UPPSALA UNIVERSITET

Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine 1443

Targeting allelic loss in colorectal cancer

VERÓNICA RENDO

ACTA
UNIVERSITATIS
UPSALIENSIS
UPPSALA
2018

ISSN 1651-6206
ISBN 978-91-513-0274-4
urn:nbn:se:uu:diva-345970

Dissertation presented at Uppsala University to be publicly examined in Rudbecksalen, Dag Hammarskjölds väg 20, Uppsala, Friday, 4 May 2018 at 09:00 for the degree of Doctor of Philosophy (Faculty of Medicine). The examination will be conducted in English. Faculty examiner: Professor Nickolas Papadopoulos (The Johns Hopkins Institutions, USA).

Abstract

Rendo, V. 2018. Targeting allelic loss in colorectal cancer. *Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine* 1443. 50 pp. Uppsala: Acta Universitatis Upsaliensis. ISBN 978-91-513-0274-4.

Targeted cancer therapy exploits molecular differences between tumor and normal cells to selectively kill cancer cells. Whereas targeting of activated oncogenes has proved clinically useful, few current therapies exploit loss-of-function mutations in tumor suppressor genes or in the genome at large. This thesis explores the consequences of allelic loss affecting tumor suppressor genes and passenger genes in colorectal cancer (CRC), aiming to identify vulnerabilities that can be exploited for therapy.

In **Paper I** we used genome editing to model inactivation of *PRDM2* and showed that *PRDM2* loss impacts cell growth and invasiveness, potentially mediated by genes involved in epithelial-to-mesenchymal transition. We confirmed the role of *PRDM2* as a tumor suppressor gene in CRC and proved that c.4467delA inactivating mutations constitute a driver event in CRC.

In **Paper II** we investigated whether the reduced allelic diversity resulting from loss of heterozygosity (LOH) in cancers could be exploited for therapy. We identified target genes by mapping prevalent alleles frequently lost in cancer and investigated *NAT2* loss in CRC. Drug discovery efforts identified a compound selectively toxic to tumor cells with reduced NAT2 activity, providing proof of concept for LOH targeting by small molecule drugs.

In **Paper III**, we aimed to widen the cohort of CRC patients eligible for NAT2 allele-selective chemotherapy. We determined NAT2 slow acetylator frequencies and LOH events in two independent cohorts by next-generation sequencing and genomic arrays. Next, we demonstrated enhanced response to allele-selective chemotherapy of tumor cells encoding additional prevalent NAT2 slow acetylator alleles, and developed a method for detection of *NAT2* allelic loss suitable for clinical use.

In **Paper IV**, we extended the search of therapeutic target genes by mining loss-of-function (LoF) alleles retained in tumors after LOH. This effort identified a prevalent splice site disruption in *CYP2D6* as a putative target and motivated the development of cell model systems to identify compounds targeting CYP2D6 loss in cancer cells.

In **Paper V** we characterized a set of 56 microsatellite stable CRCs by whole-genome sequencing in an attempt to understand the genetic causes leading to genomic instability and colorectal tumorigenesis. We confirmed the mutation frequencies of known CRC genes and identified for the first time the contribution of an unknown mutational process in 10% of the analyzed tumors.

Keywords: colorectal cancer, gene editing, loss of heterozygosity, targeted therapy, whole genome sequencing

Verónica Rendo, Department of Immunology, Genetics and Pathology, Rudbecklaboratoriet, Uppsala University, SE-751 85 Uppsala, Sweden.

ISSN 1651-6206
ISBN 978-91-513-0274-4
urn:nbn:se:uu:diva-345970 (http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-345970)

No te rindas que la vida es eso,
continuar el viaje,
perseguir tus sueños,
destrabar el tiempo,
correr los escombros,
y destapar el cielo.

Mario Benedetti

Don't give up, that's what life is
continue the journey,
follow your dreams,
unstuck time,
move the rubble,
and uncover the sky.

Mario Benedetti

List of Papers

This thesis is based on the following papers, which are referred to in the text by their Roman numerals.

I Pandzic, T., **Rendo, V.**, Lim, J., Larsson, C., Larsson, J., Stoimenov, I., Kundu, S., Akhtar Ali, M., Hellström, M., He, L., Lindroth, A. and Sjöblom, T. Somatic *PRDM2* c.4467delA mutations in colorectal cancers control histone methylation and tumor growth. (2017) *Oncotarget,* 8:98646-98659, doi.org/10.18632/oncotarget.21713.

II Stoimenov, I. *, **Rendo, V. ***, Mateus, A., Sjöberg, E., Svensson, R., Gustavsson, AL., Johansson, L., Ng, A., Artursson, P., Nygren, P., Cheong, I. and Sjöblom, T. Exploiting loss of heterozygosity for allele-selective colorectal cancer chemotherapy. *Submitted*.

III **Rendo, V.**, Kundu, S., Rameika, N., Ljungström, V., Svensson, R., Palin, K., Aaltonen, L., Sjöblom, T. and Stoimenov, I. Defining eligible patients for allele-selective chemotherapies targeting NAT2 in colorectal cancer. *Manuscript*.

