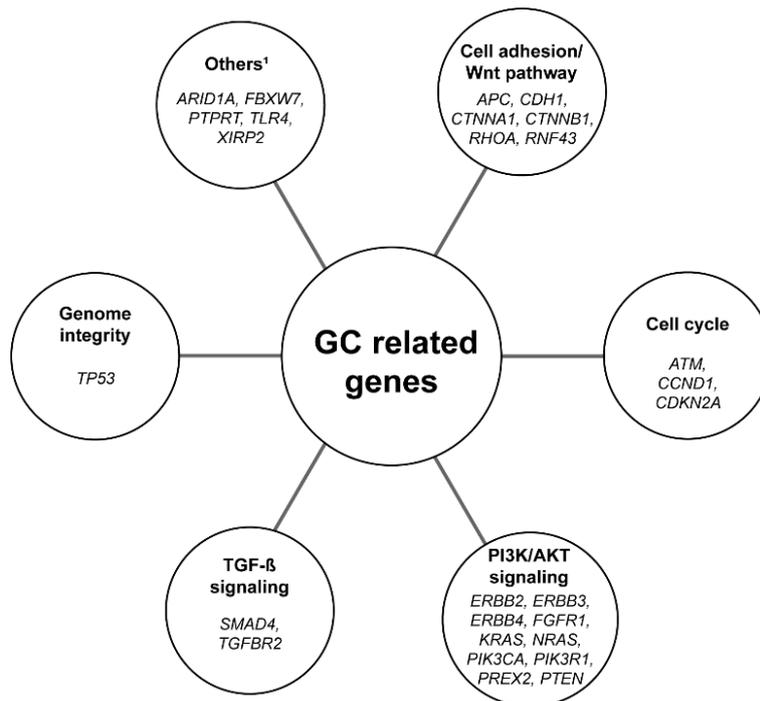


Figure 67: Mutation profile of gastric carcinomas and association with the CIN-status



**Figure 68: Occurrence of TP53 and CDH1 mutations in the four CIN-subgroups**

For a more specific analysis, the altered genes (n = 27) were clustered according to their involvement in function in cellular processes during carcinogenesis into the following groups as shown in Figure 69.

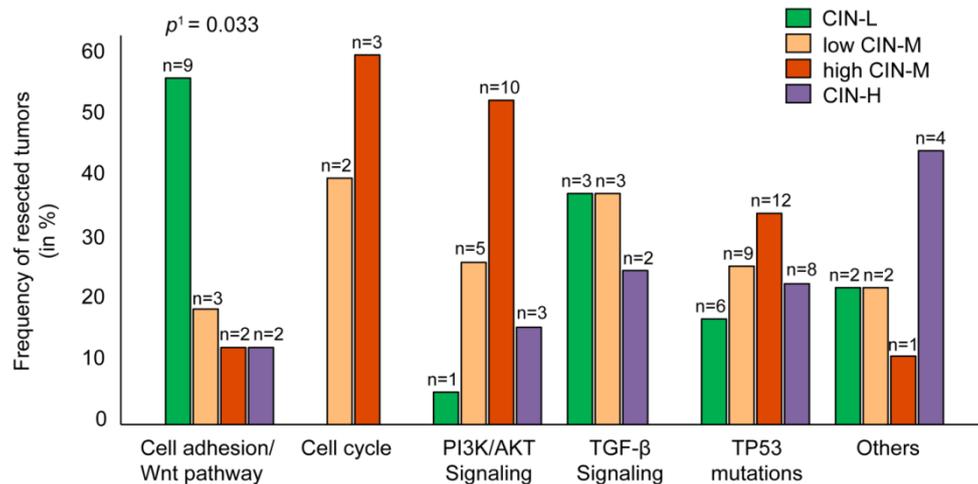


**Figure 69: Clustering of GC related genes according to their function in cellular processes during carcinogenesis**

<sup>1</sup>Among others, the genes are involved in cell growth and differentiation, chromatin remodeling, immune response and Notch signaling.

Overall, the clustering of the gastric cancer related genes was significantly associated with the four CIN-subgroups ( $p = 0.033$ , Figure 70). Mutations in genes involved in cell adhesion processes and in the Wnt pathway were more frequently identified in tumors classified as CIN-L (Figure 70).

Besides, the occurrence of mutations in genes involved in the PI3 Kinase/AKT signaling pathway (53%) and in cell cycle processes (60%) were associated with the high CIN-M subgroup compared to the remaining CIN-groups (Figure 70).



**Figure 70: Clustering of mutated gastric cancer related genes and association with CIN**  
<sup>1</sup> $p$ -value (overall) of Fisher's exact test

### 3.6.6. Prediction of copy number variants

Additionally, copy number variants (CNVs) were predicted in the 52 analyzed primary resected tumors using the summary of the coverage data of each tumor sample and amplicons generated by the Torrent Suite Software. Recurrent amplifications were identified in *CCNE1*, *CDK6*, *CCND1*, *KRAS*, *MYC*, *MET*, *ERBB1* and *ERBB2* in 13 tumors and one deletion in *CDKN2A* in one tumor. CNVs in genes involved in cell cycle processes such as *CDKN2A*, *CCNE1*, *CDK6* and *CCND1* were primarily identified in tumors classified as high CIN-M and *MET* and *MYC* amplifications seem to occur more frequently in the low CIN-M subgroup (Figure 67).

Nevertheless, no significant association was found between the four molecular CIN-groups based on quartiles of the AI ratios and the occurrence of CNVs, which could be attributable to overall small sample sizes.

### 3.7. Main characteristics of the molecular subgroups

In the following, the main characteristics of the molecular subgroups according to TCGA compared to the modified molecular subgroups are summarized (Figure 71).

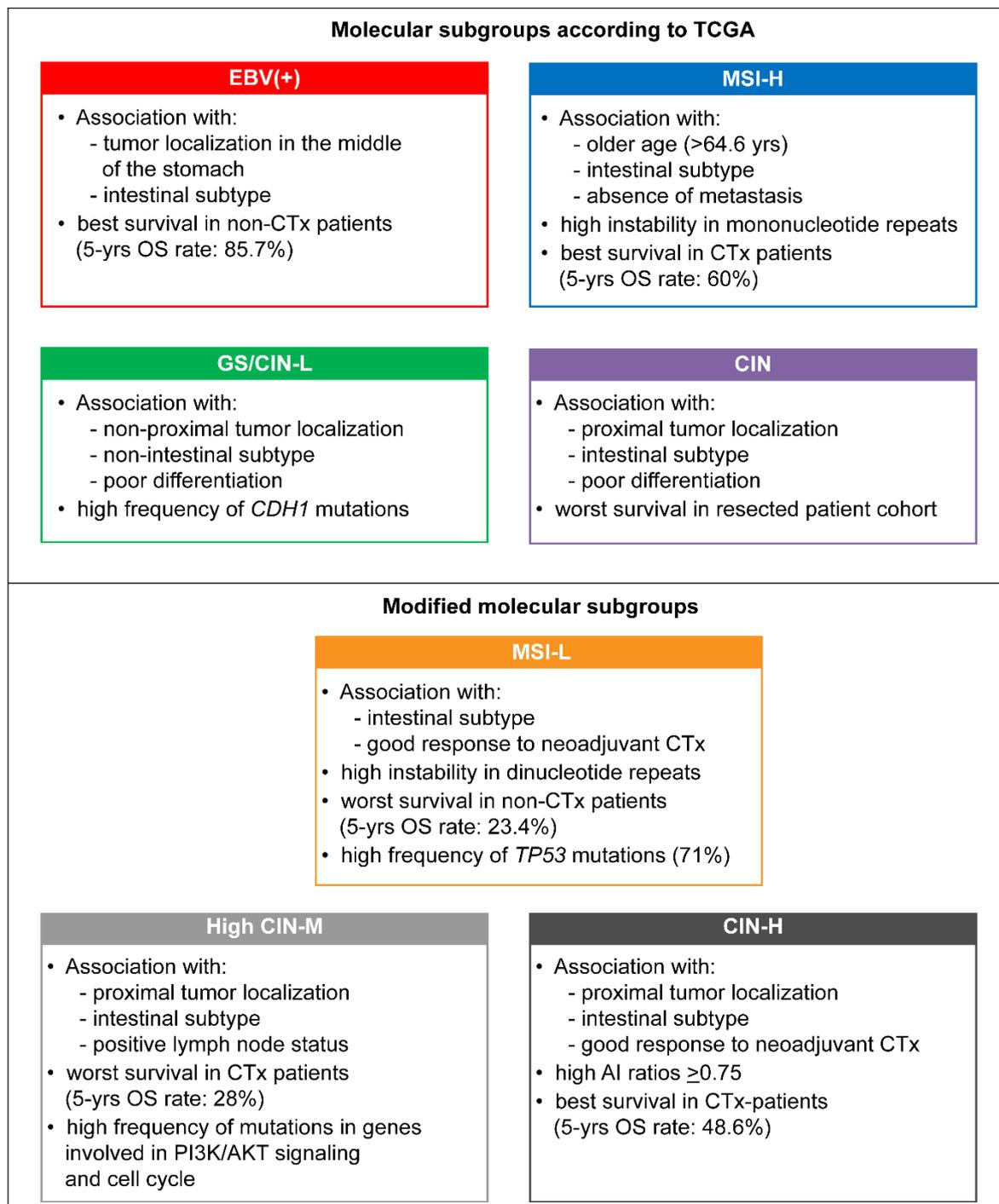


Figure 71: Main characteristics of the molecular subgroups

## 4. DISCUSSION

### 4.1. Identification of novel molecular subgroups in gastric cancer

Distinct molecular subtypes have been recently identified in GCs but knowledge about their clinical relevance in particular in the context of preoperative CTx is still limited (TCGA 2014, Cristescu et al. 2015, Sohn et al. 2017, Haag et al. 2019, Huang et al. 2019). On that account, the prognostic and predictive significance was clarified for the four classical molecular TCGA subgroups EBV(+), MSI-H, GS and CIN and in addition for a more detailed modified molecular classification system encompassing three modified subgroups namely MSI-L, high CIN-M and remaining CIN-groups apart from the EBV(+) and MSI-H subtype. Therefore, comprehensive analyses were performed in large GC cohorts including tumor biopsies of patients before platinum/5-FU based preoperative CTx and resected tumors of patients treated without or with preoperative CTx.

One of the most important achievements of the present study was the identification of novel molecular subgroups which demonstrated potentially predictive and prognostic impact in GC patients. Besides, these molecular subtypes showed distinct clinical and genomic signatures. One of these molecular subtypes represented MSI-L which was significantly associated with better response to neoadjuvant CTx in patients with biopsies before preoperative CTx. These patients also showed a tendency for increased survival even if the difference was not statistically significant, probably due to low sample size. Interestingly, MSI-L demonstrated a negative prognostic effect in the resected patient cohort treated with surgery alone, which could indicate that the MSI-L subtype has a differential prognostic role depending on the treatment of the patients. As the significant association of MSI-L with worse prognosis was also demonstrated in patients with clinically staged cT2 tumors, the determination of the MSI-L subgroup may contribute improvements for future managements of this particular patient subgroup (Kohlruss et al. 2019). A similar association of MSI-L with worse prognosis was shown previously for colorectal cancer (Lee et al. 2015).

The MSI-L phenotype has been described in various tumor entities including GCs between ranges of 4%-20% (Hatch et al. 2005, Napieralski et al. 2007, An et al. 2012, Lee et al. 2015, Moehler et al. 2015). In comparison, in this study a rather lower prevalence of 4% was detected for patients having tumors with MSI-L. It has to be mentioned that the detection rate of MSI-L strongly depends on the type and number of microsatellite markers used in a respective study. As shown in this study, it is striking that instability in MSI-L tumors occurred predominantly at one of the dinucleotide repeats which is also in line with previous findings (Hatch et al. 2005, Lee et al. 2015, Kim et al. 2017). In addition, as shown here the types of instability in the MSI-L phenotype which were restricted to the dinucleotide markers were mostly insertions with two base pairs whereas the instabilities detected at the two mononucleotide markers were found to be deletions even though instability was only found for two cases. In contrast, the most frequent instabilities in MSI-H tumors were detected in the two mononucleotide repeats BAT25 and BAT26 which were also included in the marker panel for MSI analysis.

These findings suggested that the occurrence of MSI-L may be related to a distinct mechanism in particular for the characteristic instability at the dinucleotide repeats. Contrarily to the well-known molecular background of MSI-H, the origin and biological significance of the MSI-L phenotype is still poorly understood and controversially discussed (Tomlinson et al. 2002, Hile et al. 2013). Tumors with MSI-H are usually associated with defects in one of the MMR proteins (MLH1, MSH2, MSH6 and PMS2) and demonstrated high genomic instability. By contrast, tumors which exhibit only moderate genomic instability may undergo different molecular processes compared to the MSI-H phenotype and showed instability only at selected microsatellites (Hile et al. 2013). Those tumors may display alternative forms of tumor-associated instabilities at di- or tetranucleotide microsatellite repeats which are known as the MSI-L phenotype or as elevated microsatellite alterations at selected tetranucleotides (EMAST). Furthermore, as suggested in the literature the phenomenon MSI-L may be related to moderate defects in specific DNA repair genes, to spontaneous mutations at microsatellites associated with elevated mutation rates or induction by DNA damaging agents (Hile et al. 2013, Koi et al. 2018). Some studies hypothesized that the MSI-L and EMAST phenotype are results of a defect machinery or aberrant protein expression of MSH3 in colorectal cancer (Haugen et al. 2008, Plaschke et al. 2012, Carethers 2017, Koi et al. 2018). Based on the results of the present study it is also tempting to speculate that MSI-L may be associated with a particular type of dysfunctional DNA repair mechanism and numerous proteins involved in these repair complexes may represent possible candidates in that scenario. To support this hypothesis and to further clarify the potential role of MSI-L as distinct subgroup, additional studies of the GC cohorts included in this study analyzing MSH3 expression patterns by immunohistochemistry of MSI-L tumors compared to MSS tumors may be helpful. However, MSI-L tumors are usually not distinguished from MSS tumors and in contrast to the known clinical relevance of the MSI-H phenotype; the biological significance of MSI-L remains unclear or is not yet defined (Lee et al. 2015, Nazemalhosseini Mojarad et al. 2016). Some studies analyzing colorectal cancer reported obvious differences in the clinical relevance of MSS and MSI-L tumors whereas others did not observe any differences (Kohonen-Corish et al. 2005, Wright et al. 2005, Kim et al. 2009, Azzoni et al. 2011). The negative prognostic role of the MSI-L subtype in patients treated without CTx and the good prognosis to preoperative CTx in the pre-therapeutic tumor biopsy cohort led to assume that MSI-L in the context in this study should be treated as own distinct subgroup.

The other novel identified subgroup namely high CIN-M demonstrated a prognostic significance in the resected tumor cohort. High CIN-M which represents the third quartile of the AI ratios was associated with worse survival in the resected patient cohort though this negative prognostic effect was predominantly demonstrated in patients treated with preoperative CTx. These findings led to speculate that altered CTx treatment strategies are rather necessary to improve outcome of this particular patient group having high CIN-M tumors. Regarding all resected tumors, it is striking that the CIN-H subtype with elevated AI ratios  $\geq 0.75$  demonstrated more or less survival rates similar to the two remaining lower CIN quartiles.

Interestingly, the most obvious differences, although not statistically significant, regarding survival and the two CIN-subtypes high CIN-M and CIN-H were observed in the resected patient cohort after CTx whereas the CIN-H tumors showed improved outcome in this patient group. In addition, CIN-H demonstrated a positive predictive effect of patients with tumor biopsies before preoperative CTx. The association between CIN and prognosis of the patients was analyzed in various tumor entities and controversial results were observed even though different techniques were used for the assessment of CIN (Carter et al. 2006, Walther et al. 2008, Birkbak et al. 2011, Jamal-Hanjani et al. 2015, Tijhuis et al. 2019). In a comprehensive study, a gene expression signature of CIN was determined for multiple human cancers and the authors observed that patients with tumors having high CIN signatures demonstrated worse survival compared to those having a lower CIN signature (Carter et al. 2006). In addition, increased CIN has been shown to be linked to aggressive tumor growth and progression in various cancer types (Gao et al. 2016). However, some studies also presumed that only an optimal level of CIN can lead to tumor progression and proposed a “just-right” model (Bakhoun and Compton 2012, Gronroos and Lopez-Garcia 2018). Moreover, a paradoxical non-monotonic relationship of CIN and prognosis was reported in GCs and breast carcinomas (Birkbak et al. 2011, Jamal-Hanjani et al. 2015). In addition, two different CIN-exploiting treatment strategies currently exist whereas CIN-reducing approaches aimed to suppress the rate of chromosomal alterations and to reduce the tumor growth (Bakhoun and Compton 2012, McClelland 2017, Thompson et al. 2017). Contrarily, the CIN-inducing treatment strategies generate extreme levels of CIN to induce tumor cell death (Thompson et al. 2017).

As the CIN-H group in the present study was associated with good prognosis in patients treated with preoperative CTx it is tempting to speculate in the context of the findings in the literature that cancer cells harboring elevated CIN were predominantly eliminated by particular CTx agents due to fitness loss or cell autonomous lethality. In comparison, an intermediate level of CIN seems to stabilize the cancer cell biological fitness which is related to an overall worse outcome of the patients. Support for the approach of this study was given in a retrospective and prospective validating study using the CIN70 expression signature as a surrogate measurement of CIN (Birkbak et al. 2011, Jamal-Hanjani et al. 2015). These authors demonstrated that the third CIN quartile was associated with the poorest prognosis whereas the extreme CIN-group showed improved survival of the patients. As shown here and in previous studies, the CIN-status and prognosis do not stand in a simple monotone relationship which may indicate that a more detailed classification of CIN in four subgroups is necessary to identify and depict prognostic relevant subgroups (Birkbak et al. 2011, Jamal-Hanjani et al. 2015). Moreover, the previous research and the findings of this study support that the intermediate CIN or high CIN-M phenotype as proposed here represent distinct biological molecular subgroups in GCs. Furthermore, prospective analysis in eligible *in vitro* gastric cancer cell lines or organoid models are mandatory to clarify the relation of the CIN-status with tumor progression or chemosensitivity regarding different CTx agents or checkpoint inhibitors (Thompson et al. 2017, Sansregret et al. 2018).

In addition, in this study high CIN-M tumors were associated with proximal tumor localization and intestinal subtype. As this subtype was newly proposed, it remains difficult to compare with previous reports. Nonetheless, the CIN-subgroup according to TCGA was predominantly found in the gastroesophageal junction and cardia of the stomach and was also associated with intestinal histology (TCGA 2014, TCGA 2017). In this study, the high CIN-M phenotype showed high frequencies throughout the patient cohorts in the range between 25%-32%. Furthermore, the most interesting finding regarding the genetic background of this subgroup was the high prevalence of *TP53* alterations and mutations in genes involved in the RTK/AKT signaling cascade. These findings were supported by other studies where activating mutations of components of the RTK/Ras pathway were mainly detected in highly chromosomal unstable tumors (Deng et al. 2012, TCGA 2014).

As the detection of MSI-L and high CIN-M is based on relatively simple and cost-effective multiplex PCR assays, those two subtypes could represent attractive markers for routine diagnostic analysis. In addition, the prognostic relevance of the modified molecular classification system was underlined as it emerged as an independent prognostic factor in the multivariable analysis of resected patients treated without preoperative CTx when considering the pre-therapeutically available factors.

Regarding the original molecular TCGA subgroups MSI-H and EBV(+) in this study, no predictive impact was found for both subtypes. However, patients with EBV(+) and MSI-H tumors showed the best prognosis in the resected patient cohort. The MSI-H subgroup was found to be associated with good prognosis for GC patients as reported in the majority of the studies (Cristescu et al. 2015, Marrelli et al. 2016, Pereira et al. 2018, Polom et al. 2018, Huang et al. 2019). This is essentially in line with the findings for the resected patient cohort in this analysis. However, different results in the context of CTx have been reported in various studies (An et al. 2012, Choi et al. 2014, Kim et al. 2015, Smyth et al. 2017, Haag et al. 2019). In the study of Smyth et al. (2017), a negative prognostic effect of MSI was found for patients who received preoperative CTx in the context of the MAGIC trial. These results did not support the findings of the present study, since no negative prognostic relevance of MSI-H was found in patients with tumor biopsies before CTx and additionally a good prognostic effect of MSI-H was demonstrated in all resected patients independently of the respective treatment. Furthermore, the significant associations of the MSI-H subtype found with patient age, tumor localization and status of metastasis confirm results reported by others (TCGA 2014, Pereira et al. 2018, Polom et al. 2018, Choi et al. 2019). The prevalence of MSI-H among all analyzed tumors in this study was found in a range between 10%-11% which is also similar to recent reports (Polom et al. 2018, Choi et al. 2019). In contrast, the TCGA study revealed higher frequencies of MSI tumors (22%) which may be related to a different composition of the patient cohorts especially regarding the tumor localization as in this study a relatively high number of proximal located tumors are included in the analysis (TCGA 2014, Kohlruss et al. 2019). As mentioned before, MSI was more frequently found in the corpus or lower parts of the stomach (TCGA 2014, Polom et al. 2018).

The prevalence of EBV(+) of all analyzed tumors in this study was in a range between 3.5%-4% which is essentially in line with some previous reports (Genitsch et al. 2015, Hewitt et al. 2018). However, most of the studies revealed higher frequencies of EBV-associated gastric carcinomas in the range between 5%-10% which may also be related to the composition of the patient cohorts in context of tumor localization (Lee et al. 2009, Murphy et al. 2009, Nishikawa et al. 2014, TCGA 2014). EBV(+) was associated with tumors localized in the middle of the stomach as well as intestinal histology and male sex which is similar to recent studies (TCGA 2014, Hewitt et al. 2018). Furthermore, EBV(+) was not associated with response to neoadjuvant CTx in this study, though a good prognosis was observed in EBV-associated tumors of patients treated without CTx as recorded previously (Camargo et al. 2014). In addition, in a clinical trial evaluating an immune check point inhibitor in metastatic GC, a good response was observed in patients with EBV-associated tumors as well as MSI-H tumors (Kim et al. 2018).

Due to the findings of the study that prognostic and predictive implications and evidences in molecular genetic patterns were primarily detected in the modified subgroups, a more detailed classification system beside the TCGA classification as presented here might be very useful. Of course, further extensive studies are necessary to confirm these findings and to clarify the molecular background of the MSI-L and high CIN-M phenotype in the pathogenesis of GC. Although the identification of these molecular subgroups have broaden the knowledge about the complexity of gastric cancer, the main challenge will be to effectively translate these molecular findings into improvements in classical therapy strategies or into better future management of GC patients to improve outcome and prediction.

## **4.2. Determination of AI and CIN**

CIN is known to be a major hallmark for the development of various cancers including gastric and esophageal adenocarcinomas and was described as one of the four molecular subgroups by extensive whole-genome analysis of the TCGA consortium. CIN was identified as the most common subtype having high frequencies of copy number alterations due to recurrent deletions and amplifications of specific chromosomal loci (TCGA 2014).

In most of the large studies, CIN is determined using SNP array technologies which indeed provide genome-wide information about chromosomal alterations but are not suited for routine diagnostics as they are rather cost-intensive and are often only applicable on DNA from fresh frozen tumor material (TCGA 2014, TCGA 2017, Liu et al. 2018). No adequate and cost-efficient diagnostic tool was available for the reliable determination of CIN in gastric cancer until then. This issue was addressed in a previous analysis as well as in this study and a microsatellite based multiplex PCR assay was successfully established for the reliable detection of AI as surrogate for CIN performable on DNA isolated from FFPE tumorous tissues (Kohlruss et al. 2018).

