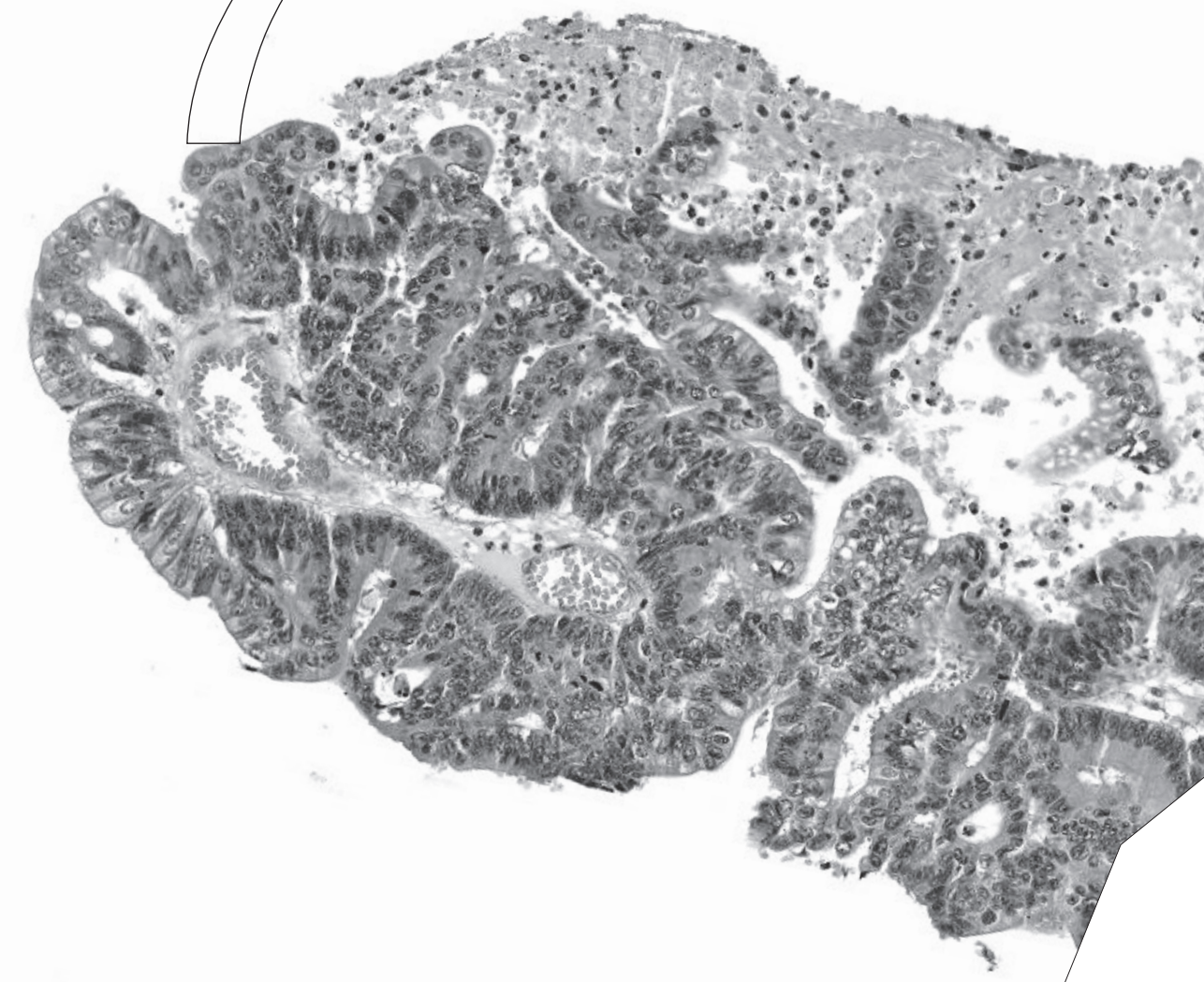