IV **Rendo, V.**, Kundu, S., Rameika, N., Svensson, R., Ljungström, V., Stoimenov, I. and Sjöblom, T. Exploiting allelic loss in cancer therapy: identification of CYP2D6 as a potential target. *Manuscript.*

V **Rendo, V.**, Osterman, E., Ljungström, V.* and Sjöblom, T.* Characterization of colorectal cancers by whole genome sequencing. *Manuscript*.

*Equally contributing authors

Other papers by the PhD candidate not included in this thesis:

I Mathot, L., Kundu, S., Ljungström, V., Svedlund, J., Moens, L., Adlerteg, T., Falk-Sörqvist, E., **Rendo, V.**, Bellomo, C., Mayrhofer, M., Cortina, C., Sundström, M., Micke, P., Botling, J., Isaksson, A., Moustakas, A., Batlle, E., Birgisson, H., Glimelius, B., Nilsson, M. and Sjöblom, T. Somatic Ephrin Receptor Mutations Are Associated with Metastasis in Primary Colorectal Cancer. (2017) *Cancer Research,* 1;77(7):1730-1740, doi: 10.1158/0008-5472.CAN-16-1921.

Contents

Abbreviations

AICDA	Activation induced cytidine deaminase
AJCC	American Joint Committee on Cancer
APOBEC	Apolipoprotein B mRNA editing enzyme
ASI	Allele specific inhibition
BC	Before Christ
BER	Base excision repair
BSC	Best supportive care
CIMP	CpG island methylator phenotype
CIN	Chromosomal instability
CMS	Consensus molecular subtype
COSMIC	Catalogue Of Somatic Mutations In Cancer
CRC	Colorectal cancer
ctDNA	Circulating tumor DNA
DNA	Deoxyribonucleic acid
EGF	Epidermal growth factor
EMT	Epithelial to mesenchymal transition
EpCAM	Epithelial cell adhesion molecule
FAP	Familial adenomatous polyposis
FDA	Food and Drug Administration
FGF	Fibroblast growth factor
FOLFIRI	Folinic acid, fluorouracil, irinotecan hydrochloride
H3K9	Histone H3 Lysine 9
HNPCC	Hereditary nonpolyposis colorectal cancer
IFL	Irinotecan, fluorouracil, leucovorin
LoF	Loss-of-function
LOH	Loss of heterozygosity
mAb	Monoclonal antibody
MAP	*MUTYH*-associated polyposis
mCRC	Metastatic CRC
MMR	Mismatch repair
MSI	Microsatellite instability
MTT	3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium
NGS	Next-generation sequencing
nsSNV	Non-synonymous SNV
OS	Overall survival
PARP	Poly (ADP-ribose) polymerase

PCR	Polymerase chain reaction
PFS	Progression-free survival
rAAV	Recombinant adeno-associated virus
RNA	Ribonucleic acid
RNAi	Interference RNA
SCNA	Somatic copy number alteration
shRNA	Short hairpin RNA
siRNA	Small interference RNA
SNV	Single nucleotide variation
STR	Short tandem repeat
TCGA	The Cancer Genome Atlas
TNM	Tumor, nodes, metastasis
TS	Targeted sequencing
TSG	Tumor suppressor gene
UICC	Union for International Cancer Control
UTR	Untranslated region
UV	Ultraviolet
VEGF	Vascular endothelial growth factor
WES	Whole exome sequencing
WGS	Whole genome sequencing

Introduction

Cancer as a genetic disease

Cancer is a genetic disease characterized by uncontrolled cell growth. It remains a leading cause of morbidity and mortality in the world, with 14.1 million new cases diagnosed every year and 8.2 million deaths reported[1]. Cancer comprises over 200 types of disease that can originate in most organs of the body, showing a high level of intra- and inter-patient heterogeneity[2].

Cancer occurs when genes involved in processes such as cell division, differentiation, DNA repair and apoptosis become altered, allowing cells to escape normal growth regulation and obtain increased proliferative capabilities. Some of these cancer-causing alterations can be inherited and predispose to the development of the disease, but the vast majority are somatic mutations acquired over time or under certain exposure conditions[3]. Cancer is a multistep process where genetic alterations accumulate and gradually lead to disease progression.

From a genetic perspective, tumor progression is seen as the result of a selection process where cell populations undergo a selection pressure from the microenvironment they reside in. Clones capable of proliferating and adapting under these dynamic conditions are positively selected and continue to expand. The mechanism by which these cells are able to adapt is the acquisition of "driver" mutations that directly promote malignant transformation. However, these cancer cells also acquire mutations that do not confer any direct selective advantage to tumor growth or development and are regarded as "passenger" mutations along the selection process[4].

Genes involved in cancer development

Most cancers cannot be linked to alterations occurring in a specific gene but rather to several mutations in different genes affecting select critical signaling pathways through which genes regulate cellular activities. There are fewer cancer-related pathways (~12) compared to cancer genes (~140), as mutations in members of the same biological pathway may be mutually ex-

clusive[5]. Two main types of genes contribute to carcinogenesis when mutated: oncogenes and tumor suppressor genes.