In several previous reports, microsatellite analysis was used to detect LOH and AI in various cancer types but a standardized cut-off value for the definition of AI is not defined and threshold values were used in the range below 0.5-0.6 or above 1.5-2.0 (Beckmann et al. 1996, Ott et al. 2003, Erill et al. 2005, Frigerio et al. 2007, Brýs et al. 2013). The multiplex PCR assays used in this study allow a more sensitive detection of AI due to experimentally determined marker-specific cut-off values (Kohlruss et al. 2018). To my knowledge, only some other studies performed microsatellite analysis using individual cut-off values in bladder carcinoma (Frigerio et al. 2007, van Tilborg et al. 2012). Although CIN is a common phenomenon in GCs, the definition of CIN is not clearly defined in the literature and has to be critically considered.

The comparison of the detected copy number gains, losses and calls of AI determined by the genome-wide OncoScan analysis with the results for AI determined by the multiplex PCR assays regarding single microsatellite loci revealed an overall concordance of 84%. Most of the discrepancies between the two methods were related to the occurrence of balanced gains detected in the OncoScan assay which are not detectable by microsatellite analysis or to AI-values very close to the cut-off borderlines of the respective microsatellite marker. It also has to be emphasized that the microsatellite analysis indicates AI and does not allow a clear distinction between copy number gains or losses at a respective chromosomal region. This however, is not essentially required for the purpose of this study to determine a more global CIN classification based on AI as surrogate marker for CIN. To analyze the genome wide extent of chromosomal alterations, a chromosomal arm was considered to be altered if at least 80% of one arm was lost, gained or demonstrated AI according to the CIN-definition of TCGA (2017). The comparison of the results regarding the AI ratios determined by the multiplex PCR assays with the ratio of genome chromosomal alterations revealed a high correlation. These findings indicate, that the AI ratios determined by the multiplex PCR assays can reliably reflect the extent of chromosomal alterations occurring on a genome-wide level. In addition, the sensitivity of the microsatellite based assay was addressed and revealed that the test can be reliably used if the tumor cell content was at least 10%. Furthermore, a rather stable performance of the CIN classification based on this assay in relation to tumor heterogeneity was demonstrated.

Analogous to the molecular classification system of the TCGA consortium, EBV(+) and MSI-H represented distinct molecular subgroups which were strictly delineated from CIN and GS tumors (TCGA 2014). Therefore, the EBV- and MSI-status was assessed in a preselection step and as a reliable determination of AI is only possible in microsatellite stable (MSS) tumors, only MSS/EBV(-) and MSI-L tumors were included in the analysis of AI and CIN using the microsatellite based multiplex PCR assay. In this study, a CIN classification in two subgroups based on genome-wide analysis and definition of the TCGA study was proposed as well as a modified four-group classification system.

According to the definition of TCGA, 80% of the tumor biopsies before CTx were classified as CIN and 77% of the resected tumors without or after neoadjuvant CTx. This reflects considerably higher frequencies especially in comparison with the findings of the TCGA study which classified 50% of the analyzed tumors as CIN which can be related to the application of different methods to detect CIN (TCGA 2014). As the number of proximal located tumors was relatively high, it may also be a reason for the high frequencies of CIN-associated tumors in this study. Nevertheless, CIN was the most frequently found subtype of the tumors in both studies. Furthermore, the CIN phenotype was associated with proximal tumor localization and with intestinal tumor type, which is similar to previous reports (Powell et al. 2005, TCGA 2014, TCGA 2017, Liu et al. 2018). The genomic stable subtype was more frequently found in tumors with higher tumor grades (G3/4), which most likely reflects the association of GS with non-intestinal type tumors which usually are poorly differentiated neoplasms. Regarding the clinical relevance, this two-subgroup based CIN-classification revealed no significant associations neither with response to neoadjuvant CTx nor with prognosis. Thus suggests that a classification system in only two CIN-groups does not adequately reflect the biological heterogeneity of GCs and a more differentiated graduation of CIN in four subgroups is necessary.

In the resected tumor cohort obvious differences were observed in survival regarding the four modified CIN-subgroups whereas the high CIN-M and CIN-H phenotypes were of outstanding clinical relevance as discussed earlier. Considering the potential prognostic value of the proposed modified CIN-classification, it may be of significant impact to implement this system in clinical trials regarding different treatment strategies. To my knowledge, only in a few number of other studies CIN was categorized in more than two subgroups as for example Watanabe et al. (2012) classified colorectal carcinomas as CIN-H or CIN-L when the LOH ratio (AI ratio) was  $\geq 33\%$  or  $< 33\%$  analyzing seven microsatellite markers of five chromosomal regions. The authors further subclassified CIN-H tumors as a mild or severe type when the LOH ratio was  $< 75\%$  or  $\geq 75\%$ . In another study, a four-part CIN classification was demonstrated in breast and ovarian cancer as well as in GCs using the CIN70 signature as surrogate measure of CIN (Birkbak et al. 2011). These findings underline the heterogeneity of the CIN phenotype and as shown here, probably different subgroups of CIN are more appropriate to reflect the heterogeneity of GC.

In addition, tumor cell plasticity in the context of platinum/5-FU based CTx was determined analyzing paired tumor biopsies before CTx and resected tumors after preoperative CTx. Therefore, pre- and post-therapeutic alterations in the four modified CIN-subgroups were compared. The analysis revealed that 50% of the tumors changed their CIN-status from a higher group to a lower one from the pre-therapeutic biopsy to the corresponding post-therapeutic resected tumor. This may indicate, that tumor cells harboring higher AI ratios were preferentially eliminated by preoperative CTx. In contrast, in the minority of the corresponding tumors a change was observed from a lower CIN-group to a higher one which may be associated with an induction of AI by CTx agents in tumor cells in this setting.

Also the differences in the CIN-status of corresponding pre- and post-therapeutic tumors could be due to different tumor cell contents. To deepen the knowledge of these findings, comparative analysis using a validation cohort with pre-therapeutic tumor biopsies without CTx and corresponding resected tumors may be helpful.

Furthermore, the results of AI analysis were evaluated regarding the single microsatellite regions and the most notable findings are emphasized as follows. Considering the frequencies of altered chromosomal regions in this study revealed that AI at 9p21, 12p12 and 17p13 were among the most frequent alterations in the tumor biopsy and resected tumor cohort. Essentially in line with these findings, various studies reported loss of heterozygosity at 9p21 and 17p13 in GCs in a broad range from 11%-57% and 33%-71% (Ott et al. 2003, Powell et al. 2005, Jiao et al. 2006, Fan et al. 2012, Choi et al. 2018). In the present study, frequencies of AI at 9p21 were found in a range between 79%-80% and at 17p13 between 57%-66% which are similar to the ranges found previously. The overall differences may be related to the usage of different methods to detect AI or to specific characteristics of the analyzed patient cohorts. As already described in the publication of the multiplex microsatellite assays, comparing the AI ratios determined by the microsatellite based PCR assays with the ratios of chromosomal alterations detected by the OncoScan assay at a single chromosomal region revealed that generally the multiplex PCR assay showed higher ratios of AI (Kohlruss et al. 2018). This higher sensitivity may be related to our individual definition of specific cut-off values for each marker. In addition, the fact that we mainly used markers of chromosomal regions which are specifically altered in GC may explain the higher AI ratios.

Regarding the results of AI in the resected tumor cohort stratified according to the treatment approach demonstrated that tumors treated without preoperative CTx showed significantly higher frequencies of AI at 18q21. This chromosomal region is known to be linked to a number of cancer-related genes and tumor suppressors (Candusso et al. 2002). Evaluating the clinical parameters in association with AI at single microsatellite loci in this study revealed that generally AI was preferentially associated with proximal tumor localization and the intestinal histological subtype. Some studies reported a similar association of AI or copy number gains at selected chromosomal regions with the Laurén histology whereas others did not find obvious differences (Jin et al. 2015, Choi et al. 2018).

It also has to be emphasized that the chromosomal regions 8q24 and 17q21 were associated with better response to preoperative CTx in the tumor biopsy cohort which may reflect potentially specific regions of genes involved in chemosensitivity toward platinum/5FU containing therapeutic regimens. A potential candidate located at 8q24 may be the oncogene *MYC* which for example was found to be associated with response to Docetaxel-based neoadjuvant CTx in breast cancer (Pereira et al. 2017).

The frequency of the resected tumors having AI at a respective microsatellite region in relation to the four modified CIN-groups revealed five chromosomal loci at which the increase of the frequencies from low to high CIN-M was 2- to 3-fold higher in comparison to those from high CIN-M to CIN-H.

These regions were covering potentially targetable genes such *CDK6* at 7q21, *MET* at 7q31 or *MYC* at 8q24. However, further fine mapping of these chromosomal regions are necessary to identify the most relevant genes of the respective chromosomal regions.

### 4.3. Mutational landscape of gastric cancer

Through extensive molecular profiling of large GC cohorts by whole exome or targeted sequencing, recurrent and novel potential cancer-driving genes were recently discovered in several studies (TCGA 2014, Wang et al. 2014, Xu et al. 2014, Chen et al. 2015, Cai et al. 2019). The importance of classic driver mutations in the pathogenesis of stomach cancer were strengthened and new driver genes which were involved in chromatin remodeling, Wnt signaling, cell motility and RTK signaling were identified (Katona and Rustgi 2017).

In the present study, mutation profiling was performed in a subset of 52 GCs to additionally obtain genetic information about the identified molecular subgroups. Therefore, a GC related gene panel consisting of 58 genes was used for targeted sequencing of the tumors and somatic mutations were found in 86.5% of the analyzed patients. The most frequent mutations among the analyzed tumors were detected in *TP53* (65%) and *CDHI* (15%) which are known to be classical cancer-driving genes in stomach cancer (Bellini et al. 2012, TCGA 2014, Cristescu et al. 2015, Pan et al. 2018, Cai et al. 2019). A high prevalence of *TP53* mutations is recorded in many human cancers and detected in the range between 33%-59% for GCs in several large studies (TCGA 2014, Cristescu et al. 2015, Cai et al. 2019). *TP53* is a key regulator of different cellular processes and ensures the genomic integrity. In this study, the most frequent alterations of *TP53* were missense substitutions occurring at the DNA-binding domain at the hotspot positions p.R273 and p.R282 which is essentially in line with the findings of the TCGA study (TCGA 2014). Mutations in *TP53* were significantly associated with proximal tumor localization which was also reported by others (Cai et al. 2019).

In this study, alterations in *TP53* were detected throughout the four modified CIN-subgroups but interestingly the prevalence of the *TP53* mutations steadily increased from the lowest CIN-subgroup to the highest one. In the context of the TCGA studies, *TP53* mutations were also enriched in the CIN subtype (TCGA 2014, TCGA 2017). However, a direct comparison of these findings concerning the TCGA CIN-subgroup and the four CIN-subgroups defined in this study is not really reliable due to the application of different methods to determine and define CIN.

*CDHI* is also a classically mutated driver gene which is usually associated with a diffuse histological pattern of GC which is in line with the findings of this study (Becker et al. 1994, TCGA 2014, Cristescu et al. 2015). A significant association was observed regarding the four modified CIN-groups, which was mainly related to the occurrence of *CDHI* mutations in the CIN-L subtype. According to the TCGA data, *CDHI* mutations were found in 37% of the genomically stable tumors which support the findings of this study (TCGA 2014).

Although the high prevalence of *TP53* and *CDHI* mutations was analyzed and confirmed in large GC cohorts in several studies, to date no altered treatment strategies for patients with tumors harboring these gene alterations exist (Katona and Rustgi 2017).

Beside the detection of these two well-known cancer-driving genes of GC, genomic alterations were also detected in the gene *RHOA* encoding a small GTPase which is involved in cell adhesion and cytoskeleton remodeling. *RHOA* mutations were identified in 6% of the analyzed tumors and were associated with the diffuse histological subtype although the sample size of tumors harboring this alteration is very low. Similarly, the association of *RHOA* mutations with diffuse type tumors was previously confirmed by various studies (Kakiuchi et al. 2014, TCGA 2014, Wang et al. 2014, Röcken et al. 2016). The prevalence of *RHOA* mutations reported in the literature ranged from 4%-25% and strongly depends on the composition of the tumor cohort in the context of the Laurén subtypes (Kakiuchi et al. 2014, TCGA 2014, Wang et al. 2014, Röcken et al. 2016). Such as, Kakiuchi et al. (2014) found high frequencies of *RHOA* mutations (25%) due to a greater number of diffuse typed tumors. In addition, *RHOA* mutations were more frequently found in genomically stable tumors as shown in this study which is in line with the findings of the TCGA consortium (TCGA 2014). The alterations of *RHOA* were all detected at position p.Tyr42Cys and are located in the effector-binding region of the gene which can lead to dysregulation of downstream RHO signaling. This region was also reported previously as hotspot region in several studies (TCGA 2014, Wang et al. 2014, Röcken et al. 2016). The role of *RHOA* alterations in the pathogenesis of GC still remains unclear but it seems to play a crucial role affecting cell motility in the development of diffuse type cancers (TCGA 2014, Katona and Rustgi 2017).

Moreover, in various sequencing studies of GC recurrent mutations were detected in chromatin remodeling genes whereby *ARID1A* is most frequently mutated in a range between 14%-27% (Wang et al. 2011, Zang et al. 2012, TCGA 2014). Indeed, the prevalence of *ARID1A* mutations in this study was a bit lower but still revealed the fourth most frequent mutated gene among all analyzed tumors. Inactivating mutations of *ARID1A* were mainly prevalent in EBV-associated tumors and hypermutated tumors with MSI (Wang et al. 2011, TCGA 2014). As the attention of this study was to obtain genetic information about the different CIN-subgroups, only MSS or MSI-L tumors were sequenced and therefore no direct comparison of the findings of these studies was possible. Apart from these findings, also known hotspot mutations with rather low frequencies between 2%-6% were found in the genes *PIK3CA*, *ERBB2*, *ERBB3*, *NRAS* and *KRAS*. Beside those well-known hotspot mutations, a number of several novel or not described sequence variants were detected in this study which should be validated by Sanger sequencing.

As mentioned before, the high CIN-M subgroup was associated with a higher prevalence of alterations in *TP53*, higher frequency of mutations found in genes involved in the PI3K/AKT signaling pathway such as *PIK3CA*, *PIK3RI* and ERBB family of genes as well as genes involved in cell cycle processes such as *CCND1* and *CDKN2A* which is similarly to the findings in the TCGA study for the CIN subtype (TCGA 2014).

As the Ion Torrent technology used in this study still is very expensive for large-scale sequencing of GC samples in the diagnostic setting, only a subset of the tumors was analyzed within the scope of this doctoral thesis. The new era of next-generation sequencing is the application of comprehensive gene panels such as the *TruSight* sequencing panel by Illumina targeting a wide spectrum of cancer-related genes. The benefits of this approach are that it is performable in large translational studies and enables a broader analysis of the genetic landscape of GCs compared to the panel based sequencing used in this study (Fisher et al. 2016, Ow et al. 2019).

To summarize, many large NGS studies provided a genome-wide knowledge about the complexity of the genetic landscape of GCs and had led to the development of novel genomically based classification schemes for this cancer type. However, to date the major challenge is to transfer these information into approaches for predictive treatment options for targeted therapy and to improve the outcome of GC patients (Katona and Rustgi 2017).

#### **4.4. Limitations of the study and perspectives**

Despite the comprehensive analysis of very large patient cohorts comprising 143 tumor biopsies before CTx and 616 resected tumors without or after neoadjuvant CTx, the present study has also some limitations which are mainly related to the retrospective nature. A main limiting factor was the requirement of DNA from non-tumorous and tumorous tissue from a respective patient as the multiplex PCR assays were based on the comparison of microsatellite patterns of corresponding normal and tumor DNA. Similarly, a main limiting factor for inclusion of patients with tumor biopsies before CTx in the analysis was the availability of DNA or sufficient tumor material. The further limitations were previously described in the respective publication and were emphasized as follows (Kohlruss et al. 2019). As the analysis in this study was not performed in the context of a randomized clinical trial testing different CTx regimens, the analysis has to be considered as an exploratory study. In addition, the samples used in this study originate from a non-homogenous patient collective from daily clinical practice of two local medical centers that naturally exhibited variations in surgical approaches and different CTx protocols.

Indeed, further comprehensive studies are needed to confirm the results found in this study especially concerning the two phenotypes MSI-L and high CIN-M to clarify and deepen the knowledge about their functional role in gastric carcinogenesis. More importantly, randomized studies are required to verify the prognostic and predictive impact of the modified molecular subtypes based on the TCGA classification as proposed in this study. In additional studies, the clinical relevance of these molecular subgroups should also be clarified in the context of the latest recommended treatment strategies in advanced GC patients such as the FLOT regimen.

## 5. CONCLUSION

In conclusion, novel molecular subgroups of a modified classification system according to TCGA were identified with predictive and prognostic impact in the context of preoperative CTx by comprehensive molecular characterization of large GC cohorts including adenocarcinomas of the esophageal junction. As the EBV- and MSI-status can easily and reliably be determined by standardized techniques, no generally diagnostic method was available for the detection of CIN until then. In this study, the reliable determination and classification of CIN was confirmed by the application of microsatellite based multiplex PCR assays which represent a simple and cost-efficient diagnostic tool for the analysis of large GC cohorts suitable in routine diagnostic settings. The original molecular TCGA subtypes EBV(+) and MSI-H were not associated with better response to preoperative CTx in patients with tumor biopsies before CTx but indicated good prognosis in resected GC patients treated with or without CTx. In contrast, the novel identified subgroups MSI-L and CIN-H were predictive of good response to CTx. GC patients having tumors with MSI-L had the worst survival in patients treated with surgery alone which indicates that the MSI-L phenotype might help to delineate patients with high benefit from preoperative platinum/5-FU based CTx treatment. High CIN-M was also identified as molecular subgroup with a distinct genetic pattern showing high prevalence of *TP53* mutations and alterations in genes involved in PI3K/AKT signaling. Besides, the high CIN-M subtype was associated with worse prognosis in resected patients treated with perioperative CTx. Finally, in this study a successful characterization of a more detailed classification system encompassing five molecular subgroups with prognostic relevance in patients with resected tumors without or after neoadjuvant CTx was demonstrated.

Due to the findings of the study that prognostic and predictive implications and evidences in molecular genetic patterns were primarily detected in the modified subgroups of GC patients, a more detailed classification system beside the TCGA classification is necessary to understand the molecular complexity and heterogeneity underlying the pathogenesis of GC. Indeed, additional studies are mandatory to confirm the results found in this study to further characterize the functional background of the two subtypes MSI-L and high CIN-M in gastric carcinogenesis. Moreover, prospective comprehensive studies are important to clarify the clinical relevance of the modified molecular subgroups in the context of alternative treatment strategies such as for example the currently recommended FLOT regimen in advanced GC patients or targeted therapy approaches using immune checkpoint inhibitors to improve the outcome of GC patients. Nevertheless, the main challenge still remains to translate the molecular data generated from the identified molecular subgroups into better treatment management of GC patients to significantly improve prediction and survival.

## 6. APPENDIX

**Supplementary Table 1: Chemotherapy regimens of the preoperatively treated patients**

	Tumor biopsies before CTx		Resected tumors after CTx		
	n	%	n	%	
<b>Total</b>	143	100	325	100	
<b>Preoperative chemotherapy regime</b>	Cis + 5-FU or Cap	117	82	124	38
	Ox + 5-FU or Cap	18	13	46	14
	Cis + 5-FU + Doc or Pac	2	1	27	8
	Ox + 5-FU + Doc	0	0	21	6
	Cis or Ox + 5-FU or Cap + Epi	5	4	84	26
	Others	1	<1	22	7
	n/a	0	0	1	<1

Abbreviations: Cis, cisplatin; Ox, oxaliplatin; 5-FU, 5-fluorouracil; Cap, capecitabine; Doc, docetaxel; Pac, paclitaxel; Epi, epirubicin; Others, combination of Cis/Ox with other agents or cross over between different treatment regimens; n/a, no data available

**Supplementary Table 2: Chromosomal regions with respective tumor-related genes included in the microsatellite based multiplex PCR assays (Kohlruess 2016)**

Chromosomal region	Gene
4q22	CCSER1
5q11	MAP3K1
6p25	FOXC1
7q21	CDK6
7q31	MET
8p23	GATA4
8q24	MYC
9p21	CDKN2A
12p12	KRAS
16q23	WWOX
17p13	TP53
17q12	ERBB2
18q21	SMAD4
19q12	CCNE1

**Supplementary Table 3: Settings for segmentation thresholds using the SNP-FASST2 algorithm implemented in the Nexus Express Software**

Settings for segmentation thresholds	
Significance threshold <sup>1</sup>	1.0e <sup>-5</sup>
High gain <sup>2</sup>	log2R > 0.6
Single copy gain <sup>1</sup>	log2R > 0.2
Big loss <sup>2</sup>	log2R < -1.0
Hemizygous loss <sup>1</sup>	log2R < -0.2
Homozygous frequency threshold <sup>2</sup>	0.85
Homozygous value threshold <sup>2</sup>	0.8
Heterozygous imbalance threshold <sup>2</sup>	0.4
Minimum LOH length (Kb) <sup>2</sup>	500

<sup>1</sup>According to Hardiman et al. (2016); <sup>2</sup>Default values defined from Nexus Express Software

**Supplementary Table 4: Published studies used for the selection of gastric cancer related genes for NGS panel design**

Published study	Number of patients	Sequencing method
TCGA (2014)	295	Whole exome
Cristescu et al. (2015)	300	Targeted
Wang et al. (2014)	100	Whole genome
Chen et al. (2015)	78	Whole exome
Kuboki et al. (2016)	121	Cancer panel (409 genes)
Dulak et al. (2013)	149 EAC	Whole exome
Li et al. (2016)	544	Targeted

EAC, Esophageal adenocarcinoma

**Supplementary Table 5: Survival data of patients with tumor biopsies before CTx stratified according to tumor regression grade (TRG) and response to CTx**

Response to neoadjuvant CTx	No.	Events	Survival probability [%]			Median survival [months] (95% CI)	<i>p</i> -value <sup>1</sup>
			1 yr	3 yrs	5 yrs		
TRG1	45	13	95.4	85.1	71.1	nr	<b>&lt;0.001</b>
TRG2	34	20	66.5	53.2	49.1	44.6 (0-94.1)	
TRG3	64	44	73.1	37.4	32.1	26.7 (12.9-40.5)	
Total	143	77	78.4	55.4	47.7	48.1 (26.2-70.1)	
Responder (TRG1)	45	13	95.4	85.1	71.1	nr	<b>&lt;0.001</b>
Non-responder (TRG2/3)	98	64	70.8	42.5	37.6	29.3 (21.2-37.4)	
Total	143	77	78.4	55.4	47.7	48.1 (26.2-70.1)	