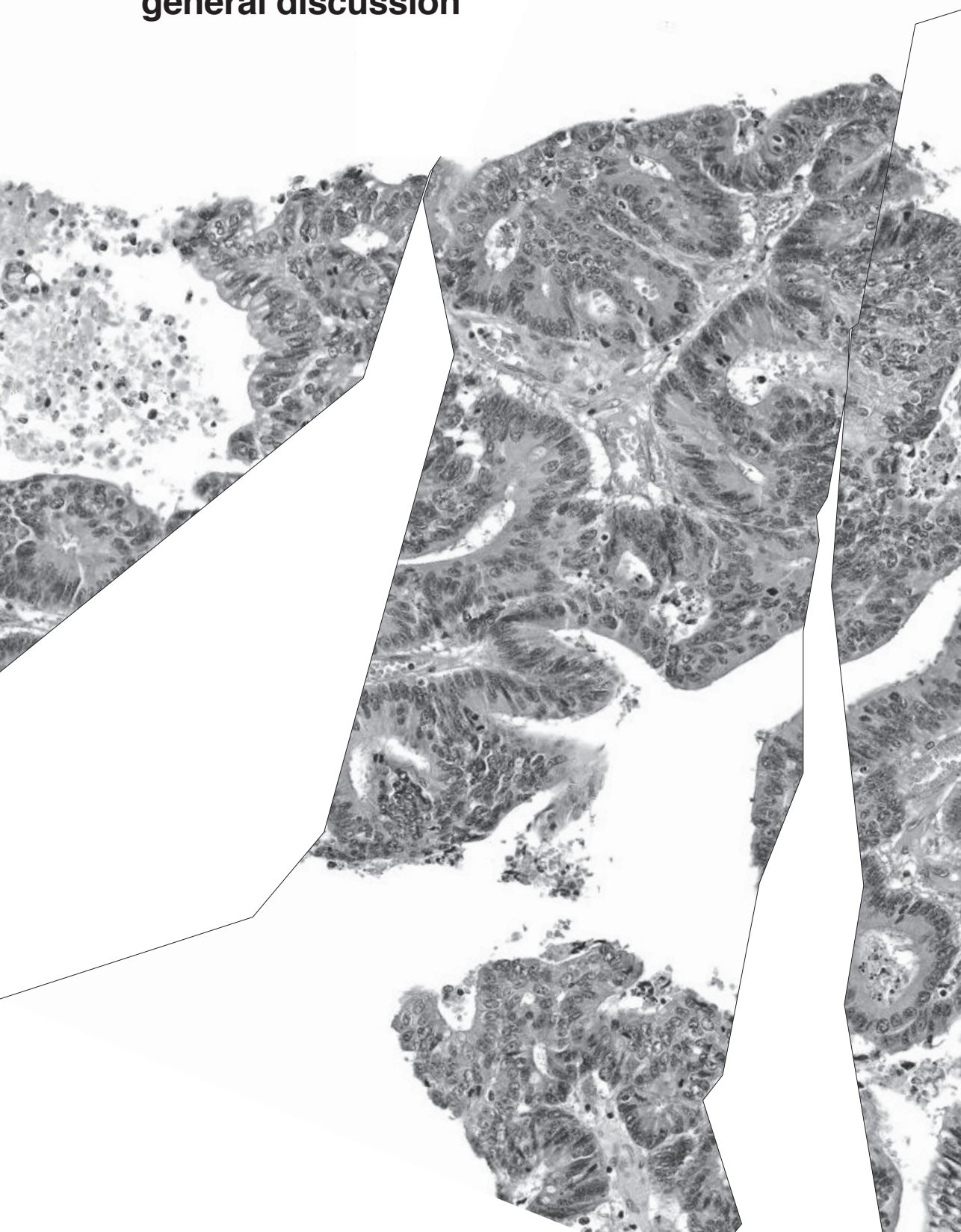
7