Oncogenes

Oncogenes originate from proto-oncogenes, encoding normally functioning proteins that regulate cell proliferation. Proto-oncogenes become oncogenic by the acquisition of gain-of-function mutations, which lead to the overexpression of normally active proteins or constitutive expression of overactive proteins. Oncogenes typically become activated by point mutations (e.g. *KRAS*, *BRAF* and *EGFR*) or by gene amplification *(*e.g. *HER2* and *C-MYC*). Alternatively, chromosomal rearrangements can generate fusion proteins or relocate single genes to genomic regions under the control of strong promoters (e.g. *BCR-ABL1* and *VTI1A-TCF7L2*). Regardless of the mechanism, only one activating mutation in one allele of an oncogene is necessary to promote aberrant cell growth[6].

Tumor suppressor genes

Tumor suppressor genes (TSGs) are required for appropriate cellular function as they mainly encode cell cycle regulators and apoptosis inducers. In contrast to oncogenes, TSGs are inactivated by loss-of-function (LoF) mutations. These may include point mutations as well as translocations, insertions, deletions and large chromosomal losses. Commonly, both alleles of a TSG must be inactivated to alter cellular function[7]. However, in some cases the deletion of one allele is enough to reduce the levels of gene product below the minimum required to maintain normal function (i.e. haploinsufficiency)[8].

One group of TSGs is specifically involved in the maintenance of genomic stability. These genes, known as caretaker or stability genes, encode members of DNA damage repair pathways such as base excision repair (e.g. *MUTYH*), mismatch repair (e.g. *MLH1* and *MSH2*) and double-strand break repair (e.g. *BRCA1* and *BRCA2*). Inactivation of caretaker genes occurs by similar mechanisms as with TSGs, and increases the rate at which spontaneous mutations accumulate in the cell. Consequently, the frequency at which mutations may occur in oncogenes and TSGs is also increased. For this reason, it is considered that loss of caretaker genes promotes tumor initiation in an indirect fashion[9].

Large-scale studies in cancer

Cancer is a multistep process where genetic alterations accumulate and gradually lead to disease progression. Somatic mutations present in cancer cells range from subtle changes (such as point mutations or small insertions and deletions) to more complex chromosomal aberrations. According to an analysis of 26 different cancer types, 33 to 66 genes contain somatic non-synonymous mutations, of which 95% account for single nucleotide variants (SNVs). From this group, 90.7% lead to missense changes, 7.6% to nonsense changes and 1.7% to the disruption of splice sites and untranslated regions (UTRs)[5]. The remaining source of genetic variation in the cancer genome (5%) originates from chromosomal translocations, copy number variation and insertions or deletions of large genomic regions.

When looking into the genomic alterations present in cancer, only a limited number of genes seem to be mutated with high frequencies whereas a larger fraction of genes are less commonly mutated[10]. Large-scale studies, mainly facilitated by next-generation sequencing (NGS) technologies, have progressively identified these less common alterations present across tumor types and identified which ones in particular might drive oncogenesis, disease progression and relapse. The contribution of several large-scale studies is summarized in Table 1.

Table 1. Examples of large-scale studies in cancer

Cancer type	Strategy	Main findings
Breast cancer Colorectal cancer	Exome-wide Sanger sequencing (11 + 11)	The genomic landscape of breast and colorectal cancers is composed by a few commonly mutated gene "mountains" and a large number of gene "hills" with lower mutation frequency[10].
Breast cancer	WGS (24)	Diversity of rearrangements present in the genome, >2,100 deletion, tandem duplication and amplification events. Novel in-frame fusion genes *ETV6-ITPR2*, *NFIA-EHF* and *SLC26A6-PRKAR2A*[11].
Breast cancer	WES (510)	Classification of breast cancer into subtypes based on HER2 protein expression and mutations in *GATA3*, *PIK3CA* and *MAP3K1*[12].
Breast cancer	WES (103) WGS (17)	Novel recurrent *CBFB* mutation and *MAGI3-AKT3* gene fusion[13].
Breast cancer	WES (100)	Novel recurrent mutations in *AKT2*, *ARID1B*, *CASP8*, *CDKN1B*, *MAP3K1*, *MAP3K13*, *NCOR1*, *SMARCD1* and *TBX3*[14].
Colorectal cancer	WES (224) WGS (97)	Novel recurrent mutations in *ARID1A*, *SOX9* and *FAM123B*. Gene amplification (*ERBB2* and *IGF2*) and translocation (*NAV2-TCF7L1*) events[15].
Colorectal cancer	WES (72) WGS (2)	Recurrent R-spondin fusion events (*EIF3E-RSPO2* and *PTPRK-RSPO3*)[16].