<sup>1</sup>*p*-values (overall) of log rank test; *p*-values < 0.05 in bold

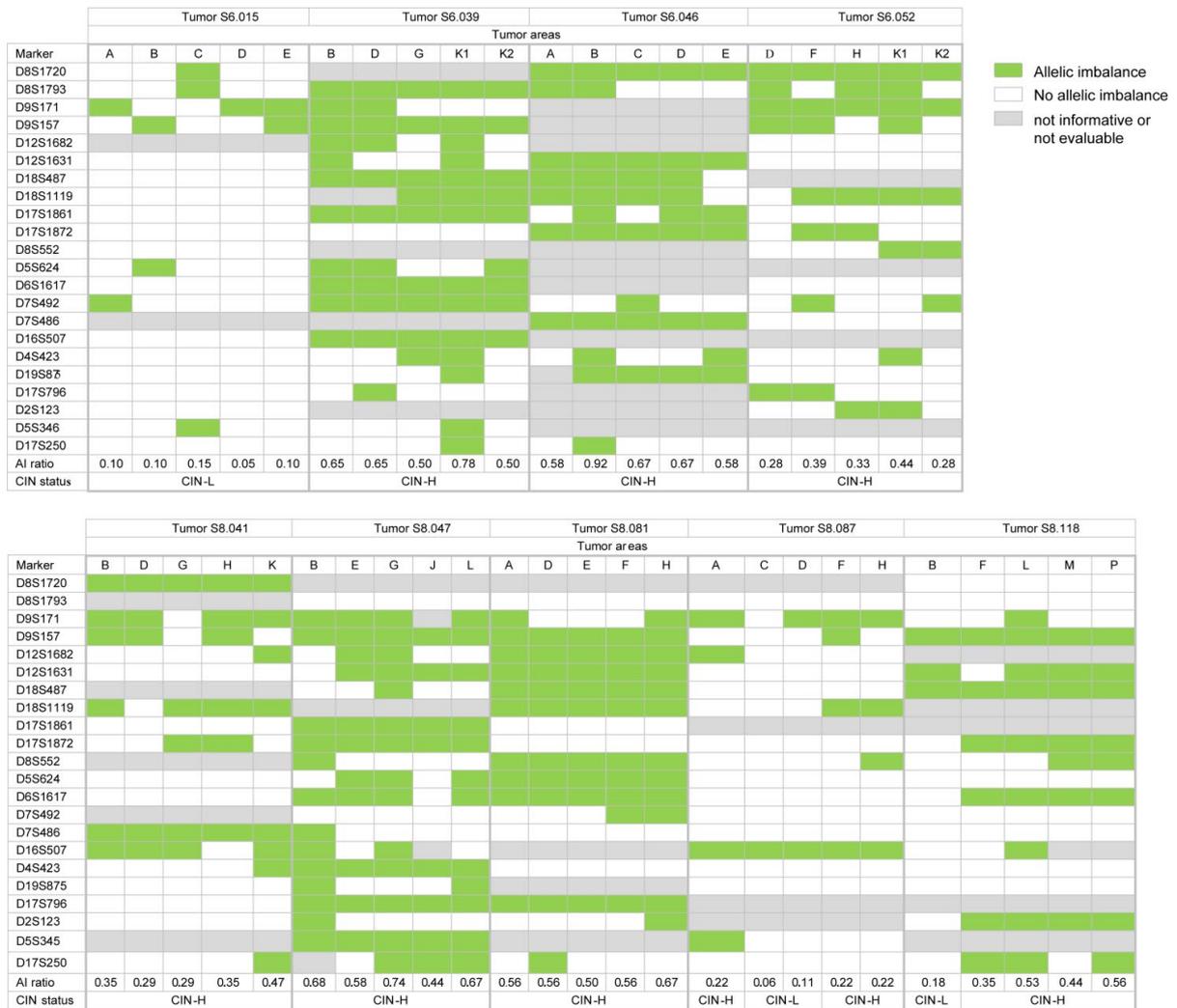
**Supplementary Table 6: Concordance of chromosomal alterations determined by the OncoScan assay compared to the results of AI detected with the multiplex PCR assays at a respective microsatellite locus**

Microsatellite marker	Chromosomal region	Number of concordant measurements	Number of informative events <sup>1</sup>	Concordance (%) <sup>2</sup>
D12S1631	12p12	27	28	96
D7S486	7q31	21	22	96
D8S552	8p23	18	20	95
D2S123	2p21	15	16	94
D9S171	9p21	19	21	90
D17S1872	17q12	25	28	89
D9S157	9p21	24	27	89
D6S1617	6p25	24	27	89
D5S346	5q21	24	27	89
D7S492	7q21	23	26	89
D17S796	17p13	20	23	87
D12S1682	12p12	20	23	87
D17S1861	17q12	25	29	86
D4S423	4q22	23	27	85
D18S487	18q21	16	19	84
D19S875	19q12	19	23	83
D5S624	5q11	20	25	80
D16S507	16q23	16	24	79
D8S1720	8q24	20	27	74
D18S1119	18q21	14	20	70
D8S1793	8q24	18	26	69
D17S250	17q21	11	21	52

<sup>1</sup>In total, 30 tumors were analyzed per microsatellite marker and only informative events generated by the multiplex PCR assays were counted. <sup>2</sup>Number of concordant measurements of both methods per informative events generated by the multiplex PCR assays multiplied by hundred.

**Supplementary Table 7: Ratios of chromosomal alterations and AI ratios of the 30 tumors analyzed with the OncoScan assay and the microsatellite based multiplex PCR assays**

Tumors	OncoScan assay		Microsatellite based multiplex PCR assays		
	Number of chromosomal arms with alteration	Ratios of chromosomal alterations	Number of markers with AI	Number of informative markers	AI ratios
S6.008	23	0.64	9	16	0.56
S6.001	22	0.61	14	18	0.78
S6.003	19	0.53	10	20	0.5
S6.046	19	0.53	8	12	0.67
S6.038	18	0.5	13	22	0.59
S6.039	18	0.5	10	18	0.56
S6.009	16	0.44	6	14	0.43
S6.005	14	0.39	11	17	0.65
S6.007	13	0.36	9	16	0.56
S6.017	12	0.33	12	18	0.67
S6.047	10	0.28	11	16	0.69
S6.054	9	0.25	7	19	0.37
S6.018	8	0.22	10	19	0.53
S6.052	8	0.22	7	19	0.37
S6.013	7	0.19	7	15	0.47
S6.056	7	0.19	9	18	0.5
S6.033	6	0.17	7	17	0.41
S6.006	5	0.14	6	17	0.35
S6.024	4	0.11	7	18	0.39
S6.060	4	0.11	4	17	0.24
S6.055	2	0.06	0	15	0
S6.015	1	0.03	2	20	0.1
S6.044	1	0.03	1	17	0.06
S6.011	0	0	1	15	0.07
S6.016	0	0	1	20	0.05
S6.023	0	0	1	21	0.05
S6.027	0	0	1	18	0.06
S6.034	0	0	1	20	0.05
S6.040	0	0	1	17	0.06
S6.061	0	0	2	18	0.11



**Supplementary Figure 1: Microsatellite based CIN classification and tumor heterogeneity.**

AI ratios and the resulting CIN-status of nine tumors each with five tumor areas were shown. The occurrence of AI at a respective microsatellite marker of each tumor area is colored in green, no AI is colored in white and not informative or not evaluable markers are colored in grey.

**Supplementary Table 8: Results of dilution series for determination of the detection limit of the microsatellite based multiplex PCR assays**

Sample S6.005			Sample S6.038		
Tumor cell content	AI ratio	CIN classification	Tumor cell content	AI ratio	CIN classification
<b>70%</b>	0.54	High	<b>60%</b>	0.47	High
60%	0.40	High	51%	0.55	High
56%	0.47	High	48%	0.41	High
50%	0.40	High	43%	0.41	High
46%	0.40	High	40%	0.35	High
42%	0.40	High	36%	0.47	High
35%	0.20	High	30%	0.35	High
28%	0.07	Low	24%	0.24	High
23%	0.07	Low	20%	0.12	Low
20%	0.07	Low	17%	0.18	Low
14%	0.00	Low	12%	0.12	Low
10%	0.00	Low	9%	0.06	Low
Sample S6.017			Sample S6.001		
Tumor cell content	AI ratio	CIN classification	Tumor cell content	AI ratio	CIN classification
<b>70%</b>	0.87	High	<b>90%</b>	0.87	High
60%	0.78	High	77%	0.87	High
56%	0.60	High	72%	0.67	High
50%	0.67	High	64%	0.80	High
46%	0.80	High	60%	0.73	High
42%	0.73	High	54%	0.87	High
35%	0.60	High	45%	0.67	High
28%	0.47	High	36%	0.33	High
23%	0.53	High	30%	0.33	High
20%	0.40	High	26%	0.27	High
14%	0.40	High	18%	0.40	High
10%	0.47	High	13%	0.33	High

Initial tumor cell contents determined by a pathologist are highlighted in bold.

**Supplementary Table 9: Frequency of AI for the non-CTx and CTx tumor cohort**

Chromosomal regions	Resected tumors without CTx (cT3/cT4, n=135)			Resected tumors after CTx (cT3/cT4, n=269)			p-value <sup>2</sup>
	Number of tumors with AI	Number of informative tumors	Frequency of AI (%) <sup>1</sup>	Number of tumors with AI	Number of informative tumors	Frequency of AI (%) <sup>1</sup>	
2p21	33	58	57	77	162	48	0.221
4q22	46	119	39	86	240	36	0.642
5q11	35	101	35	64	221	29	0.304
5q21	53	110	48	103	234	44	0.469
6p25	44	110	40	102	235	43	0.551
7q21	39	92	42	68	193	35	0.243
7q31	36	104	35	61	193	32	0.598
8p23	36	80	45	101	198	51	0.364
8q24	73	131	56	124	259	48	0.143
9p21	99	127	78	207	259	80	0.654
12p12	76	122	62	159	251	63	0.844
16q23	29	71	41	70	169	41	0.934
17p13	55	94	59	115	194	59	0.901
17q12	65	134	49	142	268	53	0.397
17q21	26	106	25	53	228	23	0.797
18q21	79	116	68	139	242	57	<b>0.053</b>
19q12	38	103	37	76	214	36	0.811

Only tumors with a clinical tumor stage cT3/cT4 were included in this analysis.

<sup>1</sup>Number of tumors with AI divided by number of informative tumors; <sup>2</sup>p-values of Chi-square test or Fisher's exact test; p-values < 0.1 in bold

**Supplementary Table 10: Frequency of AI for the resected tumors stratified according to tumor localization**

Chromosomal regions	Tumors with proximal tumor localization (n=268)			Tumors with non-proximal tumor localization (n=260)			p-value <sup>2</sup>
	Number of tumors with AI	Number of informative tumors	Frequency of AI (%) <sup>1</sup>	Number of tumors with AI	Number of informative tumors	Frequency of AI (%) <sup>1</sup>	
2p21	81	144	56	69	149	46	0.089
4q22	99	235	42	68	239	28	<b>0.002</b>
5q11	75	214	35	47	196	24	<b>0.014</b>
5q21	109	233	47	86	219	39	0.107
6p25	115	229	50	64	215	30	<b>&lt;0.001</b>
7q21	76	186	41	57	187	30	<b>0.036</b>
7q31	64	196	33	55	195	28	0.339
8p23	97	177	55	72	170	42	<b>0.020</b>
8q24	134	259	52	125	250	50	0.695
9p21	213	256	83	190	250	76	<b>0.044</b>
12p12	161	245	66	137	241	57	<b>0.045</b>
16q23	62	143	43	65	164	40	0.509
17p13	117	190	62	100	185	54	0.140
17q12	140	268	52	110	258	43	<b>0.027</b>
17q21	61	222	27	41	211	19	<b>0.049</b>
18q21	159	241	66	109	219	50	<b>&lt;0.001</b>
19q12	91	215	42	59	199	30	<b>0.007</b>

<sup>1</sup>Number of tumors with AI divided by number of informative tumors; <sup>2</sup>p-values of Chi-square test or Fisher's exact test; p-values < 0.05 in bold

**Supplementary Table 11: Frequency of AI for the resected tumors stratified according to Laurén histological subtypes**

Chromosomal regions	Tumors with intestinal subtype (n=292)			Tumors with non-intestinal subtype (n=240)			p-value <sup>2</sup>
	Number of tumors with AI	Number of informative tumors	Frequency of AI (%) <sup>1</sup>	Number of tumors with AI	Number of informative tumors	Frequency of AI (%) <sup>1</sup>	
2p21	86	151	57	66	145	46	<b>0.049</b>
4q22	109	264	41	60	213	28	<b>0.003</b>
5q11	88	229	38	34	184	18	<b>&lt;0.001</b>
5q21	127	251	51	70	204	34	<b>&lt;0.001</b>
6p25	109	239	46	72	208	35	<b>0.018</b>
7q21	86	204	42	48	173	28	<b>0.004</b>
7q31	77	213	36	43	182	24	<b>0.007</b>
8p23	97	185	52	73	165	44	0.126
8q24	155	284	55	105	229	46	<b>0.049</b>
9p21	239	286	84	168	224	75	<b>0.017</b>
12p12	182	265	69	118	225	52	<b>&lt;0.001</b>
16q23	79	164	48	50	146	34	<b>0.013</b>
17p13	141	211	67	76	167	46	<b>&lt;0.001</b>
17q12	158	290	54	94	240	39	<b>&lt;0.001</b>
17q21	66	245	27	37	192	19	0.061
18q21	176	254	69	93	209	44	<b>&lt;0.001</b>
19q12	100	235	43	50	181	28	<b>0.002</b>

<sup>1</sup>Number of tumors with AI divided by number of informative tumors; <sup>2</sup>p-values of Chi-square test or Fisher's exact test; p-values < 0.05 in bold

**Supplementary Table 12: Frequency of AI for tumor biopsies before CTx and association with response to preoperative CTx**

Chromosomal regions	Responder (n=38)			Non-responder (n=84)			p-value <sup>2</sup>
	Number of tumors with AI	Number of informative tumors	Frequency of AI (%) <sup>1</sup>	Number of tumors with AI	Number of informative tumors	Frequency of AI (%) <sup>1</sup>	
2p21	14	20	70	30	46	65	0.705
4q22	11	32	34	32	66	48	0.187
5q11	13	25	52	24	69	35	0.131
5q21	19	29	66	34	71	48	0.109
6p25	18	29	62	37	76	49	0.220
7q21	14	30	47	32	65	49	0.816
7q31	9	25	36	24	63	38	0.855
8p23	12	22	55	33	69	48	0.583
8q24	30	37	81	40	80	50	<b>0.001</b>
9p21	33	37	89	61	82	74	<b>0.067</b>
12p12	24	38	63	59	82	72	0.332
16q23	11	26	42	19	54	35	0.538
17p13	19	31	61	46	67	69	0.473
17q12	21	38	55	47	84	56	0.943
17q21	17	28	61	23	68	34	<b>0.015</b>
18q21	25	37	68	50	76	66	0.851
19q12	14	30	47	25	60	42	0.652

<sup>1</sup>Number of tumors with AI divided by number of informative tumors; <sup>2</sup>p-values of Chi-square test or Fisher's exact test; p-values < 0.1 in bold

**Supplementary Table 13: Frequency of AI for resected tumors after CTx and association with tumor regression**

Chromosomal regions	TRG2 (n=141)			TRG3 (n=143)			p-value <sup>2</sup>
	Number of tumors with AI	Number of informative tumors	Frequency of AI (%) <sup>1</sup>	Number of tumors with AI	Number of informative tumors	Frequency of AI (%) <sup>1</sup>	
2p21	42	87	48	40	85	47	0.873
4q22	38	120	32	54	135	40	0.167
5q11	31	115	27	34	116	29	0.691
5q21	53	124	43	54	124	44	0.898
6p25	50	123	41	55	124	44	0.556
7q21	42	108	39	32	96	33	0.410
7q31	30	96	31	35	108	32	0.859
8p23	58	104	56	49	104	47	0.212
8q24	61	135	45	70	139	50	0.391
9p21	102	136	75	115	138	83	<b>0.089</b>
12p12	90	136	66	80	129	62	0.480
16q23	38	85	45	36	90	40	0.529
17p13	55	101	54	64	102	63	0.231
17q12	69	140	49	76	143	53	0.516
17q21	28	120	23	26	121	21	0.731
18q21	76	131	58	69	123	56	0.758
19q12	38	109	35	45	119	38	0.643

<sup>1</sup>Number of tumors with AI divided by number of informative tumors; <sup>2</sup>p-values of Chi-square test or Fisher's exact test; p-values < 0.1 in bold

**Supplementary Table 14: Survival data of patients with tumor biopsies before CTx in association with AI-status**

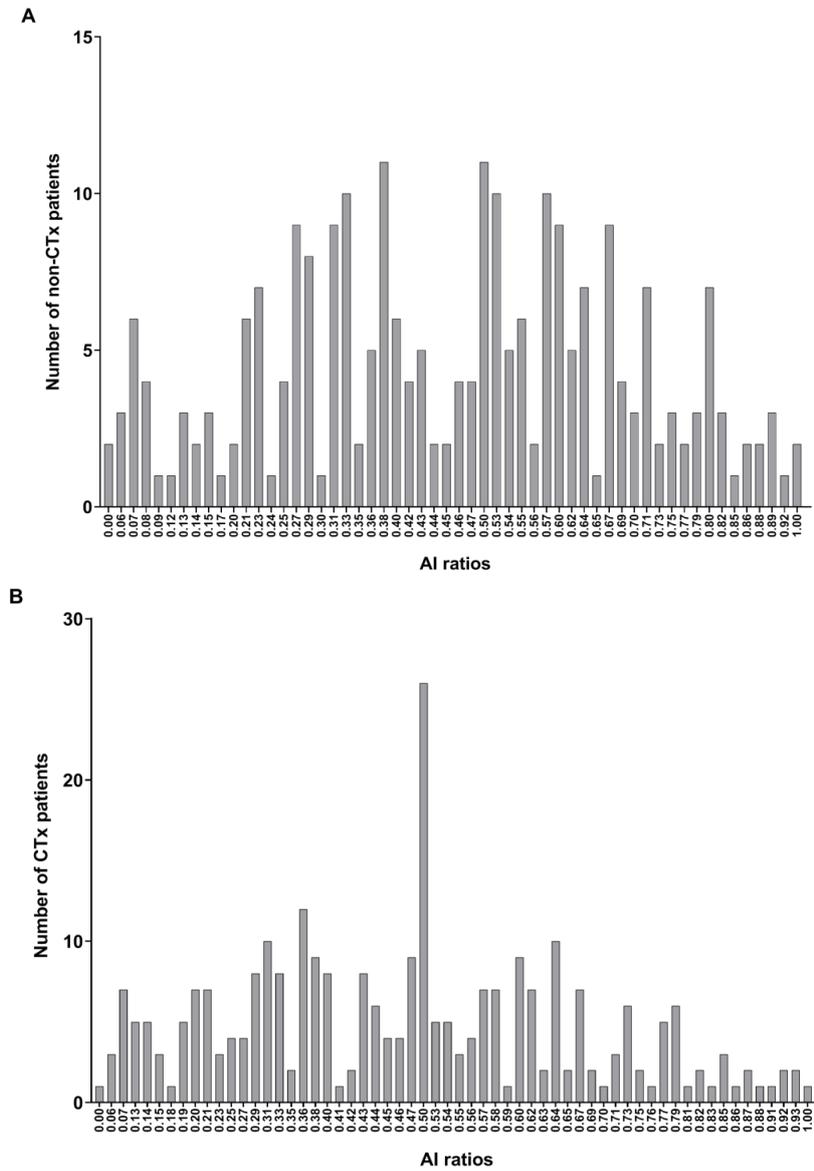
Chromosomal regions	AI-status	No. <sup>1</sup>	Events	Survival probability [%]			Median survival [months] (95% CI)	p-value <sup>2</sup>
				1 yr	3 yrs	5 yrs		
2p21	AI	44	26	79.4	50.1	45.1	38.0 (0-83.6)	0.320
	No AI	22	9	90.5	79.8	62.7	62.2 (17.0-107.5)	
	Total	66	35	75.6	54.1	47.1	61.2 (19.6-102.8)	
4q22	AI	43	23	76.3	66.4	44.8	38.7 (15.2-62.2)	0.782
	No AI	55	31	74.1	54.4	47.6	57.8 (22.5-93.1)	
	Total	98	54	75.6	54.1	47.1	44.6 (12.0-77.3)	
5q11	AI	37	19	78.4	58.4	49.7	38.7 (0-81.7)	0.184
	No AI	57	34	68.4	44.2	41.6	28.4 (14.1-42.7)	
	Total	94	53	75.6	54.1	47.1	36.6 (22.4-50.8)	
5q21	AI	53	30	73.6	55.8	47.6	48.1 (2.1-94.1)	0.450
	No AI	47	28	76.0	50.9	42.7	36.6 (17.9-55.3)	
	Total	100	58	75.6	54.1	47.1	38.7 (13.9-63.5)	
6p25	AI	55	32	77.4	56.2	48.2	57.8 (26.4-89.3)	0.362
	No AI	50	30	69.2	43.7	38.2	28.4 (7.9-43.9)	
	Total	155	62	75.6	54.1	47.1	36.6 (19.0-54.2)	
7q21	AI	46	29	71.7	56.5	57.8	44.6 (15.9-73.4)	0.362
	No AI	49	22	74.1	62.0	56.1	75.3 (29.1-121.5)	
	Total	95	51	75.6	54.1	47.1	61.2 (36.2-86.3)	
7q31	AI	33	21	72.5	50.4	40.7	36.6 (7.3-66.0)	0.441
	No AI	55	29	77.5	56.5	49.6	44.6 (13.6-75.7)	
	Total	88	50	75.6	54.1	47.1	44.6 (16.1-73.1)	
8p23	AI	45	27	75.4	51.5	43.6	36.6 (8.0-65.2)	0.640
	No AI	46	28	68.6	47.1	44.2	32.2 (13.3-51.1)	
	Total	91	55	75.6	54.1	47.1	33.8 (16.0-51.6)	
8q24	AI	70	37	76.6	56.0	45.4	38.7 (12.4-65.0)	0.620
	No AI	47	27	73.7	50.6	47.8	48.1 (5.9-90.3)	
	Total	117	64	75.6	54.1	47.1	38.7 (12.9-64.5)	
9p21	AI	94	51	74.8	56.1	48.3	48.1 (19.3-76.9)	0.677
	No AI	25	14	84.0	49.0	49.0	33.8 (0-91.3)	
	Total	119	65	75.6	54.1	47.1	48.1 (23.1-73.2)	
12p12	AI	83	45	70.4	53.7	46.6	38.7 (7.1-70.3)	0.755
	No AI	37	22	86.1	51.7	44.3	38.0 (11.0-59.6)	
	Total	120	67	75.6	54.1	47.1	38.7 (12.4-65.0)	
16q23	AI	30	22	73.3	35.6	32.0	21.9 (8.6-35.2)	<b>0.070</b>
	No AI	50	25	79.7	57.4	54.9	65.9 (27.8-104.1)	
	Total	80	47	75.6	54.1	47.1	33.8 (4.0-63.6)	
17p13	AI	65	38	69.9	49.3	40.1	31.3 (14.1-48.6)	0.248
	No AI	33	17	90.6	61.0	61.0	62.1 (30.2-94.0)	
	Total	98	55	75.6	54.1	47.1	38.0 (12.5-63.6)	
17q12	AI	68	40	86.1	48.6	41.8	32.2 (20.3-44.1)	0.265
	No AI	54	28	85.0	60.9	53.4	65.9 (41.2-90.7)	
	Total	122	68	75.6	54.1	47.1	44.6 (18.5-70.8)	
17q21	AI	40	21	64.1	52.9	47.0	38.0 (0-79.8)	0.963
	No AI	56	33	83.9	49.9	45.5	33.8 (4.2-63.4)	
	Total	96	54	75.6	54.1	47.1	38.0 (10.0-66.0)	
18q21	AI	75	43	73.7	53.1	43.6	37.9 (12.8-63.0)	0.599
	No AI	38	20	81.6	58.2	54.6	62.1 (26.6-97.7)	
	Total	113	63	75.6	54.1	47.1	44.6 (19.9-69.3)	
19q12	AI	39	22	75.6	50.0	41.2	32.2 (7.2-57.3)	0.973
	No AI	51	29	76.2	50.6	52.3	36.6 (6.8-66.4)	
	Total	90	51	75.6	54.1	47.1	36.6 (22.8-50.4)	