Summary and general discussion



INTRODUCTION AND AIM OF THIS THESIS

Colorectal cancer (CRC) is caused by somatic genomic aberrations, sometimes superseded on germline mutations. These include small nucleotide variants (SNVs), chromosomal copy number aberrations (CNAs) and structural variants (SVs) [1-3]. Somatic DNA aberrations can subsequently affect biological processes that enable tumor progression [1,4]. Consequently, comprehensive profiling of somatic DNA aberrations may contribute to a better understanding of underlying molecular tumor pathology. Ultimately, patients could benefit from improved stratification for prognosis and therapy selection by molecular tumor characterization.

The aim of the present thesis was to investigate somatic DNA aberrations in CRC with the main focus on SVs, since their biological and clinical impact on CRC progression is relatively unknown compared to somatic SNVs and CNAs.

COMPUTATIONAL METHOD FOR DETECTION OF RECURRENT BREAKPOINT GENES

DNA CNA profiles obtained by for instance array Comparative Genomic Hybridization (CGH) or (low-pass) whole genome sequencing are widely used for identifying numerical chromosomal aberrations. Since CNA-associated SVs are characterized by CNA-associated breakpoint locations, *i.e.* chromosomal locations that separate the contiguous DNA copy number segments, analysis of widely generated high-resolution CNA profiles enables identification of genes recurrently affected by SVs in (large) series of tumor samples. The user-friendly computational method, 'GeneBreak', which is described in **chapter two**, was designed to systematically identify genes that are recurrently affected by CNA-associated breakpoints from high-resolution CNA profiles. Both CNA-associated breakpoint detection and a tailored annotation approach for breakpoint-to-gene mapping were implemented. Furthermore, dedicated cohort-based statistics was incorporated, including correction for covariates that influence the probability to be a breakpoint gene and integration of multiple testing correction, enabling identification of recurrent breakpoint genes across (large) series of samples. The computational method 'GeneBreak' was applied in all studies described in this thesis when CNA-associated breakpoint gene identification was performed (**chapters 3-5**). The R-package has been made available from Bioconductor (http://www.bioconductor.org).

PREVALENCE AND RELEVANCE OF RECURRENT BREAKPOINT GENES IN CRC

CNA-associated breakpoints indicate underlying chromosomal breaks and thus genomic locations affected by SVs. In **chapter three**, a large cohort of 352 CRC primary tumor samples from CAIRO and CAIRO2 phase III clinical trials [5,6] was used to investigate what genes were affected by CNA-associated chromosomal breaks. This analysis pinpointed 748 genes that were recurrently

affected by CNA-associated breakpoints (FDR<0.1) among which 170 being affected in >3% of CRCs. This showed that recurrent breakpoint genes were highly prevalent in CRC and that frequencies of gene mutations by CNA-associated breakpoints are comparable to frequencies of gene mutation by SNVs in CRCs. Next, to distinguish CRC driver from passenger mutations *de-novo*, modules of putative driver genes were composed by Multi-Dendrix analysis. Several breakpoint genes were pinpointed as putative CRC drivers, in addition to the well-known non-synonymous SNVs, indicating the biological relevance of breakpoint genes. Moreover, the clinical relevance of breakpoint genes was demonstrated by patient stratification based on gene breakpoints and well-known non-synonymous SNVs using prior knowledge about gene interaction networks. This analysis revealed four CRC subtypes one of which was associated with a very poor prognosis in the setting of metastasized disease. Interestingly, this subtype included eight out of ten MSI CRCs and might be associated with compromised immune suppressive capacity.

Chapter four reported cancer genome profiling of somatic SNVs, CNAs and breakpoint genes on a selected series of 57 stage II and 57 stage III MSS colon cancers aiming to identify indicators of disease recurrence. None of these stage II and all stage III patients had received 5-fluorouracil-based adjuvant chemotherapy. Interestingly, mutations in *APC*, which is a key inhibitor of the WNT signaling pathway [7], were associated with worse outcome. Firstly, *APC* mutations were associated with a poor disease-free survival (DFS) in MSS stage III (5-fluorouracil treated), but not in stage II, colon cancer patients. Moreover, mutations in *APC* and *KRAS* were frequently co-occurring and the impact on DFS was further enhanced when *APC* mutations were combined with mutations in MAPK signaling pathway genes (*KRAS*, *BRAF* and *NRAS*), which suggests a synergistic effect of coinciding activation of WNT and MAPK signaling pathways. Secondly, tumors that have an *APC* and/or *TP53* mutation showed increased extent of CNAs and number of CNA-associated breakpoints, indicating a preventing role for *APC* in chromosomal instability (CIN). In addition, we demonstrated that copy number loss of chromosome 18q12.1 - 18q12.2 was associated with disease recurrence. However, loss of this region was not significantly associated with a poor DFS. Taken together, genomic aberrations that are highly prevalent in CRC (*APC*, *TP53*, *KRAS* and 18q loss) were associated with disease progression in this series of MSS colon cancers. Interestingly, these aforementioned mutations (*APC*, *TP53*, *KRAS* and 18q loss) are known to be important for CRC development and have been incorporated in the established CRC progression model [3,8].