Colorectal cancer	TS (1,134)	Novel *CTNNB1* deletions and intronic *APC* splice alterations. Enrichment of oncogenic mutations based on tumor laterality[17].
Squamous cell lung cancer	WES (178) WGS (19)	Novel altered genes include *NFE2L2/KEAP1*, squamous differentiation genes and *CDKN2A/RB1*[18].
Ovarian carcinoma	WES (316)	High-grade tumors contain *TP53* mutations. Other recurrent somatic mutations in *NF1*, *BRCA1*, *BRCA2*, *RB1* and *CDK12*[19].
Melanoma	WGS (25)	Novel recurrent mutation in *PREX2*[20].
Head and neck squamous carcinoma	WES (32)	Mutations in *NOTCH1* and *FBXW7*, which are suggested to act as tumor suppressor genes[21].
Acute myeloid leukemia	WGS (8)	Two major clonal evolution patterns identified by comparing the genomes of primary tumors and relapses[22].
Acute myeloid leukemia	WGS (24)	Most mutations are caused by random events in progenitor cells and not by an initiating mutation[23].

WGS, whole genome sequencing; WES, whole exome sequencing; TS, targeted sequencing.

The landscape of cancer genomes is composed of somatic mutations that arise over time due to different processes. As perturbations causing DNA damage differ in mechanism (endogenous vs. exogenous) and duration, each tumor will have a particular spectrum of mutations. These patterns of genetic alterations can be detected by genome-scale studies and are known as "mutational signatures[4,24]. Mutational signatures are determined according to the 96 possible mutations that can occur in trinucleotides, and 21 distinct signatures have been identified to date from the sequencing of >7,000 cancers representing 30 tumor types[25]. The most frequent endogenous signatures are 1A and 1B, found in >25 different cancer types and originating from the deamination of 5-methylcytosines present in CpG sites. These signatures are represented by an increased rate of C•G to T•A substitutions[25]. The deamination of cytosine to uracil is another frequent event mediated by the DNA editing enzymes AICDA and APOBEC, involved in BER. The enrichment of thymines preceding mutated cytosines constitutes Signatures 2 and 17, present in >16 cancer types[25]. APOBEC activation may alternatively result in C•G to G•C transversions, which characterize Signature 13. Signatures 5, 12, 16 and 21 have in common the presence of A•T to G•C transitions, believed to originate by deamination of adenine to hypoxanthine (which pairs with cytosine during DNA replication) [26,27]. Signatures 8 and 18 comprise G•C to T•A mutations and result from the interaction between reactive oxygen species and DNA, generating 8-oxo-2'-deoxyguanosine which establishes hydrogen bonds with adenine[27,28]. When it comes to exogenous processes, the most common signature is 7, arising from pyrimidine dimers formed by exposure to UV light. This signature is characterized by an excess of CC•GG to TT•AA double substitutions and can account for even 25% of a melanoma's mutation burden[25,29]. A fraction of the remaining signatures is attributed to the interaction with chemical compounds that can intercalate or bind to

the DNA. Signature 11, characterized by C•G to T•A transitions, is the result of exposure to the chemotherapy agents cyclophosphamide and temozolomide[25]. Signature 4, enriched for G•C to T•A transversions, is caused by a carcinogen found in tobacco[25,30]. Finally, Signature 22 shows T•A to A•T mutations and is caused by exposure to aristolochic acid[31]. Identifying the etiology of the remaining signatures as well as which mutational processes operate in tumor cells will help understand the mechanisms of disease progression and hopefully identify potential targets for therapeutic intervention.

Genomic instability in cancer

Human cancers exhibit a trait known as genomic instability, leading to the high rate at which mutations are acquired throughout the genome compared with normal cells. The most common type of instability is known as chromosomal instability (CIN), characterized by chromosomal aberrations in both structure and number. A fraction of tumors of specific tumor types (e.g. colorectal, endometrial, head and neck and esophageal cancers)[32] show microsatellite instability (MSI), where an increased accumulation rate of base pair mutations is observed[33-35].

In some hereditary syndromes, genomic instability is caused by mutations in DNA repair genes. An example is Lynch syndrome, where mutations in DNA mismatch repair (MMR) genes trigger MSI[36]. Similarly, mutations in the breast cancer susceptibility genes *BRCA1* and *BRCA2* lead to deficiency in double-strand break repair and can lead to hereditary breast and ovarian cancers with CIN[37]. Finally, patients with *MUTYH*-associated polyposis have a deficient base excision repair (BER) system and show increased frequencies of G•C to T•A transversions that predispose for the development of colorectal cancer (CRC)[35].

Despite the advancements in the characterization of CRC tumors, the genetic alterations causing genomic instability in CIN tumors are less understood when compared with MSI cases. Genes involved in the formation of the mitotic spindle and chromosome segregation (e.g. *BUB1* and *MAD2B*) have been implicated to drive CIN in colorectal tumors, but several human cell lines that are CIN actually contain very few mutations in these genes[38,39]. Other studies have suggested that CIN might be instead triggered by the mutation of oncogenes that interfere with DNA replication forks and generate double-strand breaks[40].

Loss of heterozygosity

A common feature of cancers is aneuploidy, characterized by abnormal gains and losses in chromosomal number. These changes are thought to provide a mechanism by which gene expression can be regulated in response to the constantly changing tumor microenvironment. In most cancer types, aneuploidy is often associated with allelic imbalances arising from loss of heterozygosity (LOH) events[41]. This leads to the loss of one of the alleles initially present in the individual's normal cells, rendering the tumor cell hemizygous for such allele. In cancer, LOH frequently occurs in regions spanning TSGs that have been already inactivated by a mutation in one allele (i.e. are heterozygous) and constitutes the second inactivating event required to fully abolish gene function. The mechanism by which LOH most commonly occurs is chromosomal loss, but it may also be the result of partial deletions, unbalanced translocations, loss and reduplications and mitotic recombination (Figure 1)[42].