<sup>1</sup>Number of patients with informative markers at respective chromosomal region; <sup>2</sup>p-values of log rank test; p-values < 0.1 in bold

**Supplementary Table 15: Survival data of all patients with resected tumors in association with AI-status**

Chromosomal regions	AI-status	No. <sup>1</sup>	Events	Survival probability [%]			Median survival [months] (95% CI)	p-value <sup>2</sup>
				1 yr	3 yrs	5 yrs		
2p21	AI	152	71	81	49.9	45.8	35.8 (7.9-63.7)	0.662
	No AI	144	68	78.5	51.2	42.9	44.6 (23.6-65.6)	
	Total	296	139	78.9	50.7	43.2	41.6 (25.4-57.8)	
4q22	AI	169	91	76.6	43.5	39.2	27.9 (21.6-34.2)	<b>0.050</b>
	No AI	308	144	80.4	55.3	47.8	46.7 (25.1-68.3)	
	Total	477	235	78.9	50.7	43.2	39.1 (25.8-52.4)	
5q11	AI	122	68	72.3	49.1	40.9	35.3 (10.5-60.1)	0.320
	No AI	291	136	81	49.4	43	35.8 (25.1-46.5)	
	Total	413	204	78.9	50.7	43.2	35.8 (26.4-45.2)	
5q21	AI	197	113	76.9	48.7	40.3	35.1 (23.2-47.0)	0.223
	No AI	258	119	78.4	50.5	43.9	37.9 (25.6-50.2)	
	Total	455	232	78.9	50.7	43.2	35.9 (27.2-44.7)	
6p25	AI	181	95	74.1	48.7	40.5	33.8 (18.9-48.7)	0.346
	No AI	266	124	83.4	52.7	45	44.6 (27.9-61.3)	
	Total	447	219	78.9	50.7	43.2	41.1 (31.2-51.0)	
7q21	AI	134	73	83.2	45.7	41	29.1 (11.9-46.3)	0.596
	No AI	243	114	79.6	51.6	41.9	41.1 (26.9-55.3)	
	Total	377	187	78.9	50.7	43.2	35.1 (24.3-45.9)	
7q31	AI	120	62	80.7	43.1	34.2	28.5 (17.7-39.3)	0.193
	No AI	275	127	80.3	55.8	48.1	45.3 (26.3-64.3)	
	Total	395	189	78.9	50.7	43.2	41.6 (28.5-54.7)	
8p23	AI	170	86	80.4	50.7	44.3	38.7 (16.3-61.1)	0.163
	No AI	180	91	77.5	43.1	32.5	32.4 (25.7-39.1)	
	Total	350	177	78.9	50.7	43.2	33.8 (24.0-43.6)	
8q24	AI	260	129	78.3	52.1	44.1	41.1 (23.3-58.9)	0.770
	No AI	253	122	79.5	50.1	43.2	39.0 (26.7-51.3)	
	Total	513	251	78.9	50.7	43.2	39.1 (27.5-50.7)	
9p21	AI	407	209	77.8	48.9	42.6	33.9 (24.1-43.7)	<b>0.042</b>
	No AI	103	39	88.1	62.3	51.3	67.7 (33.2-102.2)	
	Total	510	248	78.9	50.7	43.2	40.0 (28.4-51.6)	
12p12	AI	300	161	76.3	47.9	39.6	32.7 (24.1-41.4)	<b>0.028</b>
	No AI	190	81	83	55.4	49.9	57.1 (26.6-87.6)	
	Total	490	242	78.9	50.7	43.2	39.0 (26.8-51.2)	
16q23	AI	129	61	82.4	52.3	45.5	41.7 (10.4-73.0)	0.761
	No AI	181	81	80.6	49.6	43.9	35.9 (7.3-50.2)	
	Total	310	142	78.9	50.7	43.2	39.1 (24.2-54.0)	
17p13	AI	217	109	78.6	48.5	43.2	32.7 (16.4-49.0)	0.339
	No AI	161	73	79.7	54	45.6	44.6 (18.1-71.1)	
	Total	378	182	78.9	50.7	43.2	40.0 (26.5-53.6)	
17q12	AI	252	130	75.3	49.2	42	34.3 (20.2-48.4)	0.274
	No AI	278	133	82.4	52.2	44.5	42.2 (26.9-57.5)	
	Total	530	263	78.9	50.7	43.2	39.0 (29.4-48.6)	
17q21	AI	103	45	81.7	58.7	54.6	79.0 (29.0-129.0)	<b>0.066</b>
	No AI	334	165	81.8	50.1	41.8	39.1 (28.5-49.7)	
	Total	437	210	78.9	50.7	43.2	44.3 (30.0-58.7)	
18q21	AI	269	132	77.9	49.4	42.8	35.3 (22.3-48.3)	0.962
	No AI	194	100	79	47.6	40.1	33.8 (22.3-45.3)	
	Total	463	232	78.9	50.7	43.2	34.3 (25.7-42.9)	
19q12	AI	150	85	73.8	45.5	38.3	29.1 (15.6-42.6)	0.145
	No AI	266	129	79.1	49.1	43.8	35.3 (24.8-45.8)	
	Total	416	214	78.9	50.7	43.2	33.8 (25.2-42.4)	

<sup>1</sup>Number of patients with informative markers at respective chromosomal region; <sup>2</sup>p-values of log rank test; p-values < 0.1 in bold



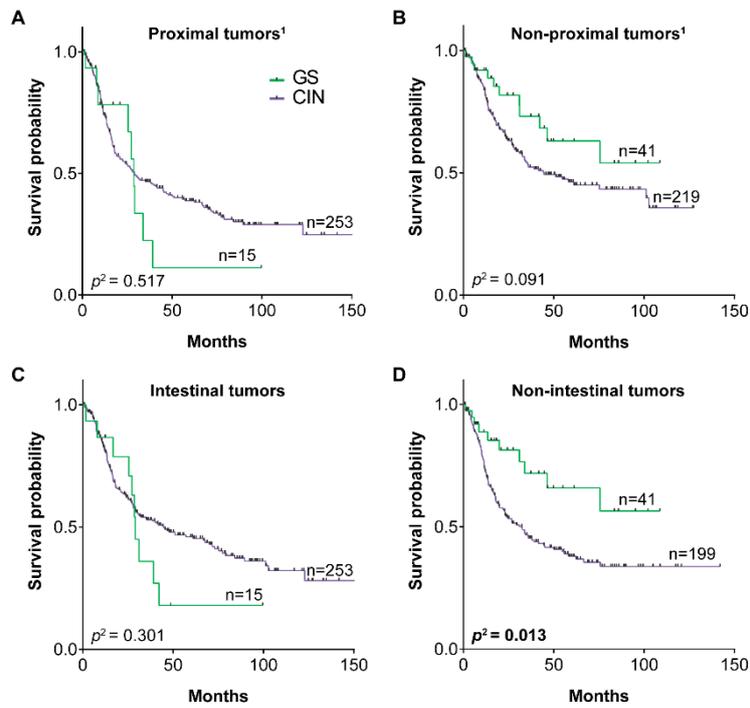
**Supplementary Figure 2: Distribution of AI ratios of resected tumors stratified according to their treatment with neoadjuvant CTx (yes/no).**

The numbers of patients treated without CTx (A) and of patients after CTx (B) are shown. Each bar represents the number of patients with a specific AI ratio measured in the range from 0 to 1.

**Supplementary Table 16: Survival data of the patient cohorts and subgroups in association with the CIN classification according to TCGA**

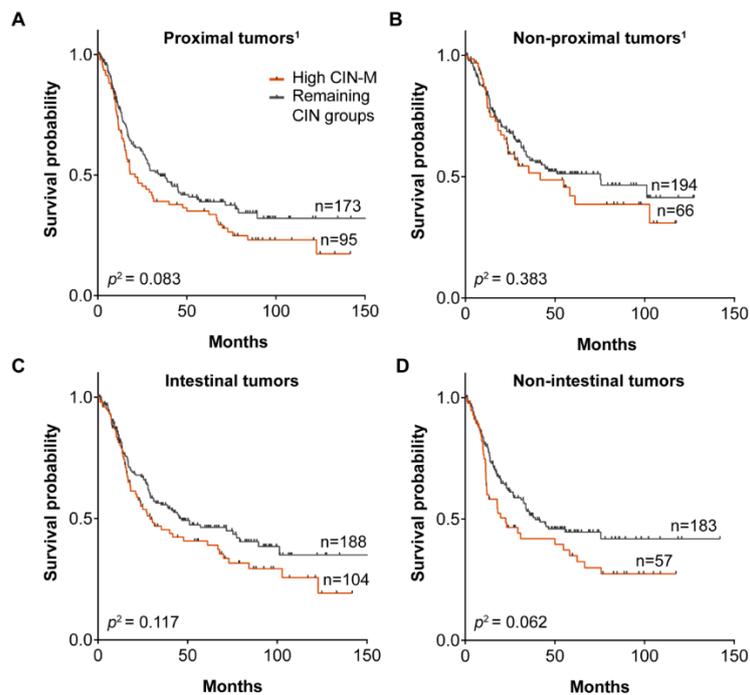
	CIN classification according to TCGA	No.	Events	Survival probability [%]			Median survival [months]	HR	<i>p</i> -value <sup>1</sup>
				1 yr	3 yrs	5 yrs	(95% CI)	(95% CI)	
<b>Tumor biopsies before CTx</b>	GS	7	3	100	60	60	65.9 (12.8-119.1)	0.69 (0.22-2.21)	0.530
	CIN	115	65	74.1	53.4	46.2	38.7 (11.8-65.7)	1 ref.	
	Total	122	68	75.6	54.1	47.1	44.6 (18.5-70.8)	-	
<b>All resected specimens</b>	GS	56	20	88.1	59.3	48.7	46.3 (0-93.0)	0.72 (0.45-1.13)	0.148
	CIN	476	244	77.9	49.7	42.6	35.9 (25.3-46.3)	1 ref.	
	Total	532	264	78.9	50.7	43.2	39.0 (29.3-48.8)	-	
<b>Resected tumors without CTx</b>	GS	26	7	91.6	79.8	57.6	nr	0.62 (0.29-1.34)	0.221
	CIN	222	101	79.5	56.7	50.2	61.1 (27.8-94.4)	1 ref.	
	Total	248	108	80.5	58.7	50.8	61.1 (27.5-94.7)	-	
<b>Resected tumors after CTx</b>	GS	30	13	85.8	39.1	39.1	31.0 (28.0-34.0)	0.78 (0.44-1.38)	0.389
	CIN	254	143	76.5	43.8	36.1	29.0 (22.3-35.7)	1 ref.	
	Total	284	156	77.5	43.7	36.6	30.3 (25.2-35.4)	-	
<b>All resected specimens Proximal<sup>2</sup></b>	GS	15	9	78.3	22.4	11.2	28.6 (24.7-32.5)	1.25 (0.64-2.46)	0.517
	CIN	253	145	74.1	46.7	38.8	29.1 (17.7-40.5)	1 ref.	
	Total	268 <sup>2</sup>	154	74.3	45.8	37.7	29.0 (19.9-38.1)	-	
<b>All resected specimens Non-proximal<sup>2</sup></b>	GS	41	11	92.0	73.2	63.0	nr	0.59 (0.32-1.1)	<b>0.091</b>
	CIN	219	98	82.3	52.4	46.3	44.6 (18.0-71.3)	1 ref.	
	Total	260 <sup>2</sup>	109	83.6	55.3	48.7	54.8 (27.6-82.0)	-	
<b>All resected specimens Intestinal</b>	GS	15	10	86.7	36.0	18.0	29.1 (25.0-33.2)	1.40 (0.74-2.67)	0.301
	CIN	277	136	81.5	53.0	46.0	44.9 (22.2-67.6)	1 ref.	
	Total	292	146	81.7	52.1	44.5	41.7 (25.4-58.0)	-	
<b>All resected specimens Non-intestinal</b>	GS	41	10	88.7	71.9	65.9	nr	0.45 (0.24-0.86)	<b>0.013</b>
	CIN	199	108	73.1	45.4	38.0	32.4 (21.9-42.9)	1 ref.	
	Total	240	118	75.4	49.0	41.7	35.8 (22.3-49.3)	-	

<sup>1</sup>*p*-values of log rank test; *p*-values < 0.1 in bold; <sup>2</sup>No clinical data were available for tumor localization in four patients



**Supplementary Figure 3: Overall survival of all resected patients discriminated by the CIN classification according to TCGA and association with tumor localization and Laurén histopathological subtypes.**

Kaplan-Meier curves are shown for resected patients with proximal located tumors (A), non-proximal located tumors (B), intestinal type tumors (C) and non-intestinal type tumors (D). <sup>1</sup>No clinical data were available for tumor localization in four patients; <sup>2</sup>*p*-values of log rank test, *p*-values < 0.05 in bold



**Supplementary Figure 4: Overall survival of all resected patients discriminated by the CIN classification according to quartiles of the AI ratios in two subgroups and association with tumor localization and Laurén histopathological subtypes.**

Kaplan-Meier curves are shown for resected patients with proximal located tumors (A), non-proximal located tumors localization (B), intestinal type tumors (C) and non-intestinal type tumors (D). <sup>1</sup>No clinical data were available for tumor localization in four patients; <sup>2</sup>*p*-values of log rank test

**Supplementary Table 17: Survival data of the patient cohorts and subgroups in association with the modified CIN classification in four subgroups based on quartiles of the AI ratios**

	Modified CIN classification	No.	Events	Survival probability [%]			Median survival [months] (95% CI)	HR (95% CI)	p-value <sup>1</sup>
				1 yr	3 yrs	5 yrs			
<b>Tumor biopsies before CTx</b>	CIN-L	10	5	90.0	64.3	64.3	62.1 (4.7-119.5)	1 ref.	-
	Low CIN-M	40	24	73.2	47.9	44.7	33.8 (4.3-63.3)	1.35 (0.51-3.54)	0.546
	High CIN-M	48	24	75.0	55.1	50.5	62.2 (-)	0.98 (0.37-2.59)	0.973
	CIN-H	24	15	73.9	55.9	37.3	37.9 (27.0-48.8)	1.32 (0.48-3.64)	0.592
	Total	122	68	75.6	54.1	47.1	44.6 (18.5-70.8)	-	-
<b>All resected specimens</b>	CIN-L	97	37	84.8	58.7	48.8	46.3 (7.4-85.2)	1 ref.	-
	Low CIN-M	214	104	80.1	51.8	43.8	41.1 (25.3-57.0)	1.25 (0.86-1.82)	0.244
	High CIN-M	161	95	74.0	44.1	38.4	27.4 (17.6-37.2)	1.57 (1.07-2.29)	<b>0.021</b>
	CIN-H	60	28	78.5	53.1	47.3	51.0 (0.2-101.8)	1.16 (0.71-1.89)	0.562
	Total	532	264	78.9	50.7	43.2	39.0 (29.3-48.8)	-	-
<b>Resected tumors without CTx</b>	CIN-L	46	14	87.7	73.3	58.7	nr	1 ref.	-
	Low CIN-M	93	39	81.9	56.5	51.6	77.3 (25.9-128.7)	1.46 (0.79-2.69)	0.227
	High CIN-M	80	40	74.4	56.2	48.3	54.8 (23.7-85.9)	1.69 (0.92-3.11)	<b>0.091</b>
	CIN-H	29	15	82.1	52.3	45.7	51.0 (0.96-101.04)	1.68 (0.81-3.47)	0.165
	Total	248	108	80.5	58.7	50.8	61.1 (27.5-94.7)	-	-
<b>Resected tumors after CTx</b>	CIN-L	51	23	82.5	43.5	38.7	31.1 (22.0-40.3)	1 ref.	-
	Low CIN-M	121	65	78.7	48.5	38.0	35.1 (21.1-49.1)	1.08 (0.67-1.75)	0.741
	High CIN-M	81	55	73.6	32.0	28.6	21.0 (15.5-36.5)	1.53 (0.94-2.49)	<b>0.087</b>
	CIN-H	31	13	74.8	54.0	48.6	38.7 (-)	0.84 (0.42-1.66)	0.611
	Total	284	156	77.5	43.7	36.6	30.3 (25.2-35.4)	-	-

<sup>1</sup>p-values of Cox's regression compared to CIN-L (ref.); p-values < 0.1 in bold

**Supplementary Table 18: Survival data of all resected patients stratified according to tumor localization and Laurén histopathological subtypes and association with the high CIN-M subgroup**

All resected specimens	CIN-status	No.	Events	Survival probability [%]			Median survival [months] (95% CI)	HR (95% CI)	p-value <sup>1</sup>
				1 yr	3 yrs	5 yrs			
<b>Proximal</b>	High CIN-M	95	64	68.7	39.1	35.1	21.0 (11.3-30.7)	1.33 (0.96-1.83)	<b>0.083</b>
	Remaining CIN-groups	173	90	77.6	49.6	38.9	35.1 (22.8-47.4)	1 ref.	
	Total	268 <sup>2</sup>	154	74.3	45.8	37.7	29 (19.9-38.1)	-	
<b>Non-proximal</b>	High CIN-M	66	31	81.8	51.5	42.1	41.7 (12.1-71.3)	1.20 (0.79-1.83)	0.383
	Remaining CIN-groups	194	78	84.2	56.6	51.2	75.3 (36.2-114.4)	1 ref.	
	Total	260 <sup>2</sup>	109	83.6	55.3	48.7	54.8 (27.6-82.0)	-	
<b>Intestinal</b>	High CIN-M	104	58	81.3	45.4	40.8	29.0 (13.5-44.5)	1.30 (0.93-1.82)	0.117
	Remaining CIN-groups	188	88	81.9	55.7	46.4	46.7 (19.2-74.2)	1 ref.	
	Total	292	146	81.7	52.1	44.5	41.7 (25.4-58.0)	-	
<b>Non-intestinal</b>	High CIN-M	57	37	62.0	41.9	34.9	22.6 (9.8-35.4)	1.45 (0.98-2.14)	<b>0.062</b>
	Remaining CIN-groups	183	81	80.0	51.5	44.6	39.0 (22.8-55.2)	1 ref.	
	Total	240	118	75.4	49.0	41.7	35.8 (22.3-49.3)	-	

<sup>1</sup>p-values of log rank test; p-values < 0.1 in bold; <sup>2</sup>No clinical data were available for tumor localization in four patients

**Supplementary Table 19: Frequency of AI at 17 chromosomal regions in the resected tumor cohort and classification in the four CIN-subgroups determined according to the quartiles of the AI ratios**

	CIN-L			Low CIN-M		
Chromosomal regions	Number of tumors with AI	Number of informative tumors	Frequency of AI (%) <sup>1</sup>	Number of tumors with AI	Number of informative tumors	Frequency of AI (%) <sup>1</sup>
2p21	14	66	21	68	126	54
4q22	9	86	10	54	194	28
5q11	5	83	6	29	169	17
5q21	9	79	11	64	184	35
6p25	11	86	13	62	179	35
7q21	3	74	4	44	159	28
7q31	7	70	10	29	165	18
8p23	14	71	20	78	150	52
8q24	19	91	21	84	208	40
9p21	36	91	40	175	208	84
12p12	23	95	24	108	193	56
16q23	8	69	12	38	117	32
17p13	13	69	19	80	154	52
17q12	16	97	16	84	213	39
17q21	5	80	6	24	182	13
18q21	18	86	21	96	184	52
19q12	4	77	5	38	168	23
	High CIN-M			CIN-H		
Chromosomal regions	Number of tumors with AI	Number of informative tumors	Frequency of AI (%) <sup>1</sup>	Number of tumors with AI	Number of informative tumors	Frequency of AI (%) <sup>1</sup>
2p21	47	78	60	23	26	88
4q22	64	147	44	42	50	84
5q11	58	122	48	30	39	77
5q21	84	144	58	40	48	83
6p25	76	135	56	32	47	68
7q21	64	110	58	23	34	68
7q31	55	119	46	29	41	71
8p23	52	92	57	26	37	70
8q24	110	156	71	47	58	81
9p21	141	154	92	55	57	96
12p12	117	148	79	52	54	96
16q23	56	94	60	27	30	90
17p13	86	116	74	38	39	97
17q12	98	160	61	54	60	90
17q21	38	125	30	36	50	72
18q21	110	141	78	45	52	87
19q12	76	131	58	32	40	80

<sup>1</sup>Number of tumors with AI divided by number of informative tumors

**Supplementary Table 20: Molecular classification according to TCGA and association with patient characteristics**

Category	Value	Resected tumors (n=616)						
		GS <sup>1</sup> [n]	EBV(+) [n]	<i>p</i> -value <sup>2</sup>	MSI-H [n]	<i>p</i> -value <sup>2</sup>	CIN [n]	<i>p</i> -value <sup>2</sup>
<b>Cases</b>	Total	56	24		59		477	
<b>Age [yrs]</b>	Median	64.9	57.4		71.4		64.1	
	Range	33.4-85.2	29.2-80.9		40.4-84.9		28.3-90.9	
<b>Age Median [yrs]</b>	<64.6	27	17	0.062	13	<b>0.003</b>	249	0.572
	≥64.6	29	7		46		228	
<b>Gender</b>	Male	40	22	<b>0.047</b>	38	0.420	352	0.704
	Female	16	2		21		125	
<b>Tumor localization</b>	Proximal	15	9	<b>0.026</b>	23	0.210	253	<b>0.001</b>
	Middle/Total	17	12		14		110	
	Distal	19	2		21		88	
	Total/limitis	5	0		1		22	
	n/a	0	1		0		4	
<b>Proximal versus non-proximal</b>	Proximal	15	9	0.278	23	0.165	253	< <b>0.001</b>
	Non-proximal	41	14		36		220	
	NA	0	1		0		4	
<b>Laurén classification</b>	Intestinal	15	15	<b>0.002</b>	39	< <b>0.001</b>	278	< <b>0.001</b>
	Non-intestinal	41	9		20		199	
<b>Tumor grade</b>	G1/2	5	0	0.313	11	0.101	108	<b>0.012</b>
	G3/4	44	18		38		299	
	n/a	7	6		10		70	
<b>Clinical tumor stage</b>	cT2	15	6	0.868	12	0.415	111	0.570
	cT3/4	41	18		47		364	
	n/a	0	0		0		2	
<b>(y)pT<sup>3</sup></b>	(y)pT1	6	3	0.887	3	0.476	44	0.817
	(y)pT2	8	3		8		60	
	(y)pT3	26	13		35		254	
	(y)pT4	16	5		13		119	
<b>(y)pT1/2 versus (y)pT3/4<sup>3</sup></b>	(y)pT1/2	14	6	1.00	11	0.409	104	0.586
	(y)pT3/4	42	18		48		373	
<b>(y)pN<sup>3</sup></b>	Negative	18	9	0.642	21	0.696	141	0.689
	Positive	38	15		38		336	
<b>Metastasis status</b>	No	45	23	0.076	56	<b>0.017</b>	409	0.283
	Yes	11	1		3		68	
<b>Resection category</b>	R0	47	23	0.140	47	0.554	352	0.098
	R1	9	1		12		125	
<b>Neoadjuvant CTx</b>	No	26	8	0.278	35	0.166	222	0.987
	Yes	30	16		24		255	