The contribution to colorectal adenoma-to-carcinoma progression of somatic non-synonymous SNVs in genes and somatic copy number aberrations (CNAs) has been widely examined. However, the impact on CRC development of genes affected by CNA-associated breakpoints has received less attention. Recurrent CNA-associated breakpoint genes that could contribute to adenoma-to-carcinoma progression were described in **chapter five**. For this analysis, high-resolution aCGH profiles from 118 colorectal adenoma [9] and 466 colorectal carcinomas were used. A total of 21 recurrent breakpoint genes were more frequently affected in colorectal carcinomas compared to adenomas ($p<0.05$). *MACROD2* was the most prominent breakpoint gene in CRCs (39.5%) while it did not show an SV in any of the 118 colorectal adenoma samples. Furthermore, clinical

impact of these 21 genes in tumor malignancy was supported by the observation that breakpoint events in one or more of these 21 genes were associated with worse DFS in stage II and stage III MSS colon cancer patients (n=114). Taken together, these observations suggest that these 21 genes, which were nearly exclusively affected in carcinomas compared to adenomas, are likely to be causally involved in malignant transformation of benign colorectal precursor lesions into CRC following well-known somatic SNVs and CNAs [3,8].

Since *MACROD2* was the most prominent candidate breakpoint gene in all studies of this thesis (**chapters 3-5**), further evaluation of the prognostic value of MACROD2 protein expression was performed by immunohistochemistry. Tissue microarrays containing core biopsies from a cohort of 386 stage II and stage III colon cancers [10] were stained and scored for MACROD2 protein expression. The result described in **chapter six** showed that low MACROD2 protein expression in microsatellite stable (MSS) stage III primary colon tumors predicts poor response to 5-fluorouracil-based adjuvant chemotherapy. Although this candidate gene was further explored following results obtained from CNA-associated breakpoint analyses, correlation between chromosomal CNA-associated breaks in *MACROD2* and loss of MACROD2 protein expression could not be investigated for technical reasons. Nevertheless, somatic aberrations including SVs that affect *MACROD2* may impact tumor biology rather than being an artifact associated with chromosome fragility since MACROD2 protein expression is a potential clinically relevant biomarker for therapy stratification in our series of stage III MSS colon cancer patients.

CONCLUSIONS AND FUTURE PERSPECTIVE

In current clinical practice, CRC patients are stratified mainly based on histopathological features of the tumor. Molecular profiling of somatic genomic aberrations is expected to improve prognosis and therapy prediction of individual patients. Comprehensive cancer genome characterization will further our understanding of how somatic DNA aberrations contribute to disease progression and may also enable biomarker identification and development of targeted drugs. In contrast to widely studied somatic non-synonymous SNVs and DNA CNAs, large-scale genome-wide analyses of SVs are scarce. This thesis provides an overview of genes that were recurrently affected by somatic CNA-associated breakpoints by genome-wide analyses of high-resolution CNA profiles from large series of CRC samples, covering the spectrum from benign to malignant (metastatic) lesions. It demonstrates that this class of gene mutations is highly prevalent in CRC. Moreover, it shows that SVs may substantially drive disease progression instead of merely being an epiphenomenon of CIN.

The results presented in this thesis underline that comprehensive molecular subtyping efforts should include SV analysis. Although our approach enables identification of a substantial subset of SVs, it will not provide an exhaustive overview of SVs in the cancer genome since the resolution of aCGH was limited with an average probe spacing of 17kb, and DNA copy number

neutral SVs could not be detected by our approach. Nevertheless, it was feasible to allocate chromosomal breaks to gene positions and to identify recurrent breakpoints of SVs by the use of large series of samples. Massive parallel sequencing (MPS) will enable detection of breakpoints at nucleotide level that are associated with CNA, and may also reveal copy number neutral SVs. At this moment, large series of CRC genomes with well-documented clinical information that are characterized by high-throughput deep whole genome sequencing are not yet available. Furthermore, MPS approaches have the advantage that it will also provide insight in the structure of the somatic chromosomal rearrangement. In addition, MPS analysis would allow detection of three distinct mutational classes, including SNVs, CNAs and SVs, which makes it an efficient technique for genotyping [11,12]. This appealing technique may enable assessment of a range of genomic aberrations in the primary tumor in order to predict prognosis and response to therapies. MPS can also be applied to genetic material from liquid biopsies such as blood, a promising method that may allow early tumor detection, monitoring of response to therapy, characterization of tumor heterogeneity and thereby identification of possible resistance mechanisms as well [13,14].