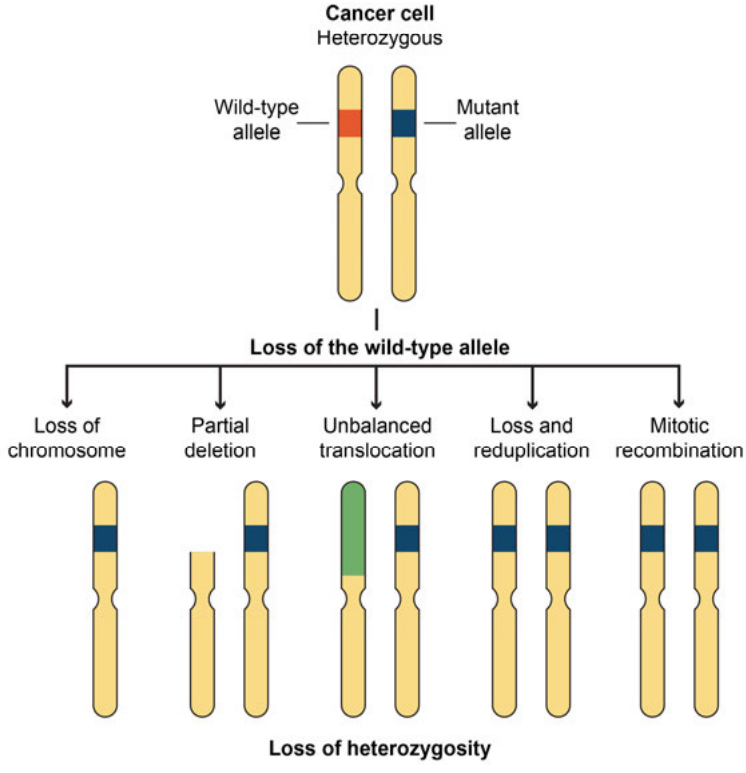

Figure 1. Mechanisms leading to loss of heterozygosity. When a cancer cell acquires its first mutation in a gene, it becomes heterozygous for that particular gene. In the context of TSGs, this mutation constitutes the first inactivating event. Loss of heterozygosity inactivates the second allele either by loss of the genomic region where the remaining functional allele is contained or by exchanging genetic material to create two copies of the initially mutated allele.

Losses due to LOH seem to be chromosome-specific and dependent on tumor type: different chromosomes are lost among cancer types, and within the same tumor some chromosomes only lose a determined region while others lose their entire arms. An explanation for this might be that allelic losses occur in particular regions where TSGs are present, and different TSGs drive the progression of different tumor types. Recent analyses suggest that LOH is the most frequent molecular alteration present in human cancers, as genomic losses affect up to 16% of the cancer genome in the average tumor (range from 1.3% in thyroid cancer to 34.4% in ovarian cancer)[43]. Further characterization of genomic regions that are prone to LOH will not only expand the map of genomic losses across tumor types but may shed light into additional genes and pathways driving tumorigenesis. In the case of CRC, genome-wide sequencing and allelotyping efforts have identified regions that are prone to genetic loss (Figure 2).

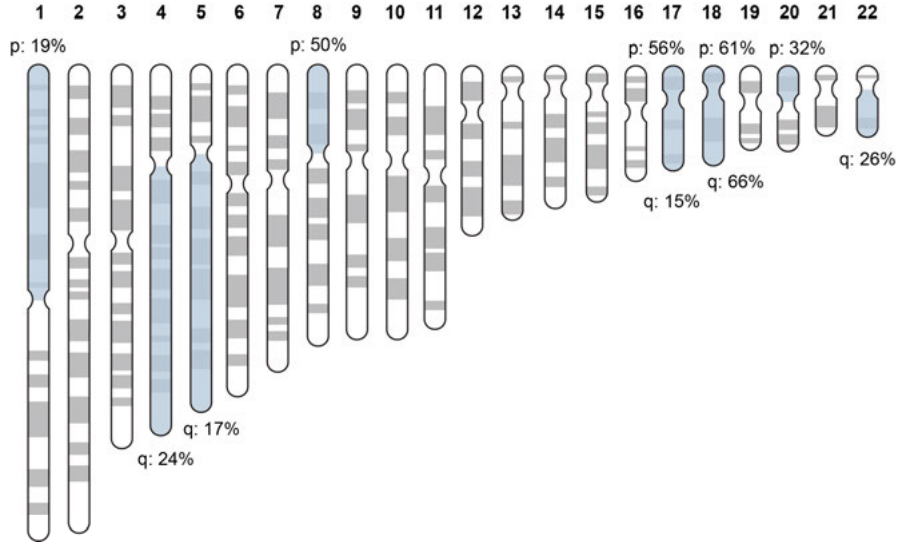

Figure 2. Map of frequent genomic losses in CRC. Chromosomal arms that are subject to genomic loss are shown in blue. Commonly deleted regions (q-value < 0.05) were obtained from TCGA's genome-scale analysis of 276 colorectal tumors[44]. The frequency at which each arm is lost is indicated for each chromosome.