<sup>1</sup>Reference group (ref.); <sup>2</sup>*p*-values of Chi-square test or Fisher's exact test compared to CIN-L (ref.), *p*-values < 0.05 in bold; <sup>3</sup>Classification according to 7<sup>th</sup> Edition UICC

**Supplementary Table 21: Modified molecular classification system based on TCGA and association with patient characteristics**

Category	Value	Resected tumors (n=615 <sup>1</sup> )								
		Remaining CIN-groups <sup>2</sup> [n]	EBV(+) [n]	<i>p</i> -value <sup>3</sup>	MSI-H [n]	<i>p</i> -value <sup>3</sup>	MSI-L [n]	<i>p</i> -value <sup>3</sup>	High CIN-M [n]	<i>p</i> -value <sup>3</sup>
<b>Cases</b>	Total	355	23		59		27		151	
<b>Age [yrs]</b>	Median	63.7	57.6		71.4		66.5		64.4	
	Range	31.7-88.3	29.2-80.9		40.4-84.9		49.3-82.2		28.3-90.9	
<b>Age Median [yrs]</b>	<64.6	189	16	0.128	13	<0.001	11	0.210	76	0.549
	≥64.6	166	7		46		16		75	
<b>Gender</b>	Male	255	22	<b>0.012</b>	38	0.246	23	0.133	114	0.396
	Female	100	1		21		4		37	
<b>Tumor localization</b>	Proximal	166	9		23		14		88	
	Middle/Total	89	12	<b>0.028</b>	14	0.075	7	0.933	31	0.131
	Distal	75	2		21		5		27	
	Total/limitis	21	0		1		1		5	
	n/a	4	0		0		0		0	
<b>Proximal vs. non-proximal</b>	Proximal	166	9		23		14		88	
	Non-proximal	185	14	0.447	36	0.236	13	0.648	63	<b>0.024</b>
	n/a	4	0		0		0		0	
<b>Laurén classification</b>	Intestinal	177	14		39		20		96	
	Non-intestinal	178	9	0.306	20	<b>0.021</b>	7	<b>0.015</b>	55	<b>0.005</b>
<b>Tumor grade</b>	G1/2	70	0	<b>0.021</b>	11	0.920	4	0.468	39	0.118
	G3/4	233	18		38		20		90	
	n/a	52	5		10		3		22	
<b>Clinical tumor stage</b>	cT2	83	5	0.851	12	0.600	5	0.558	38	0.650
	cT3/4	271	18		47		22		112	
	n/a	1	0		0		0		1	
<b>(y)pT<sup>4</sup></b>	(y)pT1	38	3		3		0		12	
	(y)pT2	42	3		8		6		20	
	(y)pT3	176	12	0.891	35	0.345	15	0.127	89	0.143
	(y)pT4	99	5		13		6		30	
	(y)pT1/2	80	6		11		6		32	
<b>(y)pN<sup>4</sup></b>	(y)pT3/4	275	17	0.694	48	0.504	21	0.970	119	0.739
	Negative	114	9		21		8		37	
	Positive	241	14	0.486	38	0.597	19	0.790	114	0.087
<b>Metastasis status</b>	No	300	22	0.145	56	<b>0.033</b>	24	0.541	130	0.648
	Yes	55	1		3		3		21	
<b>Resection category</b>	R0	263	22	<b>0.020</b>	47	0.360	20	0.999	116	0.516
	R1	92	1		12		7		35	
<b>Neoadjuvant CTx</b>	No	159	7	0.179	35	<b>0.038</b>	14	0.477	75	0.314
	Yes	196	16		24		13		76	

<sup>1</sup>One resected tumor without CTx was positive for MSI-L and EBV. This tumor was excluded in the modified molecular classification system. <sup>2</sup>Reference group (ref.); <sup>3</sup>*p*-values of Chi-square test or Fisher's exact test compared to the remaining CIN-groups (ref.), *p*-values < 0.05 in bold; <sup>4</sup>Classification according to 7<sup>th</sup> Edition UICC

**Supplementary Table 22: Molecular classification system according to TCGA and association with response to neoadjuvant CTx of tumor biopsies before CTx and tumor regression grade of resected tumors after neoadjuvant CTx**

	CIN <sup>1</sup> [n]	EBV(+) [n]	<i>p</i> -value <sup>2</sup>	MSI-H [n]	<i>p</i> -value <sup>2</sup>	GS [n]	<i>p</i> -value <sup>2</sup>
<b>Tumor biopsies before neoadjuvant CTx (n=143)</b>							
Responder (TRG1)	36	3	0.386	4	0.714	2	1.0
Non-responder (TRG2/3)	79	3		11		5	
Total	115	6		15		7	
<b>Resected tumors after neoadjuvant CTx (n=325)</b>							
TRG2	125	8	0.939	4	<b>0.002</b>	16	0.655
TRG3	130	8		20		14	
Total	255	16		24		30	

<sup>1</sup>Due to small sample size in the GS subgroup, CIN was set as reference (ref.); <sup>2</sup>*p*-values of Chi square test or Fisher's exact test compared to CIN (ref.); *p*-values < 0.05 in bold

**Supplementary Table 23: Modified molecular classification system based on TCGA with high CIN-M as own subgroup and association with response to neoadjuvant CTx of tumor biopsies before CTx and tumor regression grade of resected tumors after neoadjuvant CTx**

	Remaining CIN- groups <sup>1</sup> [n]	EBV(+) [n]	<i>p</i> - value <sup>2</sup>	MSI-H [n]	<i>p</i> - value <sup>2</sup>	MSI- L [n]	<i>p</i> - value <sup>2</sup>	High CIN-M [n]	<i>p</i> - value <sup>2</sup>
<b>Tumor biopsies before neoadjuvant CTx (n=142)</b>									
Responder (TRG1)	22	2	0.648	4	0.741	5	<b>0.018</b>	11	0.447
Non-responder (TRG2/3)	49	3		11		1		34	
Total	71	5		15		6		45	
<b>Resected tumors after neoadjuvant CTx (n=325)</b>									
TRG2	93	8	0.844	4	<b>0.004</b>	6	0.928	42	0.247
TRG3	103	8		20		7		34	
Total	196	16		24		13		76	

<sup>1</sup>Reference group (ref.); <sup>2</sup>*p*-values of Chi-square test or Fisher's exact test compared to the remaining CIN-groups (ref.); *p*-values < 0.05 in bold

**Supplementary Table 24: Modified molecular classification system based on TCGA with CIN-H as own subgroup and association with response to neoadjuvant CTx of tumor biopsies before CTx and tumor regression grade of resected tumors after neoadjuvant CTx**

	CIN-L and CIN-M <sup>1</sup> [n]	EBV(+) [n]	<i>p</i> - value <sup>2</sup>	MSI-H [n]	<i>p</i> - value <sup>2</sup>	MSI-L [n]	<i>p</i> - value <sup>2</sup>	CIN-H [n]	<i>p</i> - value <sup>2</sup>
<b>Tumor biopsies before neoadjuvant CTx (n=142)</b>									
Responder (TRG1)	23	2	0.598	4	0.855	5	<b>0.006</b>	10	<b>0.050</b>
Non-responder (TRG2/3)	71	3		11		1		12	
Total	94	5		15		6		22	
<b>Resected tumors after neoadjuvant CTx (n=325)</b>									
TRG2	125	8	0.924	4	<b>0.001</b>	6	0.721	10	0.120
TRG3	119	8		20		7		18	
Total	244	16		24		13		28	

<sup>1</sup>Reference group (ref.); <sup>2</sup>*p*-values of Chi-square test or Fisher's exact test compared to CIN-L and CIN-M (ref.); *p*-values < 0.05 in bold

**Supplementary Table 25: List of mutations identified by targeted sequencing of gastric carcinomas using the custom designed GC related gene panel**

Gene	Exon	cDNA <sup>1</sup>	Protein <sup>2</sup>	Type of mutation	Entry in database <sup>3</sup>
APC	17	c.4348C>T	p.R1450*	Nonsense	COSM13127
ARID1A	20	c.5965C>T	p.R1989*	Nonsense	COSM51425
ARID1A	20	c.6017_6033del	p.G2006fs	Frameshift del.	-
ARID1A	5	c.2077C>T	p.R693*	Nonsense	COSM184236
ATM	29	c.4385C>G	A1462G	Missense	-
CCND1	5	c.772_787del	p.S258fs	Frameshift del.	-
CDH1	13	c.2145_2164del	p.G715fs	Frameshift del.	-
CDH1	8	c.1088T>A	p.I363N	Missense	-
CDH1	7	c.975_998del	p.V325_L333del	Nonframeshift del.	-
CDH1	8	c.1009-2A>G	X337_splice	Splice variant	cBioPortal
CDH1	1	c.11G>A	p.W4*	Nonsense	-
CDH1	5	c.641T>C	p.L214P	Missense	COSM5055032
CDH1	3	c.220C>T	p.R74*	Nonsense	COSM25265
CDH1	7	c.947_955del	p.M316_T318del	Nonframeshift del.	-
CDH1	10	c.1756-4GTAA del	X522_splice	Splice variant	cBioPortal
CDKN2A	1	c.148_164del	p.Q50fs	Frameshift del.	COSM6903270
CDKN2A	2	c.162_164del	p.S4_55del	Nonframeshift del.	-
CDKN2A	2	c.204_218del	p.68_73del	Nonframeshift del.	-
CTNNA1	6	c.643C>T	p.Q215*	Nonsense	-
CTNNB1	3	c.134C>T	p.S45F	Missense	COSM5667
ERBB2	20	c.2329G>C	p.V777L	Missense	COSM436500
ERBB2	8	c.929C>T	p.S310F	Missense	COSM48358
ERBB3	7	c.C734>T	p.A245V	Missense	COSM941485
ERBB3	7	c.850G>A	p.G284R	Missense	COSM48360
ERBB4	23	c.2762A>T	p.Y921F	Missense	-
FBXW7	9	c.1322G>A	p.R441Q	Missense	MU1843281 (ICGC)
FBXW7	10	c.1513C>T	p.R505C	Missense	COSM108572
FGFR1 <sup>4</sup>	8	c.1010G>A	p.G337E	Missense	rs1064793122
FGFR1 <sup>4</sup>	8	c.1042G>A	p.G348R	Missense	rs886037634
FGFR1 <sup>4</sup>	10	c.1334G>A	p.R445Q	Missense	MU1514106 (ICGC)
KRAS	2	c.34G>A	p.G12S	Missense	COSM517
KRAS	2	c.35G>T	p.G12V	Missense	COSM520
NRAS	2	c.34G>T	p.G12C	Missense	COSM562
PIK3CA	10	c.1624G>A	p.E542K	Missense	COSM760
PIK3CA	10	c.1634A>C	p.E545A	Missense	COSM12458
PIK3CA	10	c.1633G>A	p.E545K	Missense	COSM763
PIK3R1	13	c.1709_1714GGA	p.L570_Q572 Indel	Nonframeshift subst.	-

PREX2	2	c.148T>G	p.L50V	Missense	COSM268334
PREX2	2	c.190A>C	p.N64H	Missense	-
PREX2	24	c.2738G>T	p.R913M	Missense	-
PTEN	5	c.335T>A	p.L112Q	Missense	COSM13574
PTPR1	8	c.1317G>T	p.E439D	Missense	-
RHOA	2	c.125A>G	p.Y42C	Missense	COSM2849892
RNF43	8	c.935G>A	p.C312Y	Missense	-
SMAD4	9	c.1049T>G	p.V350G	Missense	COSM6909225
SMAD4	9	c.1067C>A	p.P356H	Missense	COSM6922713
TGFBR2	6	c.1405_1408del	p.Q469fs	Frameshift del.	-
TGFBR2	5	c.1243G>A	p.D415N	Missense	COSM5864275
TGFBR2	8	c.1649C>G	p.P550R	Missense	-
TGFBR2	5	c.1226_1227ins TATCCTCG	p.N409fs	Frameshift insertion	-
TGFBR2	5	c.1259T>A	p.L420Q	Missense	-
TGFBR2	5	c.1217A>G	p.K406R	Missense	-
TLR4	3	c.1459T>G	p.F487V	Missense	COSM1267921
TP53	6	c.567_577del	p.A189fs	Frameshift del.	-
TP53	5	c.404G>A	p.C135Y	Missense	COSM10801
TP53	5	c.527G>A	p.C176Y	Missense	COSM10687
TP53	8	c.824G>A	p.C275Y	Missense	COSM10893
TP53	8	c.892G>T	p.E298*	Nonsense	COSM10710
TP53	4	c.338dupT	p.F113fs	Frameshift insertion	IACR TP53 database
TP53	7	c.764_766delTCA	p.I255_256del	Nonframeshift del.	COSM1480061
TP53	9	c.989T>G	p.L330R	Missense	COSM4384914
TP53	10	c.1043T>G	p.L348W	Missense	COSM5013841
TP53	5	c.392A>T	p.N131I	Missense	COSM44794
TP53	5	c.524G>A	p.R175H	Missense	COSM10648
TP53	5	c.541C>T	p.R181C	Missense	COSM11090
TP53	7	c.743G>A	p.R248Q	Missense	COSM1640830
TP53	8	c.817C>T	p.R273C	Missense	COSM10659
TP53	8	c.818G>A	p.R273H	Missense	COSM10660
TP53	8	c.818G>T	p.R273L	Missense	COSM1640828
TP53	8	c.844C>T	p.R282W	Missense	COSM10704
TP53	8	c.916C>T	p.R306*	Nonsense	COSM10663
TP53	10	c.1009C>T	p.R337C	Missense	COSM11071
TP53	7	c.722C>T	p.S241F	Missense	COSM1649402
TP53	5	c.437G>A	p.W146*	Nonsense	COSM43609
TP53	6	c.659A>G	p.Y220C	Missense	COSM10758
TP53	6	c.560_562-7del CTGGATTCTC	X187_splice	Splice variant	cBioPortal
TP53	7	c.673-1G>C	X225_splice	Splice variant	-
TP53	9	c.920-2A>G	X307_splice	Splice variant	rs397516439
XIRP2	9	c.4127_4129del	p.Q1376_E1377del	Nonframeshift del.	COSM5021196

<sup>1</sup>The change that has occurred in the nucleotide sequence. <sup>2</sup>The change that has occurred in the peptide sequence. <sup>3</sup>COSMIC database (Tate et al. 2019), cBioPortal (Cerami et al. 2012, Gao et al. 2013), ICGC Data Portal (ICGC, Nature 2010), IACR TP53 database (Bouaoun et al. 2016); <sup>4</sup>Sequence variants should be validated by Sanger sequencing.

## 7. BIBLIOGRAPHY

- Al-Batran SE, Hofheinz RD, Pauligk C, Kopp HG, Haag GM, Luley KB, Meiler J, Homann N, Lorenzen S, Schmalenberg H, Probst S, Koenigsmann M, Egger M, Prasnikař N, Caca K, Trojan J, Martens UM, Block A, Fischbach W, Mahlberg R, Clemens M, Illerhaus G, Zirlik K, Behringer DM, Schmiegel W, Pohl M, Heike M, Ronellenfitsch U, Schuler M, Bechstein WO, Konigsrainer A, Gaiser T, Schirmacher P, Hozaeel W, Reichart A, Goetze TO, Sievert M, Jager E, Monig S and Tannapfel A (2016). Histopathological regression after neoadjuvant docetaxel, oxaliplatin, fluorouracil, and leucovorin versus epirubicin, cisplatin, and fluorouracil or capecitabine in patients with resectable gastric or gastro-oesophageal junction adenocarcinoma (FLOT4-AIO): results from the phase 2 part of a multicentre, open-label, randomised phase 2/3 trial. *Lancet Oncol* 17(12): 1697-1708.
- Al-Batran SE, Homann N, Pauligk C, Goetze TO, Meiler J, Kasper S, Kopp HG, Mayer F, Haag GM, Luley K, Lindig U, Schmiegel W, Pohl M, Stoehlmacher J, Folprecht G, Probst S, Prasnikař N, Fischbach W, Mahlberg R, Trojan J, Koenigsmann M, Martens UM, Thuss-Patience P, Egger M, Block A, Heinemann V, Illerhaus G, Moehler M, Schenk M, Kullmann F, Behringer DM, Heike M, Pink D, Teschendorf C, Lohr C, Bernhard H, Schuch G, Rethwisch V, von Weikersthal LF, Hartmann JT, Kneba M, Daum S, Schulmann K, Weniger J, Belle S, Gaiser T, Oduncu FS, Guntner M, Hozaeel W, Reichart A, et al. (2019). Perioperative chemotherapy with fluorouracil plus leucovorin, oxaliplatin, and docetaxel versus fluorouracil or capecitabine plus cisplatin and epirubicin for locally advanced, resectable gastric or gastro-oesophageal junction adenocarcinoma (FLOT4): a randomised, phase 2/3 trial. *Lancet* 393(10184): 1948-1957.
- Alkan C, Coe BP and Eichler EE (2011). Genome structural variation discovery and genotyping. *Nat Rev Genet* 12(5): 363-376.
- Amin MB, Greene FL, Edge SB, Compton CC, Gershenwald JE, Brookland RK, Meyer L, Gress DM, Byrd DR and Winchester DP (2017). The Eighth Edition AJCC Cancer Staging Manual: Continuing to build a bridge from a population-based to a more "personalized" approach to cancer staging. *CA Cancer J Clin* 67(2): 93-99.
- An JY, Kim H, Cheong JH, Hyung WJ, Kim H and Noh SH (2012). Microsatellite instability in sporadic gastric cancer: its prognostic role and guidance for 5-FU based chemotherapy after R0 resection. *Int J Cancer* 131(2): 505-511.
- Arienti C, Pignatta S and Tesei A (2019). Epidermal Growth Factor Receptor Family and its Role in Gastric Cancer. *Front Oncol* 9: 1308-1308.
- Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, Marchini JL, McCarthy S, McVean GA and Abecasis GR (2015). A global reference for human genetic variation. *Nature* 526(7571): 68-74.
- Azzoni C, Bottarelli L, Cecchini S, Silini EM, Bordi C and Sarli L (2011). Sporadic colorectal carcinomas with low-level microsatellite instability: a distinct subgroup with specific clinicopathological and molecular features. *Int J Colorectal Dis* 26(4): 445-453.
- Bakhoun SF and Compton DA (2012). Chromosomal instability and cancer: a complex relationship with therapeutic potential. *J Clin Invest* 122(4): 1138-1143.
- Balakrishnan M, George R, Sharma A and Graham DY (2017). Changing Trends in Stomach Cancer Throughout the World. *Curr Gastroenterol Rep* 19(8): 36.
- Bang YJ, Van Cutsem E, Feyereislova A, Chung HC, Shen L, Sawaki A, Lordick F, Ohtsu A, Omuro Y, Satoh T, Aprile G, Kulikov E, Hill J, Lehle M, Ruschhoff J and Kang YK (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(9742): 687-697.
- Baudrin LG, Deleuze J-F and How-Kit A (2018). Molecular and Computational Methods for the Detection of Microsatellite Instability in Cancer. *Front Oncol* 8: 621-621.
- Bauer L, Hapfelmeier A, Blank S, Reiche M, Slotta-Huspenina J, Jesinghaus M, Novotny A, Schmidt T, Grosser B, Kohlruss M, Weichert W, Ott K and Keller G (2018). A novel pretherapeutic gene expression-based risk score for treatment guidance in gastric cancer. *Ann Oncol* 29(1): 127-132.
- Becker K, Langer R, Reim D, Novotny A, Meyer zum Buschenfelde C, Engel J, Friess H and Hofler H (2011). Significance of histopathological tumor regression after neoadjuvant chemotherapy in gastric adenocarcinomas: a summary of 480 cases. *Ann Surg* 253(5): 934-939.

- Becker KF, Atkinson MJ, Reich U, Becker I, Nekarda H, Siewert JR and Hofler H (1994). E-cadherin gene mutations provide clues to diffuse type gastric carcinomas. *Cancer Res* 54(14): 3845-3852.
- Beckmann MW, Picard F, An HX, van Roeyen CR, Dominik SI, Mosny DS, Schnurch HG, Bender HG and Niederacher D (1996). Clinical impact of detection of loss of heterozygosity of BRCA1 and BRCA2 markers in sporadic breast cancer. *Br J Cancer* 73(10): 1220-1226.
- Bellini MF, Cadamuro ACT, Succi M, Proença MA and Silva AE (2012). Alterations of the TP53 gene in gastric and esophageal carcinogenesis. *J Biomed Biotechnol* 2012: 891961.
- Berlth F, Bollschweiler E, Drebber U, Hoelscher AH and Moenig S (2014). Pathohistological classification systems in gastric cancer: diagnostic relevance and prognostic value. *World J Gastroenterol* 20(19): 5679-5684.
- Bertuccio P, Rosato V, Andreano A, Ferraroni M, Decarli A, Edefonti V and La Vecchia C (2013). Dietary patterns and gastric cancer risk: a systematic review and meta-analysis. *Ann Oncol* 24(6): 1450-1458.
- Birkbak NJ, Eklund AC, Li Q, McClelland SE, Endesfelder D, Tan P, Tan IB, Richardson AL, Szallasi Z and Swanton C (2011). Paradoxical relationship between chromosomal instability and survival outcome in cancer. *Cancer Res* 71(10): 3447-3452.
- Blank S, Stange A, Sisic L, Roth W, Grenacher L, Sterzing F, Burian M, Jager D, Buchler M and Ott K (2013). Preoperative therapy of esophagogastric cancer: the problem of nonresponding patients. *Langenbecks Arch Surg* 398(2): 211-220.
- Boland CR, Thibodeau SN, Hamilton SR, Sidransky D, Eshleman JR, Burt RW, Meltzer SJ, Rodriguez-Bigas MA, Fodde R, Ranzani GN and Srivastava S (1998). A National Cancer Institute Workshop on Microsatellite Instability for cancer detection and familial predisposition: development of international criteria for the determination of microsatellite instability in colorectal cancer. *Cancer Res* 58(22): 5248-5257.
- Bouaoun L, Sonkin D, Ardin M, Hollstein M, Byrnes G, Zavadil J and Olivier M (2016). TP53 Variations in Human Cancers: New Lessons from the IARC TP53 Database and Genomics Data. *Hum Mutat* 37(9): 865-876.
- Brar G and Shah MA (2019). The role of pembrolizumab in the treatment of PD-L1 expressing gastric and gastroesophageal junction adenocarcinoma. *Therap Adv Gastroenterol* 12: 1756284819869767.
- Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA and Jemal A (2018). Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 68(6): 394-424.
- Brenner H, Rothenbacher D and Arndt V (2009). Epidemiology of stomach cancer. *Methods Mol Biol* 472: 467-477.
- Broman KW, Murray JC, Sheffield VC, White RL and Weber JL (1998). Comprehensive human genetic maps: individual and sex-specific variation in recombination. *Am J Hum Genet* 63(3): 861-869.
- Brýs M, Migdalska-Sek M, Pastuszak-Lewandoska D, Forma E, Czarnecka K, Domanska D, Nawrot E, Wilkosz J, Rozanski W and Brzezianska E (2013). Diagnostic value of DNA alteration: loss of heterozygosity or allelic imbalance-promising for molecular staging of prostate cancers. *Med Oncol* 30(1): 391.
- Cai H, Jing C, Chang X, Ding D, Han T, Yang J, Lu Z, Hu X, Liu Z, Wang J, Shang L, Wu S, Meng P, Lin L, Zhao J, Nie M and Yin K (2019). Mutational landscape of gastric cancer and clinical application of genomic profiling based on target next-generation sequencing. *J Transl Med* 17(1): 189.
- Camargo MC, Kim WH, Chiaravalli AM, Kim KM, Corvalan AH, Matsuo K, Yu J, Sung JJ, Herrera-Goepfert R, Meneses-Gonzalez F, Kijima Y, Natsugoe S, Liao LM, Lissowska J, Kim S, Hu N, Gonzalez CA, Yatabe Y, Koriyama C, Hewitt SM, Akiba S, Gulley ML, Taylor PR and Rabkin CS (2014). Improved survival of gastric cancer with tumour Epstein-Barr virus positivity: an international pooled analysis. *Gut* 63(2): 236-243.
- Candusso ME, Luinetti O, Villani L, Alberizzi P, Klersy C, Fiocca R, Ranzani GN and Solcia E (2002). Loss of heterozygosity at 18q21 region in gastric cancer involves a number of cancer-related genes and correlates with stage and histology, but lacks independent prognostic value. *J Pathol* 197(1): 44-50.
- Carethers JM (2017). Microsatellite Instability Pathway and EMAS in Colorectal Cancer. *Curr Colorectal Cancer Rep* 13(1): 73-80.