Although DNA mutations are considered causal for tumor development en thus tumor behavior, the work presented in **chapter 6** demonstrates that protein expression may not directly correlate with the mutation status of *MACROD2* at DNA level. This illustrates that cancer biology is a complex mechanism where also other mechanisms than DNA alterations may influence gene expression. Although it is still uncertain what influenced the altered protein expression of MACROD2, it may turn out to be a relevant biomarker for response to chemotherapy prediction in stage III CRC. Besides evaluation of irreversible DNA mutations, integration analysis of the (epi-)genome, transcriptome and proteome may lead to better understanding of the tumor biology complexity. Independent external validation of this candidate biomarker is still needed to confirm its predictive value. Also clinical value of several other promising candidate markers identified in the studies described in this thesis awaits further validation.

In conclusion, the results obtained in this thesis demonstrate the high prevalence and putative biological and clinical impact of somatic CNA-associated breakpoint genes on CRC progression. This work thereby makes a contribution to the molecular classification of CRC for patient stratification in prognosis and therapy prediction.

REFERENCES

1. Stratton MR, Campbell PJ, Futreal PA. The cancer genome. Nature. 2009;458: 719–724.
2. Cancer Genome Atlas Network. Comprehensive molecular characterization of human colon and rectal cancer. Nature. 2012;487: 330–337.
3. Fearon ER, Vogelstein B. A genetic model for colorectal tumorigenesis. Cell. 1990;61: 759–767.
4. Hanahan D, Weinberg RA. Hallmarks of Cancer: The Next Generation. Cell. 2011;144: 646–674.
5. Koopman M, Antonini NF, Douma J, Wals J, Honkoop AH, Erdkamp FLG, et al. Sequential versus combination chemotherapy with capecitabine, irinotecan, and oxaliplatin in advanced colorectal cancer (CAIRO): a phase III randomised controlled trial. Lancet. 2007;370: 135–142.
6. Tol J, Koopman M, Cats A, Rodenburg CJ, Creemers GJM, Schrama JG, et al. Chemotherapy, bevacizumab, and cetuximab in metastatic colorectal cancer. N Engl J Med. 2009;360: 563–572.
7. Clevers H. Wnt/beta-catenin signaling in development and disease. Cell. 2006;127: 469–480.
8. Fearon ER. Molecular genetics of colorectal cancer. Annu Rev Pathol. 2011;6: 479–507.
9. Voorham QJM, Carvalho B, Spiertz AJ, van Grieken NCT, Mongera S, Rondagh EJA, et al. Chromosome 5q loss in colorectal flat adenomas. Clin Cancer Res. 2012;18: 4560–4569.
10. Belt EJT, Fijneman RJA, van den Berg EG, Bril H, Delis-van Diemen PM, Tijssen M, et al. Loss of lamin A/C expression in stage II and III colon cancer is associated with disease recurrence. Eur J Cancer. 2011;47: 1837–1845.
11. Meyerson M, Gabriel S, Getz G. Advances in understanding cancer genomes through second-generation sequencing. Nat Rev Genet. 2010;11: 685–696.
12. Alkan C, Coe BP, Eichler EE. Genome structural variation discovery and genotyping. Nat Rev Genet. 2011;12: 363–376.
13. Ding L, Ley TJ, Larson DE, Miller CA, Koboldt DC, Welch JS, et al. Clonal evolution in relapsed acute myeloid leukaemia revealed by whole-genome sequencing. Nature. 2012;481: 506–509.
14. Bettegowda C, Sausen M, Leary RJ, Kinde I, Wang Y, Agrawal N, et al. Detection of Circulating Tumor DNA in Early- and Late-Stage Human Malignancies. Sci Transl Med. 2014;6: 224ra24.

7