Colorectal cancer

Colorectal cancer (CRC) is the third most common form of cancer in the world, with 1.2 million cases diagnosed and over 600,000 deaths reported annually[1]. The median age at diagnosis in developed countries is 70 years, and incidence strongly increases with age. Risk factors that account for the development of CRC include male gender, family history of disease, inflammatory bowel disease, smoking, high consumption of red and processed meats, as well as obesity and diabetes. In contrast, the strongest preventive factors involve physical activity, use of hormone replacement therapy, aspirin consumption and removal of precancerous lesions through endoscopy[45].

Colorectal tumors are classified according to the TNM system[46], where the local invasion depth of the tumor (T), the involvement of lymph nodes (N) and the presence of distant metastases (M) determine the stage of each patient's tumor (Table 2)[47].

Table 2. Staging of colorectal cancers according to the TNM classification system.

AJCC stage*	TNM stage			TNM stage criteria for CRC
	T	N	M	
Stage 0	Tis	N0	M0	Tis: tumor is confined to mucosa, cancer *in situ*
Stage I	T1	N0	M0	T1: invasion of *submucosa*
	T2	N0	M0	T2: invasion of *muscularis propia*
Stage II-A	T3	N0	M0	T3: invasion of *subserosa*
Stage II-B	T4	N0	M0	T4: tumor has invaded adjacent organs or perforated the visceral peritoneum
Stage III-A	T1-2	N1	M0	N1: metastasis to 1-3 regional lymph nodes. T1 or T2 present.
Stage III-B	T3-4	N1	M0	N1 and T3-4 present.
Stage III-C	Any	N2	M0	N2: metastasis to 4+ regional lymph nodes. Any T.
Stage IV	Any	Any	M1	M1: distant metastases present. Any T, any N.

*AJCC, American Joint Committee on Cancer. Together with the International Union Against Cancer (UICC), the AJCC updates and promotes the TNM classification throughout the world.

Genetics basis of colorectal cancer

Around 10-15% of patients have an inherited predisposition to develop CRC. Hereditary non-polyposis colorectal cancer (HNPCC), also known as Lynch syndrome, is the most common inherited syndrome predisposing to colorectal cancer (2-5% of all CRC cases). It is an autosomal dominant disease caused by germline mutations in MMR genes[48]. Another form of hereditary CRC is *MUTYH*-associated polyposis (MAP), caused by germline mutation in the BER gene *MUTYH*. Patients with MAP have an 80% risk of developing CRC, with a disease onset ranging between 40-60 years[49].

Predisposition to developing CRC has also been attributed to germline mutations in the DNA polymerase ε (*POLE*) and δ (*POLD1*) genes[50]. These mutations are associated with the development of multiple colorectal adenomas. Tumors arising in patients with *POLE* or *POLD1* mutations are MSI and show an enrichment for G•C to T•A and A•T to C•G transversions[51]. To date, the reported number of cases of polymerase proofreading-associated polyposis has been too few to assign this disorder an accurate clinical phenotype. Familial adenomatous polyposis (FAP) is another hereditary syndrome causing predisposition to CRC and accounts for 1% of all cases. FAP is inherited in an autosomal dominant fashion by germline mutations in the *APC* gene[52]. In patients with FAP that go on to develop CRC, *APC* becomes inactivated by a second mutation and accumulation of β-catenin triggers uncontrolled cell proliferation[53]. Patients with FAP develop hundreds of polyps during their second decade of life that accumulate across the gastrointestinal tract. If not detected early, there is a 100% risk of developing CRC 10 years after the appearance of the first polyps. The molecular cause of these CRC syndromes is known due to the high penetrance of the inherited mutations and their clinical presentation. However, the etiology of the remaining ~10% of CRCs attributed to hereditary causes is unknown. It is thought that these may be attributed to inherited syndromes caused by germline mutations in genes that show reduced penetrance or by the additive effect of alterations located in multiple susceptibility *loci*[54]. For these tumors, determining the genetic causes of inheritance will help improve surveillance and develop better diagnostic and prevention strategies.

The majority of CRC cases arise sporadically during an individual's lifetime through diverse molecular mechanisms. In 1990, Fearon and Vogelstein described sporadic CRC as a multistep genetic process, where the accumulation of mutations in both oncogenes and TSGs lead to the gradual transformation of the epithelial cells in the colon[55]. In this model, *APC* inactivation constitutes the initial event in the adenoma-carcinoma sequence. Subsequent driver events include activating mutations in members of the Ras-MAPK

and PI3K-Akt pathways as well as losses of *TP53* and other TSGs (Figure 3).