- Carter SL, Eklund AC, Kohane IS, Harris LN and Szallasi Z (2006). A signature of chromosomal instability inferred from gene expression profiles predicts clinical outcome in multiple human cancers. *Nat Genet* 38(9): 1043-1048.
- Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, Jacobsen A, Byrne CJ, Heuer ML, Larsson E, Antipin Y, Reva B, Goldberg AP, Sander C and Schultz N (2012). The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov* 2(5): 401-404.
- Chen K, Yang D, Li X, Sun B, Song F, Cao W, Brat DJ, Gao Z, Li H, Liang H, Zhao Y, Zheng H, Li M, Buckner J, Patterson SD, Ye X, Reinhard C, Bhatena A, Joshi D, Mischel PS, Croce CM, Wang YM, Raghavakaimal S, Li H, Lu X, Pan Y, Chang H, Ba S, Luo L, Cavenee WK, Zhang W and Hao X (2015). Mutational landscape of gastric adenocarcinoma in Chinese: implications for prognosis and therapy. *Proc Natl Acad Sci USA* 112(4): 1107-1112.
- Chernik MR (2008). *Bootstrap Methods: A Guide for Practitioners and Researchers*. John Wiley & Sons, Inc. New Jersey.
- Choi WH, Lee S and Cho S (2018). Microsatellite Alterations and Protein Expression of 5 Major Tumor Suppressor Genes in Gastric Adenocarcinomas. *Transl Oncol* 11(1): 43-55.
- Choi YY, Bae JM, An JY, Kwon IG, Cho I, Shin HB, Eiji T, Aburahmah M, Kim HI, Cheong JH, Hyung WJ and Noh SH (2014). Is microsatellite instability a prognostic marker in gastric cancer? A systematic review with meta-analysis. *J Surg Oncol* 110(2): 129-135.
- Choi YY, Kim H, Shin SJ, Kim HY, Lee J, Yang HK, Kim WH, Kim YW, Kook MC, Park YK, Kim HH, Lee HS, Lee KH, Gu MJ, Choi SH, Hong S, Kim JW, Hyung WJ, Noh SH and Cheong JH (2019). Microsatellite Instability and Programmed Cell Death-Ligand 1 Expression in Stage II/III Gastric Cancer: Post Hoc Analysis of the CLASSIC Randomized Controlled study. *Ann Surg* 270(2): 309-316.
- Cinti C, Vindigni C, Zamparelli A, La Sala D, Epistolato MC, Marrelli D, Cevenini G and Tosi P (2008). Activated Akt as an indicator of prognosis in gastric cancer. *Virchows Arch* 453(5): 449-455.
- Cisło M, Filip AA, Arnold Offerhaus GJ, Ciseł B, Rawicz-Pruszyński K, Skierucha M and Polkowski WP (2018). Distinct molecular subtypes of gastric cancer: from Laurén to molecular pathology. *Oncotarget* 9(27): 19427-19442.
- Clevers H and Nusse R (2012). Wnt/beta-catenin signaling and disease. *Cell* 149(6): 1192-1205.
- Collins K, Jacks T and Pavletich NP (1997). The cell cycle and cancer. *Proc Natl Acad Sci USA* 94(7): 2776-2778.
- Correa P (1992). Human gastric carcinogenesis: a multistep and multifactorial process--First American Cancer Society Award Lecture on Cancer Epidemiology and Prevention. *Cancer Res* 52(24): 6735-6740.
- Cristescu R, Lee J, Nebozhyn M, Kim KM, Ting JC, Wong SS, Liu J, Yue YG, Wang J, Yu K, Ye XS, Do IG, Liu S, Gong L, Fu J, Jin JG, Choi MG, Sohn TS, Lee JH, Bae JM, Kim ST, Park SH, Sohn I, Jung SH, Tan P, Chen R, Hardwick J, Kang WK, Ayers M, Hongyue D, Reinhard C, Loboda A, Kim S and Aggarwal A (2015). Molecular analysis of gastric cancer identifies subtypes associated with distinct clinical outcomes. *Nat Med* 21(5): 449-456.
- Cunningham D, Allum WH, Stenning SP, Thompson JN, Van de Velde CJ, Nicolson M, Scarffe JH, Lofts FJ, Falk SJ, Iveson TJ, Smith DB, Langley RE, Verma M, Weeden S, Chua YJ and Participants MT (2006). Perioperative chemotherapy versus surgery alone for resectable gastroesophageal cancer. *N Engl J Med* 355(1): 11-20.
- Dang HZ, Yu Y and Jiao SC (2012). Prognosis of HER2 over-expressing gastric cancer patients with liver metastasis. *World J Gastroenterol* 18(19): 2402-2407.
- Deng N, Goh LK, Wang H, Das K, Tao J, Tan IB, Zhang S, Lee M, Wu J, Lim KH, Lei Z, Goh G, Lim QY, Tan AL, Sin Poh DY, Riahi S, Bell S, Shi MM, Linnartz R, Zhu F, Yeoh KG, Toh HC, Yong WP, Cheong HC, Rha SY, Boussioutas A, Grabsch H, Rozen S and Tan P (2012). A comprehensive survey of genomic alterations in gastric cancer reveals systematic patterns of molecular exclusivity and co-occurrence among distinct therapeutic targets. *Gut* 61(5): 673-684.
- Diaz-Cano SJ and Brady SP (1997). DNA extraction from formalin-fixed, paraffin-embedded tissues: protein digestion as a limiting step for retrieval of high-quality DNA. *Diagn Mol Pathol* 6(6): 342-346.

- Dib C, Faure S, Fizames C, Samson D, Drouot N, Vignal A, Millasseau P, Marc S, Hazan J, Seboun E, Lathrop M, Gyapay G, Morissette J and Weissenbach J (1996). A comprehensive genetic map of the human genome based on 5,264 microsatellites. *Nature* 380(6570): 152-154.
- Dulak AM, Stojanov P, Peng S, Lawrence MS, Fox C, Stewart C, Bandla S, Imamura Y, Schumacher SE, Shefler E, McKenna A, Carter SL, Cibulskis K, Sivachenko A, Saksena G, Voet D, Ramos AH, Auclair D, Thompson K, Sougnez C, Onofrio RC, Guiducci C, Beroukhir R, Zhou Z, Lin L, Lin J, Reddy R, Chang A, Landrenau R, Pennathur A, Ogino S, Luketich JD, Golub TR, Gabriel SB, Lander ES, Beer DG, Godfrey TE, Getz G and Bass AJ (2013). Exome and whole-genome sequencing of esophageal adenocarcinoma identifies recurrent driver events and mutational complexity. *Nat Genet* 45(5): 478-486.
- Efron B and Tibshirani RJ (1994). An Introduction to the Bootstrap. CRC Press LLC Boca Raton, Florida.
- Endris V, Penzel R, Warth A, Muckenhuber A, Schirmacher P, Stenzinger A and Weichert W (2013). Molecular diagnostic profiling of lung cancer specimens with a semiconductor-based massive parallel sequencing approach: feasibility, costs, and performance compared with conventional sequencing. *J Mol Diagn* 15(6): 765-775.
- Erill N, Colomer A, Calvo M, Vidal A, Román R, Verdú M, Cordón-Cardó C and Puig X (2005). A novel multiplexing, polymerase chain reaction-based assay for the analysis of chromosome 18q status in colorectal cancer. *J Mol Diagn* 7(4): 478-485.
- Fan B, Dachrut S, Coral H, Yuen ST, Chu KM, Law S, Zhang L, Ji J, Leung SY and Chen X (2012). Integration of DNA copy number alterations and transcriptional expression analysis in human gastric cancer. *PLoS One* 7(4): e29824.
- Fang WL, Chang SC, Lan YT, Huang KH, Chen JH, Lo SS, Hsieh MC, Li AF, Wu CW and Chiou SH (2012). Microsatellite instability is associated with a better prognosis for gastric cancer patients after curative surgery. *World J Surg* 36(9): 2131-2138.
- Fenoglio-Preiser CM, Wang J, Stemmermann GN and Noffsinger A (2003). TP53 and gastric carcinoma: a review. *Hum Mutat* 21(3): 258-270.
- Fisher KE, Zhang L, Wang J, Smith GH, Newman S, Schneider TM, Pillai RN, Kudchadkar RR, Owonikoko TK, Ramalingam SS, Lawson DH, Delman KA, El-Rayes BF, Wilson MM, Sullivan HC, Morrison AS, Balci S, Adsay NV, Gal AA, Sica GL, Saxe DF, Mann KP, Hill CE, Khuri FR and Rossi MR (2016). Clinical Validation and Implementation of a Targeted Next-Generation Sequencing Assay to Detect Somatic Variants in Non-Small Cell Lung, Melanoma, and Gastrointestinal Malignancies. *J Mol Diagn* 18(2): 299-315.
- Fitzgerald RC, Hardwick R, Huntsman D, Carneiro F, Guilford P, Blair V, Chung DC, Norton J, Ragnath K, Van Krieken JH, Dwerryhouse S, Caldas C and International Gastric Cancer Linkage C (2010). Hereditary diffuse gastric cancer: updated consensus guidelines for clinical management and directions for future research. *J Med Genet* 47(7): 436-444.
- Fontana E and Smyth EC (2016). Novel targets in the treatment of advanced gastric cancer: a perspective review. *Ther Adv Med Oncol* 8(2): 113-125.
- Forman D, Newell DG, Fullerton F, Yarnell JW, Stacey AR, Wald N and Sitas F (1991). Association between infection with *Helicobacter pylori* and risk of gastric cancer: evidence from a prospective investigation. *BMJ* 302(6788): 1302-1305.
- Foschi R, Lucenteforte E, Bosetti C, Bertuccio P, Tavani A, La Vecchia C and Negri E (2008). Family history of cancer and stomach cancer risk. *Int J Cancer* 123(6): 1429-1432.
- Foster I (2008). Cancer: A cell cycle defect. *Radiography* 14(2): 144-149.
- Foster JM, Oumie A, Togneri FS, Vasques FR, Hau D, Taylor M, Tinkler-Hundal E, Southward K, Medlow P, McGreeghan-Crosby K, Halfpenny I, McMullan DJ, Quirke P, Keating KE, Griffiths M, Spink KG and Brew F (2015). Cross-laboratory validation of the OncoScan® FFPE Assay, a multiplex tool for whole genome tumour profiling. *BMC Med Genomics* 8: 5.
- Fried M, Manns MP and Rogler G (2013). Magen-Darm-Trakt. Springer-Verlag Berlin Heidelberg.
- Frigerio S, Padberg BC, Strebel RT, Lenggenhager DM, Messthaler A, Abdou MT, Moch H and Zimmermann DR (2007). Improved detection of bladder carcinoma cells in voided urine by standardized microsatellite analysis. *Int J Cancer* 121(2): 329-338.
- Gao C, Su Y, Koeman J, Haak E, Dykema K, Essenberg C, Hudson E, Petillo D, Khoo SK and Vande Woude GF (2016). Chromosome instability drives phenotypic switching to metastasis. *Proc Natl Acad Sci USA* 113(51): 14793-14798.

- Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, Sun Y, Jacobsen A, Sinha R, Larsson E, Cerami E, Sander C and Schultz N (2013). Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal* 6(269): p11.
- Garrido-Cardenas JA, Garcia-Maroto F, Alvarez-Bermejo JA and Manzano-Agugliaro F (2017). DNA Sequencing Sensors: An Overview. *Sensors (Basel)* 17(3): 588.
- Geigl JB, Obenauf AC, Schwarzbraun T and Speicher MR (2008). Defining 'chromosomal instability'. *Trends Genet* 24(2): 64-69.
- Genitsch V, Novotny A, Seiler CA, Kröll D, Walch A and Langer R (2015). Epstein-barr virus in gastroesophageal adenocarcinomas - single center experiences in the context of current literature. *Front Oncol* 5: 73-73.
- Goldstein J, Tran B, Ensor J, Gibbs P, Wong HL, Wong SF, Vilar E, Tie J, Broaddus R, Kopetz S, Desai J and Overman MJ (2014). Multicenter retrospective analysis of metastatic colorectal cancer (CRC) with high-level microsatellite instability (MSI-H). *Ann Oncol* 25(5): 1032-1038.
- Goode EF and Smyth EC (2016). Immunotherapy for Gastroesophageal Cancer. *J Clin Med* 5(10).
- Gravalos C and Jimeno A (2008). HER2 in gastric cancer: a new prognostic factor and a novel therapeutic target. *Ann Oncol* 19(9): 1523-1529.
- Gronroos E and Lopez-Garcia C (2018). Tolerance of Chromosomal Instability in Cancer: Mechanisms and Therapeutic Opportunities. *Cancer Res* 78(23): 6529-6535.
- Gurzu S, Kadar Z, Bara T, Bara T, Jr., Tamasi A, Azamfirei L and Jung I (2015). Mixed adenoneuroendocrine carcinoma of gastrointestinal tract: report of two cases. *World J Gastroenterol* 21(4): 1329-1333.
- Haag GM, Czink E, Ahadova A, Schmidt T, Sisic L, Blank S, Heger U, Apostolidis L, Berger AK, Springfield C, Lasitschka F, Jager D, von Knebel Doeberitz M and Kloor M (2019). Prognostic significance of microsatellite-instability in gastric and gastroesophageal junction cancer patients undergoing neoadjuvant chemotherapy. *Int J Cancer* 144(7): 1697-1703.
- Hardiman KM, Ulintz PJ, Kuick RD, Hovelson DH, Gates CM, Bhasi A, Rodrigues Grant A, Liu J, Cani AK, Greenson JK, Tomlins SA and Fearon ER (2016). Intra-tumor genetic heterogeneity in rectal cancer. *Lab Invest* 96(1): 4-15.
- Harvey AJ (2019). Overview of Cell Signaling Pathways in Cancer. Predictive Biomarkers in Oncology: Applications in Precision Medicine. Sunil Badve and George Louis Kumar. Cham, Springer International Publishing: p. 167-182.
- Hatch SB, Lightfoot HM, Jr., Garwacki CP, Moore DT, Calvo BF, Woosley JT, Sciarrotta J, Funkhouser WK and Farber RA (2005). Microsatellite instability testing in colorectal carcinoma: choice of markers affects sensitivity of detection of mismatch repair-deficient tumors. *Clin Cancer Res* 11(6): 2180-2187.
- Haugen AC, Goel A, Yamada K, Marra G, Nguyen TP, Nagasaka T, Kanazawa S, Koike J, Kikuchi Y, Zhong X, Arita M, Shibuya K, Oshimura M, Hemmi H, Boland CR and Koi M (2008). Genetic instability caused by loss of MutS homologue 3 in human colorectal cancer. *Cancer Res* 68(20): 8465-8472.
- Hause RJ, Pritchard CC, Shendure J and Salipante SJ (2016). Classification and characterization of microsatellite instability across 18 cancer types. *Nat Med* 22(11): 1342-1350.
- Hewitt LC, Inam IZ, Saito Y, Yoshikawa T, Quaas A, Hoelscher A, Bollschweiler E, Fazzi GE, Melotte V, Langley RE, Nankivell M, Cunningham D, Allum W, Hutchins GG and Grabsch HI (2018). Epstein-Barr virus and mismatch repair deficiency status differ between oesophageal and gastric cancer: A large multi-centre study. *Eur J Cancer* 94: 104-114.
- Hile SE, Shabashev S and Eckert KA (2013). Tumor-specific microsatellite instability: do distinct mechanisms underlie the MSI-L and EMAS phenotypes? *Mutat Res* 743-744: 67-77.
- Holleley CE and Geerts PG (2009). Multiplex Manager 1.0: a cross-platform computer program that plans and optimizes multiplex PCR. *BioTechniques* 46(7): 511-517.
- Hu B, El Hajj N, Sittler S, Lammert N, Barnes R and Meloni-Ehrig A (2012). Gastric cancer: Classification, histology and application of molecular pathology. *J Gastrointest Oncol* 3(3): 251-261.
- Huang SC, Ng KF, Yeh TS, Cheng CT, Lin JS, Liu YJ, Chuang HC and Chen TC (2019). Subtraction of Epstein-Barr virus and microsatellite instability genotypes from the Lauren histotypes: Combined molecular and histologic subtyping with clinicopathological and prognostic significance validated in a cohort of 1,248 cases. *Int J Cancer* 145(12): 3218-3230.

- Huber M, Pavlova B, Muhlberger H, Hollaus P and Lintner F (2002). Detection of the Epstein-Barr virus in primary adenocarcinoma of the lung with Signet-ring cells. *Virchows Arch* 441(1): 25-30.
- Hudson TJ, Anderson W, Aretz A, Barker AD, Bell C, Bernabé RR, Bhan MK, Calvo F, Eerola I, Gerhard DS, Guttmacher A, Guyer M, Hemsley FM, Jennings JL, Kerr D, Klatt P, Kolar P, Kusuda J, Lane DP, Laplace F, Lu Y, Nettekoven G, Ozenberger B, Peterson J, Rao TS, Remalec J, Schafer AJ, Shibata T, Stratton MR, Vockley JG, Watanabe K, Yang H, Yuen MMF, Knoppers BM, Bobrow M, Cambon-Thomsen A, Dressler LG, Dyke SOM, Joly Y, Kato K, Kennedy KL, Nicolás P, Parker MJ, Rial-Sebbag E, Romeo-Casabona CM, Shaw KM, Wallace S, Wiesner GL, Zeps N, Lichter P, et al. (2010). International network of cancer genome projects. *Nature* 464(7291): 993-998.
- Huntsman DG, Carneiro F, Lewis FR, MacLeod PM, Hayashi A, Monaghan KG, Maung R, Seruca R, Jackson CE and Caldas C (2001). Early gastric cancer in young, asymptomatic carriers of germline E-cadherin mutations. *N Engl J Med* 344(25): 1904-1909.
- Hynes NE and Lane HA (2005). ERBB receptors and cancer: the complexity of targeted inhibitors. *Nat Rev Cancer* 5(5): 341-354.
- Iizasa H, Nanbo A, Nishikawa J, Jinushi M and Yoshiyama H (2012). Epstein-Barr Virus (EBV)-associated gastric carcinoma. *Viruses* 4(12): 3420-3439.
- Jamal-Hanjani M, A'Hern R, Birkbak NJ, Gorman P, Gronroos E, Ngang S, Nicola P, Rahman L, Thanopoulou E, Kelly G, Ellis P, Barrett-Lee P, Johnston SR, Bliss J, Roylance R and Swanton C (2015). Extreme chromosomal instability forecasts improved outcome in ER-negative breast cancer: a prospective validation cohort study from the TACT trial. *Ann Oncol* 26(7): 1340-1346.
- Japanese Gastric Cancer Association (2017). Japanese gastric cancer treatment guidelines 2014 (ver. 4). *Gastric cancer : official journal of the International Gastric Cancer Association and the Japanese Gastric Cancer Association* 20(1): 1-19.
- Jefford CE and Irminger-Finger I (2006). Mechanisms of chromosome instability in cancers. *Crit Rev Oncol Hematol* 59(1): 1-14.
- Jiao YF, Sugai T, Habano W, Uesugi N, Takagane A and Nakamura S (2006). Clinicopathological significance of loss of heterozygosity in intestinal- and solid-type gastric carcinomas: a comprehensive study using the crypt isolation technique. *Mod Pathol* 19(4): 548-555.
- Jin DH, Park SE, Lee J, Kim KM, Kim S, Kim DH and Park J (2015). Copy Number Gains at 8q24 and 20q11-q13 in Gastric Cancer Are More Common in Intestinal-Type than Diffuse-Type. *PLoS One* 10(9): e0137657.
- Juarez-Salcedo LM, Sokol L, Chavez JC and Dalia S (2018). Primary Gastric Lymphoma, Epidemiology, Clinical Diagnosis, and Treatment. *Cancer Control* 25(1): 1073274818778256.
- Jung H-S, Lefferts JA and Tsongalis GJ (2017). Utilization of the oncoscan microarray assay in cancer diagnostics. *Applied Cancer Research* 37(1): 1.
- Kakiuchi M, Nishizawa T, Ueda H, Gotoh K, Tanaka A, Hayashi A, Yamamoto S, Tatsuno K, Katoh H, Watanabe Y, Ichimura T, Ushiku T, Funahashi S, Tateishi K, Wada I, Shimizu N, Nomura S, Koike K, Seto Y, Fukayama M, Aburatani H and Ishikawa S (2014). Recurrent gain-of-function mutations of RHOA in diffuse-type gastric carcinoma. *Nat Genet* 46(6): 583-587.
- Kandoth C, McLellan MD, Vandin F, Ye K, Niu B, Lu C, Xie M, Zhang Q, McMichael JF, Wyczalkowski MA, Leiserson MDM, Miller CA, Welch JS, Walter MJ, Wendl MC, Ley TJ, Wilson RK, Raphael BJ and Ding L (2013). Mutational landscape and significance across 12 major cancer types. *Nature* 502(7471): 333-339.
- Katona BW and Rustgi AK (2017). Gastric Cancer Genomics: Advances and Future Directions. *Cell Mol Gastroenterol Hepatol* 3(2): 211-217.
- Kaurah P, MacMillan A, Boyd N, Senz J, De Luca A, Chun N, Suriano G, Zaor S, Van Manen L, Gilpin C, Nikkel S, Connolly-Wilson M, Weissman S, Rubinstein WS, Sebold C, Greenstein R, Stroop J, Yim D, Panzini B, McKinnon W, Greenblatt M, Wirtzfeld D, Fontaine D, Coit D, Yoon S, Chung D, Lauwers G, Pizzuti A, Vaccaro C, Redal MA, Oliveira C, Tischkowitz M, Olschwang S, Gallinger S, Lynch H, Green J, Ford J, Pharoah P, Fernandez B and Huntsman D (2007). Founder and recurrent CDH1 mutations in families with hereditary diffuse gastric cancer. *JAMA* 297(21): 2360-2372.
- Kelderman S, Schumacher TN and Kvistborg P (2015). Mismatch Repair-Deficient Cancers Are Targets for Anti-PD-1 Therapy. *Cancer Cell* 28(1): 11-13.

- Keller G, Rotter M, Vogelsang H, Bischoff P, Becker KF, Mueller J, Brauch H, Siewert JR and Hofler H (1995). Microsatellite instability in adenocarcinomas of the upper gastrointestinal tract. Relation to clinicopathological data and family history. *Am J Pathol* 147(3): 593-600.
- Keller G (2002). Hereditary aspects of gastric cancer. *Pathologica* 94(5): 229-233.
- Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM and Haussler D (2002). The human genome browser at UCSC. *Genome Res* 12(6): 996-1006.
- Kim JG, Shin S and Park J (2017). Comparison between mononucleotide and dinucleotide marker panels in gastric cancer with loss of hMLH1 or hMSH2 expression. *Int J Biol Markers* 32(3): e352-e356.
- Kim ST, Cristescu R, Bass AJ, Kim KM, Odegaard JI, Kim K, Liu XQ, Sher X, Jung H, Lee M, Lee S, Park SH, Park JO, Park YS, Lim HY, Lee H, Choi M, Talasaz A, Kang PS, Cheng J, Loboda A, Lee J and Kang WK (2018). Comprehensive molecular characterization of clinical responses to PD-1 inhibition in metastatic gastric cancer. *Nat Med* 24(9): 1449-1458.
- Kim SY, Choi YY, An JY, Shin HB, Jo A, Choi H, Seo SH, Bang HJ, Cheong JH, Hyung WJ and Noh SH (2015). The benefit of microsatellite instability is attenuated by chemotherapy in stage II and stage III gastric cancer: Results from a large cohort with subgroup analyses. *Int J Cancer* 137(4): 819-825.
- Kim YH, Min BH, Choi HK, Kim SJ, Kim KM, Kim JY, Chang DK, Son HJ, Rhee PL, Kim JJ, Rhee JC and Chun HK (2009). Sporadic colorectal carcinomas with low-level microsatellite instability in Korea: do they form a distinct subgroup with distinguished clinicopathological features? *J Surg Oncol* 99(6): 351-355.
- Knudson AG, Jr. (1971). Mutation and cancer: statistical study of retinoblastoma. *Proc Natl Acad Sci USA* 68(4): 820-823.
- Kohlross M (2016). Etablierung eines Mikrosatelliten-basierten Tests zur Detektion chromosomaler Instabilität in Magenkarzinomen. Master thesis at the Institute of Pathology, Technische Universität München.
- Kohlross M, Reiche M, Jesinghaus M, Grosser B, Slotta-Huspenina J, Hapfelmeier A, Bauer L, Novotny A, Weichert W and Keller G (2018). A microsatellite based multiplex PCR method for the detection of chromosomal instability in gastric cancer. *Sci Rep* 8: 12551.
- Kohlross M, Grosser B, Krenauer M, Slotta-Huspenina J, Jesinghaus M, Blank S, Novotny A, Reiche M, Schmidt T, Ismani L, Hapfelmeier A, Mathias D, Meyer P, Gaida MM, Bauer L, Ott K, Weichert W and Keller G (2019). Prognostic implication of molecular subtypes and response to neoadjuvant chemotherapy in 760 gastric carcinomas: role of Epstein-Barr virus infection and high- and low-microsatellite instability. *J Pathol Clin Res* 5(4): 227-239.
- Kohonen-Corish MR, Daniel JJ, Chan C, Lin BP, Kwun SY, Dent OF, Dhillon VS, Trent RJ, Chapuis PH and Bokey EL (2005). Low microsatellite instability is associated with poor prognosis in stage C colon cancer. *J Clin Oncol* 23(10): 2318-2324.
- Koi M, Tseng-Rogenski SS and Carethers JM (2018). Inflammation-associated microsatellite alterations: Mechanisms and significance in the prognosis of patients with colorectal cancer. *World J Gastrointest Oncol* 10(1): 1-14.
- Kono S and Hirohata T (1996). Nutrition and stomach cancer. *Cancer Causes Control* 7(1): 41-55.
- Krebs in Deutschland für 2015/2016. 12. Ausgabe. Robert Koch-Institut (Hrsg) und die Gesellschaft der epidemiologischen Krebsregister in Deutschland e.V. (Hrsg). Berlin, 2019.
- Krenauer M (2018). Molekulare Klassifikation von Magenkarzinomen und ihre prädiktive und prognostische Relevanz für eine Platin/5FU basierte neoadjuvante Chemotherapie. Master thesis at the Institute of Pathology, Technische Universität München.
- Kuboki Y, Yamashita S, Niwa T, Ushijima T, Nagatsuma A, Kuwata T, Yoshino T, Doi T, Ochiai A and Ohtsu A (2016). Comprehensive analyses using next-generation sequencing and immunohistochemistry enable precise treatment in advanced gastric cancer. *Ann Oncol* 27(1): 127-133.
- Kugler P (2017). Der menschliche Körper - Anatomie Physiologie Pathologie. Elsevier München.
- Ladeiras-Lopes R, Pereira AK, Nogueira A, Pinheiro-Torres T, Pinto I, Santos-Pereira R and Lunet N (2008). Smoking and gastric cancer: systematic review and meta-analysis of cohort studies. *Cancer Causes Control* 19(7): 689-701.
- Lagergren J, Bergstrom R and Nyren O (1999). Association between body mass and adenocarcinoma of the esophagus and gastric cardia. *Ann Intern Med* 130(11): 883-890.
- Lane DP (1992). Cancer. p53, guardian of the genome. *Nature* 358(6381): 15-16.

- Langer R and Becker K (2018). Tumor regression grading of gastrointestinal cancers after neoadjuvant therapy. *Virchows Arch* 472(2): 175-186.
- Lauren P (1965). The two histological main types of gastric carcinoma: diffuse and so-called intestinal-type carcinoma. An attempt at a histo-clinical classification. *Acta Pathol Microbiol Scand* 64: 31-49.
- Le DT, Uram JN, Wang H, Bartlett BR, Kemberling H, Eyring AD, Skora AD, Luber BS, Azad NS, Laheru D, Biedrzycki B, Donehower RC, Zaheer A, Fisher GA, Crocenzi TS, Lee JJ, Duffy SM, Goldberg RM, de la Chapelle A, Koshiji M, Bhajee F, Huebner T, Hruban RH, Wood LD, Cuka N, Pardoll DM, Papadopoulos N, Kinzler KW, Zhou S, Cornish TC, Taube JM, Anders RA, Eshleman JR, Vogelstein B and Diaz LA (2015). PD-1 Blockade in Tumors with Mismatch-Repair Deficiency. *New Engl J Med* 372(26): 2509-2520.
- Le DT, Durham JN, Smith KN, Wang H, Bartlett BR, Aulakh LK, Lu S, Kemberling H, Wilt C, Luber BS, Wong F, Azad NS, Rucki AA, Laheru D, Donehower R, Zaheer A, Fisher GA, Crocenzi TS, Lee JJ, Greten TF, Duffy AG, Ciombor KK, Eyring AD, Lam BH, Joe A, Kang SP, Holdhoff M, Danilova L, Cope L, Meyer C, Zhou S, Goldberg RM, Armstrong DK, Bever KM, Fader AN, Taube J, Housseau F, Spetzler D, Xiao N, Pardoll DM, Papadopoulos N, Kinzler KW, Eshleman JR, Vogelstein B, Anders RA and Diaz LA, Jr. (2017). Mismatch repair deficiency predicts response of solid tumors to PD-1 blockade. *Science* 357(6349): 409-413.
- Lee JH, Kim SH, Han SH, An JS, Lee ES and Kim YS (2009). Clinicopathological and molecular characteristics of Epstein-Barr virus-associated gastric carcinoma: a meta-analysis. *J Gastroenterol Hepatol* 24(3): 354-365.
- Lee SY, Kim DW, Lee HS, Ihn MH, Oh HK, Min BS, Kim WR, Huh JW, Yun JA, Lee KY, Kim NK, Lee WY, Kim HC and Kang SB (2015). Low-Level Microsatellite Instability as a Potential Prognostic Factor in Sporadic Colorectal Cancer. *Medicine (Baltimore)* 94(50): e2260.
- Leitlinienprogramm Onkologie (2019). S3-Leitlinie Magenkarzinom. Diagnostik und Therapie der Adenokarzinome des Magens und ösophagogastralen Übergangs. Langversion 2.0. AWMF-Registernummer: 032/009OL.
- Lengauer C, Kinzler KW and Vogelstein B (1998). Genetic instabilities in human cancers. *Nature* 396(6712): 643-649.
- Lepage CC, Morden CR, Palmer MCL, Nachtigal MW and McManus KJ (2019). Detecting Chromosome Instability in Cancer: Approaches to Resolve Cell-to-Cell Heterogeneity. *Cancers (Basel)* 11(2).
- Li X, Wu WK, Xing R, Wong SH, Liu Y, Fang X, Zhang Y, Wang M, Wang J, Li L, Zhou Y, Tang S, Peng S, Qiu K, Chen L, Chen K, Yang H, Zhang W, Chan MT, Lu Y, Sung JJ and Yu J (2016). Distinct Subtypes of Gastric Cancer Defined by Molecular Characterization Include Novel Mutational Signatures with Prognostic Capability. *Cancer Res* 76(7): 1724-1732.
- Lin XJ, Wang CP, Liu XD, Yan KK, Li S, Bao HH, Zhao LY and Liu X (2014). Body mass index and risk of gastric cancer: a meta-analysis. *Jpn J Clin Oncol* 44(9): 783-791.
- Lin Y, Wu Z, Guo W and Li J (2015). Gene mutations in gastric cancer: a review of recent next-generation sequencing studies. *Tumour Biol* 36(10): 7385-7394.
- Liu Y, Sethi NS, Hinoue T, Schneider BG, Cherniack AD, Sanchez-Vega F, Seoane JA, Farshidfar F, Bowlby R, Islam M, Kim J, Chatila W, Akbani R, Kanchi RS, Rabkin CS, Willis JE, Wang KK, McCall SJ, Mishra L, Ojesina AI, Bullman S, Peadarallu CS, Lazar AJ, Sakai R, Thorsson V, Bass AJ and Laird PW (2018). Comparative Molecular Analysis of Gastrointestinal Adenocarcinomas. *Cancer Cell* 33(4): 721-735.
- Loeb LA (2001). A mutator phenotype in cancer. *Cancer Res* 61(8): 3230-3239.
- Lorenzen S, Thuss-Patience P, Al-Batran SE, Lordick F, Haller B, Schuster T, Pauligk C, Luley K, Bichev D, Schumacher G and Homann N (2013). Impact of pathologic complete response on disease-free survival in patients with esophagogastric adenocarcinoma receiving preoperative docetaxel-based chemotherapy. *Ann Oncol* 24(8): 2068-2073.
- Luo J, Chen X-Q and Li P (2019). The Role of TGF- $\beta$  and Its Receptors in Gastrointestinal Cancers. *Transl Oncol* 12(3): 475-484.
- Ma C, Patel K, Singhi AD, Ren B, Zhu B, Shaikh F and Sun W (2016). Programmed Death-Ligand 1 Expression Is Common in Gastric Cancer Associated With Epstein-Barr Virus or Microsatellite Instability. *Am J Surg Pathol* 40(11): 1496-1506.

- Ma K, Baloch Z, He T-T and Xia X (2017). Alcohol Consumption and Gastric Cancer Risk: A Meta-Analysis. *Med Sci Monit* 23: 238-246.
- Maleki SS and Röcken C (2017). Chromosomal Instability in Gastric Cancer Biology. *Neoplasia* 19(5): 412-420.
- Malkin D, Li FP, Strong LC, Fraumeni JF, Jr., Nelson CE, Kim DH, Kassel J, Gryka MA, Bischoff FZ, Tainsky MA and et al. (1990). Germ line p53 mutations in a familial syndrome of breast cancer, sarcomas, and other neoplasms. *Science* 250(4985): 1233-1238.
- Marqués-Lespier JM, González-Pons M and Cruz-Correa M (2016). Current Perspectives on Gastric Cancer. *Gastroenterol Clin North Am* 45(3): 413-428.
- Marrelli D, Polom K, Pascale V, Vindigni C, Piagnerelli R, De Franco L, Ferrara F, Roviello G, Garosi L, Petrioli R and Roviello F (2016). Strong Prognostic Value of Microsatellite Instability in Intestinal Type Non-cardia Gastric Cancer. *Ann Surg Oncol* 23(3): 943-950.
- Massarrat S and Stolte M (2014). Development of gastric cancer and its prevention. *Arch Iran Med* 17(7): 514-520.
- Matsuoka T and Yashiro M (2014). The Role of PI3K/Akt/mTOR Signaling in Gastric Carcinoma. *Cancers (Basel)* 6(3): 1441-1463.
- McClelland SE (2017). Role of chromosomal instability in cancer progression. *Endocr Relat Cancer* 24(9): T23-T31.
- Merriman B and Rothberg JM (2012). Progress in ion torrent semiconductor chip based sequencing. *Electrophoresis* 33(23): 3397-3417.
- Meyer HJ and Wilke H (2011). Treatment strategies in gastric cancer. *Dtsch Arztebl Int* 108(41): 698-705.
- Mitchell RJ, Farrington SM, Dunlop MG and Campbell H (2002). Mismatch repair genes hMLH1 and hMSH2 and colorectal cancer: a HuGE review. *Am J Epidemiol* 156(10): 885-902.
- Moehler M, Al-Batran S-E, Andus T and ANthuber M (2011). S3-Leitlinie „Magenkarzinom“ – Diagnostik und Therapie der Adenokarzinome des Magens und ösophagogastralen Übergangs (AWMF-Regist.-Nr. 032-009-OL). *Z Gastroenterol* 49: 461–531.
- Moehler M, Baltin CT, Ebert M, Fischbach W, Gockel I, Grenacher L, Holscher AH, Lordick F, Malfertheiner P, Messmann H, Meyer HJ, Palmqvist A, Rocken C, Schuhmacher C, Stahl M, Stuschke M, Vieth M, Wittekind C, Wagner D and Monig SP (2015). International comparison of the German evidence-based S3-guidelines on the diagnosis and multimodal treatment of early and locally advanced gastric cancer, including adenocarcinoma of the lower esophagus. *Gastric Cancer* 18(3): 550-563.
- Moghbeli M, Makhdoumi Y, Soltani Delgosha M, Aarabi A, Dadkhah E, Memar B, Abdollahi A and Abbaszadegan MR (2019). ErbB1 and ErbB3 co-over expression as a prognostic factor in gastric cancer. *Biol Res* 52(1): 2-2.
- Mukaisho K-I, Nakayama T, Hagiwara T, Hattori T and Sugihara H (2015). Two distinct etiologies of gastric cardia adenocarcinoma: interactions among pH, Helicobacter pylori, and bile acids. *Front Microbiol* 6: 412-412.
- Murphy G, Pfeiffer R, Camargo MC and Rabkin CS (2009). Meta-analysis shows that prevalence of Epstein-Barr virus-positive gastric cancer differs based on sex and anatomic location. *Gastroenterology* 137(3): 824-833.
- Napieralski R, Ott K, Kremer M, Becker K, Boulesteix AL, Lordick F, Siewert JR, Hofler H and Keller G (2007). Methylation of tumor-related genes in neoadjuvant-treated gastric cancer: relation to therapy response and clinicopathologic and molecular features. *Clin Cancer Res* 13(17): 5095-5102.
- Nazemalhosseini Mojarad E, Kuppen PJ, Aghdaei HA and Zali MR (2013). The CpG island methylator phenotype (CIMP) in colorectal cancer. *Gastroenterol Hepatol Bed Bench* 6(3): 120-128.
- Nazemalhosseini Mojarad E, Kashfi SMH, Mirtalebi H, Taleghani MY, Azimzadeh P, Savabkar S, Pourhoseingholi MA, Jalaeikhoo H, Asadzadeh Aghdaei H, Kuppen PJK and Zali MR (2016). Low Level of Microsatellite Instability Correlates with Poor Clinical Prognosis in Stage II Colorectal Cancer Patients. *J Oncol* 2016: 2196703.
- Nishikawa J, Yoshiyama H, Iizasa H, Kanehiro Y, Nakamura M, Nishimura J, Saito M, Okamoto T, Sakai K, Suehiro Y, Yamasaki T, Oga A, Yanai H and Sakaida I (2014). Epstein-barr virus in gastric carcinoma. *Cancers (Basel)* 6(4): 2259-2274.

- Nishino Y, Inoue M, Tsuji I, Wakai K, Nagata C, Mizoue T, Tanaka K and Tsugane S (2006). Tobacco smoking and gastric cancer risk: an evaluation based on a systematic review of epidemiologic evidence among the Japanese population. *Jpn J Clin Oncol* 36(12): 800-807.
- O'Keefe C, McDevitt MA and Maciejewski JP (2010). Copy neutral loss of heterozygosity: a novel chromosomal lesion in myeloid malignancies. *Blood* 115(14): 2731-2739.
- Oliveira C, Senz J, Kaurah P, Pinheiro H, Sanges R, Haegert A, Corso G, Schouten J, Fitzgerald R, Vogelsang H, Keller G, Dwerryhouse S, Grimmer D, Chin SF, Yang HK, Jackson CE, Seruca R, Roviello F, Stupka E, Caldas C and Huntsman D (2009). Germline CDH1 deletions in hereditary diffuse gastric cancer families. *Hum Mol Genet* 18(9): 1545-1555.
- Oliveira C, Pinheiro H, Figueiredo J, Seruca R and Carneiro F (2015). Familial gastric cancer: genetic susceptibility, pathology, and implications for management. *Lancet Oncol* 16(2): e60-70.
- Olivier M, Hollstein M and Hainaut P (2010). TP53 mutations in human cancers: origins, consequences, and clinical use. *Cold Spring Harb Perspect Biol* 2(1): a001008.
- Ott K, Vogelsang H, Mueller J, Becker K, Muller M, Fink U, Siewert JR, Hofler H and Keller G (2003). Chromosomal instability rather than p53 mutation is associated with response to neoadjuvant cisplatin-based chemotherapy in gastric carcinoma. *Clin Cancer Res* 9(6): 2307-2315.
- Ott K, Bader FG, Lordick F, Feith M, Bartels H and Siewert JR (2009). Surgical factors influence the outcome after Ivor-Lewis esophagectomy with intrathoracic anastomosis for adenocarcinoma of the esophagogastric junction: a consecutive series of 240 patients at an experienced center. *Ann Surg Oncol* 16(4): 1017-1025.
- Ow SGW, Tan KT, Yang H, Yap HL, Sapari NSB, Ong PY, Soong R and Lee SC (2019). Next Generation Sequencing Reveals Novel Mutations in Mismatch Repair Genes and Other Cancer Predisposition Genes in Asian Patients with Suspected Lynch Syndrome. *Clin Colorectal Cancer* 18(4): e324-e334.
- Pan X, Ji X, Zhang R, Zhou Z, Zhong Y, Peng W, Sun N, Xu X, Xia L, Li P, Lu J and Tu J (2018). Landscape of somatic mutations in gastric cancer assessed using next-generation sequencing analysis. *Oncol Lett* 16(4): 4863-4870.
- Park SC and Chun HJ (2013). Chemotherapy for advanced gastric cancer: review and update of current practices. *Gut Liver* 7(4): 385-393.
- Patten KT and Thibodeau GA (2016). Anatomy and Physiology, 9th edition. Elsevier
- Peleteiro B, Bastos A, Ferro A and Lunet N (2014). Prevalence of Helicobacter pylori infection worldwide: a systematic review of studies with national coverage. *Dig Dis Sci* 59(8): 1698-1709.
- Pereira CBL, Leal MF, Abdelhay E, Demachki S, Assumpcao PP, de Souza MC, Moreira-Nunes CA, Tanaka A, Smith MC and Burbano RR (2017). MYC Amplification as a Predictive Factor of Complete Pathologic Response to Docetaxel-based Neoadjuvant Chemotherapy for Breast Cancer. *Clin Breast Cancer* 17(3): 188-194.
- Pereira MA, Ramos M, Faraj SF, Dias AR, Yagi OK, Zilberstein B, Cecconello I, Alves VAF, de Mello ES and Ribeiro U, Jr. (2018). Clinicopathological and prognostic features of Epstein-Barr virus infection, microsatellite instability, and PD-L1 expression in gastric cancer. *J Surg Oncol* 117(5): 829-839.
- Petitjean A, Achatz MI, Borresen-Dale AL, Hainaut P and Olivier M (2007). TP53 mutations in human cancers: functional selection and impact on cancer prognosis and outcomes. *Oncogene* 26(15): 2157-2165.
- Pfarr N, Stenzinger A, Penzel R, Warth A, Dienemann H, Schirmacher P, Weichert W and Endris V (2016). High-throughput diagnostic profiling of clinically actionable gene fusions in lung cancer. *Genes Chromosomes Cancer* 55(1): 30-44.
- Pfarr N, Penzel R, Endris V, Lier C, Flechtenmacher C, Volckmar AL, Kirchner M, Budczies J, Leichsenring J, Herpel E, Noske A, Weichert W, Schneeweiss A, Schirmacher P, Sinn HP and Stenzinger A (2017a). Targeted next-generation sequencing enables reliable detection of HER2 (ERBB2) status in breast cancer and provides ancillary information of clinical relevance. *Genes Chromosomes Cancer* 56(4): 255-265.
- Pfarr N, Darb-Esfahani S, Leichsenring J, Taube E, Boxberg M, Braicu I, Jesinghaus M, Penzel R, Endris V, Noske A, Weichert W, Schirmacher P, Denkert C and Stenzinger A (2017b). Mutational profiles of Brenner tumors show distinctive features uncoupling urothelial carcinomas and ovarian carcinoma with transitional cell histology. *Genes Chromosomes Cancer* 56(10): 758-766.