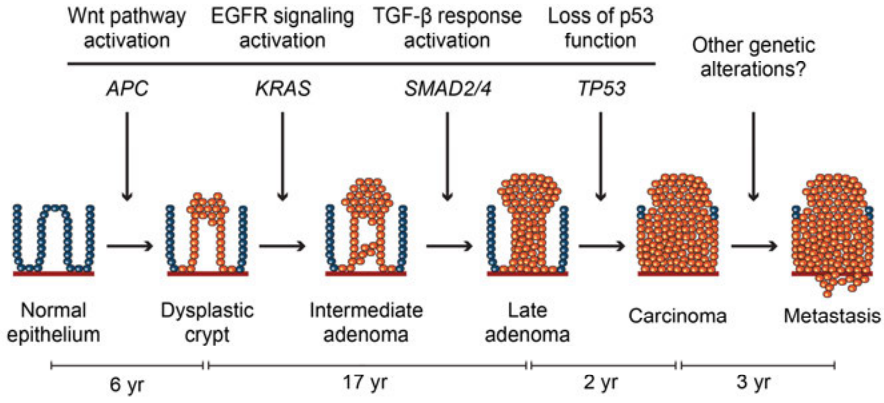

Figure 3. Simplified model of colorectal carcinogenesis. In the adenoma-carcinoma sequence, disease progression is mediated by accumulation of mutations in genes regulating proliferation and apoptosis pathways. Time estimates are shown for each stage of tumor progression. Image modified from[55-57].

Instability pathways in colorectal cancer

Colorectal tumors develop through three main instability pathways: CIN, MSI and the CpG island methylation pathway (CIMP).

Chromosomal instability pathway

The vast majority of CRCs (70-85%) are CIN, where frequent aneuploidy events and point mutations alter the function of key oncogenes and TSGs[34]. A common feature of these cancers is inactivation of *APC*, present in70-80% of CRCs[58]. Other frequent events include somatic mutation of *KRAS*, present in ~40% of all cases, and *PIK3CA* (15-25%)[59]. About 85% of tumors exhibit allelic loss in chromosomes 18 and 17p, targeting the tumor suppressors *TP53* (40-50% of CRCs) and *SMAD4* (10-15% of CRCs)[60]. Another frequently inactivated TSG is *FBXW7* (~20%)[61]. Sporadic CRCs contain an average of 80 somatic mutations per tumor, but it is only a handful of genes that are mutated in a high fraction of patients[10]. Whole-genome sequencing efforts have enabled the discovery of some of these CRC genes that are mutated at lower frequencies but equally drive tumor progression. Examples include mutation of *ARID1A*, *SOX9* and *FAM123B,* as well as amplification of *ERBB2* and *IGF2*[44]. The formation of the fusion proteins *NAV2/TCF2L1* and *VTI1A/TCF2L2* has also been reported in CRC after major chromosomal translocation events[62].

Microsatellite instability pathway

A fraction of sporadic CRCs (15-20%) have MSI, where genetic hypermutability is observed after disruption of MMR genes. Mutations tend to accumulate across the genome, where short tandem repeats are prone to insertions and deletions. When these mutations occur in protein-coding regions of the genome, frameshifts can lead to truncated proteins and functional loss. This is the mechanism by which TSGs, such as *PRDM2*, are typically inactivated in MSI tumors[63]. A common feature of MSI tumors is the presence of $BRAF^{V600E}$ mutations (found in 60% of cases as opposed to 5-10% of CIN tumors), and silencing of *MLH1* due to promoter hypermethylation[64,65]. Sporadic MSI tumors typically arise in the right side of the colon and have a particular phenotype characterized by a mucinous cell type and presence of tumor-infiltrating lymphocytes[66]. Although the spectrum of genes mutated in sporadic CRCs differs between CIN and MSI tumors, the pathways that are affected by these mutations are usually the same. An example is the absence of *SMAD4* mutations in MSI tumors, where disruption of the TGF-β signaling pathway is instead achieved by *TGFBR2* mutation[67].

CpG island methylation pathway

CRCs that also exhibit methylation in CpG islands are classified as CIMP positive and are divided into the subtypes 'CIMP high' (CIMP-H) or 'CIMP low' (CIMP-L) based on *BRAF* mutational status. CIMP tumors maintain the genetic makeup of MSI tumors, including *MLH1* gene promoter methylation. However, these tumors develop in the proximal site of the colon instead[68]. Tumors arising from serrated precursor lesions that follow the "serrated polyp" pathway usually present hypermethylation of numerous genes according to the CIMP phenotype[69].

The instability pathways observed in CRC have served as basis for different tumor classifications. The Cancer Genome Atlas (TCGA), for example, groups tumors into three groups: hypermutated (13%), ultramutated (3%) and CIN tumors (84%). Hypermutated CRCs have MSI due to MMR deficiency and *MLH1* methylation. In over 80% cases, *BRAF* V600E mutations are present. This group includes tumors with a CIMP phenotype. Ultramutated tumors contain *POLE* or *POLD1* mutations that lead to a high frequency of C to A transversions. CIN tumors have a high rate of somatic copy number alterations (SCNAs), a low mutation rate (>8/Mb), are microsatellite stable and show deregulation of the Wnt pathway (usually by *APC* mutation)[15,70]. CRCs have been alternatively classified by the Consensus Molecular Subtype (CMS) Consortium into four groups: CMS1 or MSI-immune (14%), CMS2 or canonical (37%), CMS3 or metabolic (13%) and CMS4 or mesenchymal (23%). The remaining cases (13%) are classified into a mixed

feature group. CMS1 tumors are equivalent to TCGA's hypermutated CRCs, showing MSI, *BRAF* mutation and immune infiltration. The remaining groups are sub-classifications of TCGA's CIN tumors. CMS2 tumors have a high number of SCNAs and show Wnt and Myc signaling activation. CMS3 tumors have a low number of SCNAs, are *KRAS* mutant and show metabolic deregulation. CMS4 CRCs have a high number of SCNAs and additionally present activation of the TGF-β pathway. These tumors show stromal infiltration, angiogenesis, matrix remodeling and an EMT phenotype[70,71].