- Pharoah PD, Guilford P and Caldas C (2001). Incidence of gastric cancer and breast cancer in CDH1 (E-cadherin) mutation carriers from hereditary diffuse gastric cancer families. *Gastroenterology* 121(6): 1348-1353.
- Plaschke J, Preussler M, Ziegler A and Schackert HK (2012). Aberrant protein expression and frequent allelic loss of MSH3 in colorectal cancer with low-level microsatellite instability. *Int J Colorectal Dis* 27(7): 911-919.
- Polom K, Marano L, Marrelli D, De Luca R, Roviello G, Savelli V, Tan P and Roviello F (2018). Meta-analysis of microsatellite instability in relation to clinicopathological characteristics and overall survival in gastric cancer. *Br J Surg* 105(3): 159-167.
- Powell EL, Leoni LM, Canto MI, Forastiere AA, Iocobuzio-Donahue CA, Wang JS, Maitra A and Montgomery E (2005). Concordant loss of MTAP and p16/CDKN2A expression in gastroesophageal carcinogenesis: evidence of homozygous deletion in esophageal noninvasive precursor lesions and therapeutic implications. *Am J Surg Pathol* 29(11): 1497-1504.
- Qinghai Z, Yanying W, Yunfang C, Xukui Z and Xiaoqiao Z (2014). Effect of interleukin-17A and interleukin-17F gene polymorphisms on the risk of gastric cancer in a Chinese population. *Gene* 537(2): 328-332.
- Ratti M, Lampis A, Hahne JC, Passalacqua R and Valeri N (2018). Microsatellite instability in gastric cancer: molecular bases, clinical perspectives, and new treatment approaches. *Cell Mol Life Sci* 75(22): 4151-4162.
- Rawla P and Barsouk A (2019). Epidemiology of gastric cancer: global trends, risk factors and prevention. *Prz Gastroenterol* 14(1): 26-38.
- Riquelme I, Saavedra K, Espinoza JA, Weber H, García P, Nervi B, Garrido M, Corvalán AH, Roa JC and Bizama C (2015). Molecular classification of gastric cancer: Towards a pathway-driven targeted therapy. *Oncotarget* 6(28): 24750-24779.
- Robinson JT, Thorvaldsdottir H, Winckler W, Guttman M, Lander ES, Getz G and Mesirov JP (2011). Integrative genomics viewer. *Nat Biotechnol* 29(1): 24-26.
- Robinson JT, Thorvaldsdottir H, Wenger AM, Zehir A and Mesirov JP (2017). Variant Review with the Integrative Genomics Viewer. *Cancer Res* 77(21): e31-e34.
- Röcken C, Behrens HM, Boger C and Kruger S (2016). Clinicopathological characteristics of RHOA mutations in a Central European gastric cancer cohort. *J Clin Pathol* 69(1): 70-75.
- Rothberg JM, Hinz W, Rearick TM, Schultz J, Mileski W, Davey M, Leamon JH, Johnson K, Milgrew MJ, Edwards M, Hoon J, Simons JF, Marran D, Myers JW, Davidson JF, Branting A, Nobile JR, Puc BP, Light D, Clark TA, Huber M, Branciforte JT, Stoner IB, Cawley SE, Lyons M, Fu Y, Homer N, Sedova M, Miao X, Reed B, Sabina J, Feierstein E, Schorn M, Alanjary M, Dimalanta E, Dressman D, Kasinskas R, Sokolsky T, Fianza JA, Namsaraev E, McKernan KJ, Williams A, Roth GT and Bustillo J (2011). An integrated semiconductor device enabling non-optical genome sequencing. *Nature* 475(7356): 348-352.
- Ryan E, Sheahan K, Creavin B, Mohan HM and Winter DC (2017). The current value of determining the mismatch repair status of colorectal cancer: A rationale for routine testing. *Crit Rev Oncol Hematol* 116: 38-57.
- Ryland GL, Doyle MA, Goode D, Boyle SE, Choong DY, Rowley SM, Li J, Bowtell DD, Tothill RW, Campbell IG and Gorringer KL (2015). Loss of heterozygosity: what is it good for? *BMC Med Genomics* 8: 45.
- Saeki H, Kitao H, Yoshinaga K, Nakanoko T, Kubo N, Kakeji Y, Morita M and Maehara Y (2011). Copy-neutral loss of heterozygosity at the p53 locus in carcinogenesis of esophageal squamous cell carcinomas associated with p53 mutations. *Clin Cancer Res* 17(7): 1731-1740.
- Salipante SJ, Scroggins SM, Hampel HL, Turner EH and Pritchard CC (2014). Microsatellite instability detection by next generation sequencing. *Clin Chem* 60(9): 1192-1199.
- Sansregret L, Vanhaesebroeck B and Swanton C (2018). Determinants and clinical implications of chromosomal instability in cancer. *Nat Rev Clin Oncol* 15(3): 139-150.
- Scaltriti M and Baselga J (2006). The epidermal growth factor receptor pathway: a model for targeted therapy. *Clin Cancer Res* 12(18): 5268-5272.
- Schmidt T, Sicic L, Blank S, Becker K, Weichert W, Bruckner T, Parakonthun T, Langer R, Buchler MW, Siewert JR, Lordick F and Ott K (2014). Prognostic value of histopathological regression in 850 neoadjuvantly treated oesophagogastric adenocarcinomas. *Br J Cancer* 110(7): 1712-1720.

- Siewert JR and Stein HJ (1998). Classification of adenocarcinoma of the oesophagogastric junction. *Br J Surg* 85(11): 1457-1459.
- Sitarz R, Skierucha M, Mielko J, Offerhaus GJA, Maciejewski R and Polkowski WP (2018). Gastric cancer: epidemiology, prevention, classification, and treatment. *Cancer Manag Res* 10: 239-248.
- Smyth EC, Verheij M, Allum W, Cunningham D, Cervantes A and Arnold D (2016). Gastric cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann Oncol* 27(suppl 5): v38-v49.
- Smyth EC, Wotherspoon A, Peckitt C, Gonzalez D, Hulkki-Wilson S, Eltahir Z, Fassan M, Rugge M, Valeri N, Okines A, Hewish M, Allum W, Stenning S, Nankivell M, Langley R and Cunningham D (2017). Mismatch Repair Deficiency, Microsatellite Instability, and Survival: An Exploratory Analysis of the Medical Research Council Adjuvant Gastric Infusional Chemotherapy (MAGIC) Trial. *JAMA Oncol* 3(9): 1197-1203.
- Smyth EC and Moehler M (2019). Late-line treatment in metastatic gastric cancer: today and tomorrow. *Ther Adv Med Oncol* 11: 1758835919867522.
- Sohn BH, Hwang JE, Jang HJ, Lee HS, Oh SC, Shim JJ, Lee KW, Kim EH, Yim SY, Lee SH, Cheong JH, Jeong W, Cho JY, Kim J, Chae J, Lee J, Kang WK, Kim S, Noh SH, Ajani JA and Lee JS (2017). Clinical Significance of Four Molecular Subtypes of Gastric Cancer Identified by The Cancer Genome Atlas Project. *Clin Cancer Res* 23(15): 4441-4449.
- Sondka Z, Bamford S, Cole CG, Ward SA, Dunham I and Forbes SA (2018). The COSMIC Cancer Gene Census: describing genetic dysfunction across all human cancers. *Nat Rev Cancer* 18(11): 696-705.
- Stelzer G, Rosen N, Plaschkes I, Zimmerman S, Twik M, Fishilevich S, Stein TI, Nudel R, Lieder I, Mazor Y, Kaplan S, Dahary D, Warshawsky D, Guan-Golan Y, Kohn A, Rappaport N, Safran M and Lancet D (2016). The GeneCards Suite: From Gene Data Mining to Disease Genome Sequence Analyses. *Curr Protoc Bioinformatics* 54: 1.30.31-31.30.33.
- Stolte M and Meining A (1998). Helicobacter pylori and Gastric Cancer. *Oncologist* 3(2): 124-128.
- Takahashi T, Saikawa Y and Kitagawa Y (2013). Gastric cancer: current status of diagnosis and treatment. *Cancers (Basel)* 5(1): 48-63.
- Tan P and Yeoh KG (2015). Genetics and Molecular Pathogenesis of Gastric Adenocarcinoma. *Gastroenterology* 149(5): 1153-1162 e1153.
- Tapia O, Riquelme I, Leal P, Sandoval A, Aedo S, Weber H, Letelier P, Bellolio E, Villaseca M, Garcia P and Roa JC (2014). The PI3K/AKT/mTOR pathway is activated in gastric cancer with potential prognostic and predictive significance. *Virchows Arch* 465(1): 25-33.
- Tate JG, Bamford S, Jubb HC, Sondka Z, Beare DM, Bindal N, Boutselakis H, Cole CG, Creatore C, Dawson E, Fish P, Harsha B, Hathaway C, Jupe SC, Kok CY, Noble K, Ponting L, Ramshaw CC, Rye CE, Speedy HE, Stefancsik R, Thompson SL, Wang S, Ward S, Campbell PJ and Forbes SA (2019). COSMIC: the Catalogue Of Somatic Mutations In Cancer. *Nucleic Acids Res* 47(D1): D941-d947.
- TCGA (2014). Comprehensive molecular characterization of gastric adenocarcinoma. *Nature* 513(7517): 202-209.
- TCGA (2017). Integrated genomic characterization of oesophageal carcinoma. *Nature* 541(7636): 169-175.
- Thies S and Langer R (2013). Tumor regression grading of gastrointestinal carcinomas after neoadjuvant treatment. *Front Oncol* 3: 262-262.
- Thompson LL, Jeusset LM, Lepage CC and McManus KJ (2017). Evolving Therapeutic Strategies to Exploit Chromosome Instability in Cancer. *Cancers (Basel)* 9(11).
- Tijhuis AE, Johnson SC and McClelland SE (2019). The emerging links between chromosomal instability (CIN), metastasis, inflammation and tumour immunity. *Mol Cytogenet* 12: 17-17.
- Tomlinson I, Halford S, Aaltonen L, Hawkins N and Ward R (2002). Does MSI-low exist? *J Pathol* 197(1): 6-13.
- UICC (2017). TNM classification of malignant tumors. Eight edition. Edited by James D. Brierley, Mary K. Gospodarowicz and Christian Wittekind. Oxford UK, Hoboken NJ. John Wiley & Sons
- van Beek J, zur Hausen A, Klein Kranenbarg E, van de Velde CJ, Middeldorp JM, van den Brule AJ, Meijer CJ and Bloemena E (2004). EBV-positive gastric adenocarcinomas: a distinct clinicopathologic entity with a low frequency of lymph node involvement. *J Clin Oncol* 22(4): 664-670.

- van der Post RS, Vogelaar IP, Carneiro F, Guilford P, Huntsman D, Hoogerbrugge N, Caldas C, Schreiber KEC, Hardwick RH, Ausems MGEM, Bardram L, Benusiglio PR, Bisseling TM, Blair V, Bleiker E, Boussioutas A, Cats A, Coit D, DeGregorio L, Figueiredo J, Ford JM, Heijkoop E, Hermens R, Humar B, Kaurah P, Keller G, Lai J, Ligtenberg MJL, O'Donovan M, Oliveira C, Pinheiro H, Ragnath K, Rasenberg E, Richardson S, Roviello F, Schackert H, Seruca R, Taylor A, Ter Huurne A, Tischkowitz M, Joe STA, van Dijck B, van Grieken NCT, van Hillegersberg R, van Sandick JW, Vehof R, van Krieken JH and Fitzgerald RC (2015). Hereditary diffuse gastric cancer: updated clinical guidelines with an emphasis on germline CDH1 mutation carriers. *J Med Genet* 52(6): 361-374.
- van Tilborg AAG, Kompier LC, Lurkin I, Poort R, El Bouazzaoui S, van der Keur K, Zuiverloon T, Dyrskjot L, Orntoft TF, Roobol MJ and Zwarthoff EC (2012). Selection of microsatellite markers for bladder cancer diagnosis without the need for corresponding blood. *PLoS one* 7(8): e43345-e43345.
- Vanderwalde A, Spetzler D, Xiao N, Gatalica Z and Marshall J (2018). Microsatellite instability status determined by next-generation sequencing and compared with PD-L1 and tumor mutational burden in 11,348 patients. *Cancer Med* 7(3): 746-756.
- Velho S, Fernandes MS, Leite M, Figueiredo C and Seruca R (2014). Causes and consequences of microsatellite instability in gastric carcinogenesis. *World J Gastroenterol* 20(44): 16433-16442.
- Walther A, Houlston R and Tomlinson I (2008). Association between chromosomal instability and prognosis in colorectal cancer: a meta-analysis. *Gut* 57(7): 941-950.
- Wang K, Li M and Hakonarson H (2010). ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* 38(16): e164.
- Wang K, Kan J, Yuen ST, Shi ST, Chu KM, Law S, Chan TL, Kan Z, Chan AS, Tsui WY, Lee SP, Ho SL, Chan AK, Cheng GH, Roberts PC, Rejto PA, Gibson NW, Pocalyko DJ, Mao M, Xu J and Leung SY (2011). Exome sequencing identifies frequent mutation of ARID1A in molecular subtypes of gastric cancer. *Nat Genet* 43(12): 1219-1223.
- Wang K, Yuen ST, Xu J, Lee SP, Yan HH, Shi ST, Siu HC, Deng S, Chu KM, Law S, Chan KH, Chan AS, Tsui WY, Ho SL, Chan AK, Man JL, Foglizzo V, Ng MK, Chan AS, Ching YP, Cheng GH, Xie T, Fernandez J, Li VS, Clevers H, Rejto PA, Mao M and Leung SY (2014). Whole-genome sequencing and comprehensive molecular profiling identify new driver mutations in gastric cancer. *Nat Genet* 46(6): 573-582.
- Wang X-Q, Terry P-D and Yan H (2009). Review of salt consumption and stomach cancer risk: epidemiological and biological evidence. *World J Gastroenterol* 15(18): 2204-2213.
- Wang Z, Duan J, Cai S, Han M, Dong H, Zhao J, Zhu B, Wang S, Zhuo M, Sun J, Wang Q, Bai H, Han J, Tian Y, Lu J, Xu T, Zhao X, Wang G, Cao X, Li F, Wang D, Chen Y, Bai Y, Zhao J, Zhao Z, Zhang Y, Xiong L, He J, Gao S and Wang J (2019). Assessment of Blood Tumor Mutational Burden as a Potential Biomarker for Immunotherapy in Patients With Non-Small Cell Lung Cancer With Use of a Next-Generation Sequencing Cancer Gene Panel. *JAMA Oncol* 5(5): 696-702.
- Warren JR and Marshall B (1983). Unidentified curved bacilli on gastric epithelium in active chronic gastritis. *Lancet* 1(8336): 1273-1275.
- Watanabe T, Kobunai T, Yamamoto Y, Matsuda K, Ishihara S, Nozawa K, Yamada H, Hayama T, Inoue E, Tamura J, Inuma H, Akiyoshi T and Muto T (2012). Chromosomal instability (CIN) phenotype, CIN high or CIN low, predicts survival for colorectal cancer. *J Clin Oncol* 30(18): 2256-2264.
- Welker ME and Kulik G (2013). Recent syntheses of PI3K/Akt/mTOR signaling pathway inhibitors. *Bioorg Med Chem* 21(14): 4063-4091.
- WHO (1994). International Agency for Research on Cancer, IARC Monographs on the Evaluation of Carcinogenic Risks to Humans Volume 61, Schistosomes, Liver Flukes and Helicobacter pylori, Lyon, June 7-14. 61: 1-241.
- WHO (2019). Digestive System Tumours, WHO Classification of Tumours, 5th Edition, Volume 1, WHO Classification of Tumours Editorial Board.
- Wilfinger WW, Mackey K and Chomczynski P (1997). Effect of pH and ionic strength on the spectrophotometric assessment of nucleic acid purity. *BioTechniques* 22(3): 474-476, 478-481.

- Wright CM, Dent OF, Newland RC, Barker M, Chapuis PH, Bokey EL, Young JP, Leggett BA, Jass JR and Macdonald GA (2005). Low level microsatellite instability may be associated with reduced cancer specific survival in sporadic stage C colorectal carcinoma. *Gut* 54(1): 103-108.
- Xu Z, Huo X, Ye H, Tang C, Nandakumar V, Lou F, Zhang D, Dong H, Sun H, Jiang S, Zhang G, Liu Z, Dong Z, Guo B, He Y, Yan C, Wang L, Su Z, Li Y, Gu D, Zhang X, Wu X, Wei X, Hong L, Zhang Y, Yang J, Gong Y, Tang C, Jones L, Huang XF, Chen SY and Chen J (2014). Genetic mutation analysis of human gastric adenocarcinomas using ion torrent sequencing platform. *PLoS One* 9(7): e100442.
- Yamamoto H and Imai K (2019). An updated review of microsatellite instability in the era of next-generation sequencing and precision medicine. *Semin Oncol* 46(3): 261-270.
- Yuasa Y (2003). Control of gut differentiation and intestinal-type gastric carcinogenesis. *Nat Rev Cancer* 3(8): 592-600.
- Zang ZJ, Cutcutache I, Poon SL, Zhang SL, McPherson JR, Tao J, Rajasegaran V, Heng HL, Deng N, Gan A, Lim KH, Ong CK, Huang D, Chin SY, Tan IB, Ng CC, Yu W, Wu Y, Lee M, Wu J, Poh D, Wan WK, Rha SY, So J, Salto-Tellez M, Yeoh KG, Wong WK, Zhu YJ, Futreal PA, Pang B, Ruan Y, Hillmer AM, Bertrand D, Nagarajan N, Rozen S, Teh BT and Tan P (2012). Exome sequencing of gastric adenocarcinoma identifies recurrent somatic mutations in cell adhesion and chromatin remodeling genes. *Nat Genet* 44(5): 570-574.
- Zerbino DR, Achuthan P, Akanni W, Amode MR, Barrell D, Bhai J, Billis K, Cummins C, Gall A, Giron CG, Gil L, Gordon L, Haggerty L, Haskell E, Hourlier T, Izuogu OG, Janacek SH, Juettemann T, To JK, Laird MR, Lavidas I, Liu Z, Loveland JE, Maurel T, McLaren W, Moore B, Mudge J, Murphy DN, Newman V, Nuhn M, Ogeh D, Ong CK, Parker A, Patricio M, Riat HS, Schuilenburg H, Sheppard D, Sparrow H, Taylor K, Thormann A, Vullo A, Walts B, Zadissa A, Frankish A, Hunt SE, Kostadima M, Langridge N, Martin FJ, Muffato M, Perry E, et al. (2018). Ensembl 2018. *Nucleic Acids Res* 46(D1): D754-D761.
- Zhang H-P, Li S-Y, Wang J-P and Lin J (2018). Clinical significance and biological roles of cyclins in gastric cancer. *Onco Targets Ther* 11: 6673-6685.

## 8. PUBLICATIONS AND CONGRESS CONTRIBUTIONS

### 8.1. Publications

Kohlruss M, Reiche M, Jesinghaus M, Grosser B, Slotta-Huspenina J, Hapfelmeier A, Bauer L, Novotny A, Weichert W and Keller G (2018). A microsatellite based multiplex PCR method for the detection of chromosomal instability in gastric cancer. *Sci Rep* 8: 12551.

Kohlruss M, Grosser B, Krenauer M, Slotta-Huspenina J, Jesinghaus M, Blank S, Novotny A, Reiche M, Schmidt T, Ismani L, Hapfelmeier A, Mathias D, Meyer P, Gaida MM, Bauer L, Ott K, Weichert W and Keller G (2019). Prognostic implication of molecular subtypes and response to neoadjuvant chemotherapy in 760 gastric carcinomas: role of Epstein-Barr virus infection and high- and low-microsatellite instability. *J Pathol Clin Res* 5(4): 227-239.

### 8.2. Congress contributions

Kohlruss M, Reiche M, Bauer L, Jesinghaus M, Slotta-Huspenina J, Novotny A, Hapfelmeier A, Weichert W and Keller G (2017). An optimized multiplex microsatellite assay for the determination of chromosomal instability in gastric cancer (Oral presentation).

**101. Jahrestagung der Deutschen Gesellschaft für Pathologie**, Erlangen 2017

Kohlruss M, Reiche M, Grosser B, Bauer L, Jesinghaus M, Slotta-Huspenina J, Novotny A, Hapfelmeier A, Weichert W and Keller G (2018). Molecular TCGA classification and clinical relevance in gastric cancer (Poster presentation).

**102. Jahrestagung der Deutschen Gesellschaft für Pathologie**, Berlin 2018

Kohlruss M, Reiche M, Jesinghaus M, Grosser B, Slotta-Huspenina J, Hapfelmeier A, Bauer L, Novotny A, Weichert W and Keller G (2019). A microsatellite based multiplex PCR method for the detection of chromosomal instability in gastric cancer (Poster presentation).

**13th International Gastric Cancer Congress IGCC**, Prag 2019

Kohlruss M, Grosser B, Krenauer M, Slotta-Huspenina J, Jesinghaus M, Blank S, Novotny A, Reiche M, Schmidt T, Ismani L, Hapfelmeier A, Mathias D, Meyer P, Gaida MM, Bauer L, Ott K, Weichert W and Keller G (2019). Epstein-Barr virus infection and microsatellite instability as potential prognostic and predictive markers in Gastric Cancer.

**13th International Gastric Cancer Congress IGCC**, Prag 2019

Kohlruss M, Grosser B, Krenauer M, Slotta-Huspenina J, Jesinghaus M, Blank S, Novotny A, Reiche M, Schmidt T, Ismani L, Hapfelmeier A, Mathias D, Meyer P, Gaida MM, Bauer L, Ott K, Weichert W and Keller G (2019). Prognostic implication of molecular subtypes and response to neoadjuvant chemotherapy in gastric cancer: role of EBV infection and high and low microsatellite instability (Poster presentation).

**9. Mildred Scheel Cancer Conference**, Bonn 2019

Kohlruss M, Grosser B, Krenauer M, Slotta-Huspenina J, Jesinghaus M, Blank S, Novotny A, Reiche M, Schmidt T, Ismani L, Hapfelmeier A, Mathias D, Meyer P, Gaida MM, Bauer L, Ott K, Weichert W and Keller G (2019). Molekulare Subgruppen beim Magenkarzinom: Prognostische und prädiktive Relevanz von EBV-Infektion und hoher und niedriger Mikrosatelliteninstabilität (Oral presentation).

**103. Jahrestagung der Deutschen Gesellschaft für Pathologie**, Frankfurt a.M. 2019

## 9. ACKNOWLEDGEMENT

First of all, I would like to thank my supervisor Prof. Gisela Keller for the excellent support during the entire doctoral thesis. Thanks for always giving me motivating impulses and constructive feedback even in challenging periods, thanks for all the opportunities you offered to me to extend my skills as scientist and thanks for always sharing your expert knowledge with me. Thank you Gisela that “your door” was always open to address any kind of problem or question.

I further would like to thank Prof. Wilko Weichert for being the mentor of my thesis, for inspiring discussions and for giving me the opportunity to perform my work at the Institute of Pathology. I also would like to thank my second supervisor Prof. Michael Groll for reviewing my doctoral thesis.

Further, I would like to thank my colleagues of the Institute of Pathology who supported my work in various ways. A special thanks goes to Nicole Pfarr for teaching and explaining me the NGS method as well as for sharing the NGS expertise and to Julia Slotta-Huspenina and Moritz Jesinghaus for answering pathological and histological questions. I also would like to thank our cooperation partners Alexander Hapfelmeier for supporting me in all the statistical problems and Alexander Novotny for answering clinical questions about the patient cohorts.

I specially would like to thank my colleagues and friends Bianca, Marie and Dona for assisting me with data collection of the many samples, for having enjoyable coffee and lunch breaks and for the great time we had together in the lab. And of course, I also would like to thank all current and former “AG-Keller” members especially Lukas and Magda who supported me at the beginning of my dissertation and laid the foundation for the analysis of the large patient cohorts.

Further I would like to thank all the members of the routine diagnostic lab. Thank you, Daniela for sharing lots of time with me in front of our beloved sequencer, for the amusing conversations and for solving any kind of lab problem. Thanks Susanne, Yoko and Mirjam for the nice and quick chats in the lab or on the stairs and also for your help in any kind of lab question.

I also would like to thank my wonderful office mates Gwen, Elena and Sarah for the humorous and inspiring conversations or discussions and for keeping cool or warm at 32°C or -5° in our “cold-blooded” office.

I specially would like to thank my two “biology” friends Svenja and Sandra for reviewing my thesis, for giving constructive and motivating feedback and thank you for the enjoyable lunch dates.

Finally, and most importantly, I would like to thank all my friends and my family. Thank you dad, Anke and Wencke for encouraging and motivating me all the time. Thank you Nico for your emotional support, for listening to my scientific problems and for being patient with me all the time. Thank you all for always believing in me and for supporting me whenever I needed your help.