A look into the classification systems available for CRC reflects the heterogeneity present between tumors. Perhaps one of the most general classification attempts is the use of mutational signatures, where CRC shows a mutational pattern (signatures 1, 5, 6 and 10) that differs from the rest of cancers[25]. Otherwise, the classification of CRC will probably increase in complexity as other aspects of tumor biology (e.g. gene expression analysis and immunologic profiling of tumors) become incorporated in the definition of distinct subtypes.

Targeted therapies in colorectal cancer

Standard clinical practice for the treatment of stages I and II CRCs comprises surgical therapy where affected sections of the bowel are resected and associated lymph nodes removed. In cancers where adjacent organs are invaded (late stage II onwards), the addition of adjuvant chemotherapy is strongly recommended as a complementary line of treatment. Patients usually receive regimens of irinotecan, oxaliplatin or 5-fluorouracil, drugs that act either as topoisomerase inhibitors, DNA replication blockers or antimetabolites. Although overall survival is increased with these chemotherapy agents, only 30-40% of patients with unresectable tumors benefit from treatment[72], and their lack of selectivity for tumor cells over normal tissues often lead to severe adverse effects. These observations suggest that chemotherapy alone is not enough to reduce tumor burden and that complementary therapy is needed.

Cancer treatment has slowly adopted the use of "targeted therapy" agents, acting upon molecular alterations arising specifically in tumors and driving disease progression. Targeted therapies are particularly valuable to patients with stage IV metastatic CRC, where tumors may no longer be resectable and treatment response is not improved by chemotherapy alone. Targeted therapy agents include both monoclonal antibodies (mAb) and small-molecule inhibitors. There are currently seven FDA-approved targeted therapies used in the treatment of metastatic CRCs, mostly in combination with chemotherapy (Table 3).

Table 3. FDA-approved targeted therapies for the treatment of CRC.

Drug	Class	Target	Regimen	Improvement
Bevacizumab	mAb	VEGF-A	IFL	OS (15.6 – 20.3 m)[73]
Cetuximab	mAb	EGFR	Irinotecan	Median time to disease progression (1.5 – 4.1 m)[74]
Panitumumab	mAb	EGFR	BSC	Reduction in the relative risk of disease (46 %)[75]
Ramucirumab	mAb	VEGF-R2	FOLFIRI	OS (11.7 – 13.3 m)[76]
Aflibercept	Fusion Ab	VEGF ligand	FOLFIRI	OS (11.9 – 13.8 m)[77]
Regorafenib	Multikinase	VEGF/TIE2	–	OS (5.0 – 6.4 m) PFS (1.7 – 1.9 m)[78]
Nivolumab	mAb	PD-1	–	OS (12.0 m) PFS (9.0 – 12.0 m)[79]

IFL, chemotherapy regimen that includes irinotecan (I), fluorouracil (F) and leucovorin (L); BSC, best supportive care; FOLFIRI, chemotherapy regimen that includes folinic acid (FOL), fluorouracil (F) and irinotecan hydrochloride (IRI); OS, overall survival; PFS, progression-free survival.

Despite the progress in the development of targeted therapies, much remains to be done to improve the clinical prognosis of patients with metastatic CRC (mCRC). Due to tumor heterogeneity, response to anti-EGFR and anti-VEGFR treatment is quite variable among patients, and in most cases leads to resistance. This may occur by the acquisition of additional mutations in members of the same signaling pathway or by the alteration of other signaling pathways with equivalent biological functions as the blocked cellular process. Resistance to anti-VEGFR in mCRCs remains poorly understood. However, the reduction of tumor vasculature levels in pancreatic cancers increases the overall levels of hypoxia and induces new pro-angiogenic growth factors like the FGF family of ligands[80]. In contrast, the mechanisms leading to resistance to anti-EGFR agents have been well characterized in CRC patients with metastatic disease. Primary relapse to cetuximab is mostly attributed to alterations in the MAPK and PI3K pathways, such as mutation of *KRAS* (~30% of cases), *BRAF* (~26% of cases), *NRAS* (~8% of cases) and *PIK3CA* (11% of cases), or amplification of *ERBB2* and *MET* (3-4% of cases)[81-84]. These mutations are usually acquired with therapy, but can also be already present in a subpopulation of resistant clones that is initially latent but expands after selective pressure of the treatment[85,86].

Although several oncogenes are known to drive CRC progression, only a few contain alterations that can be easily targeted with mAb or small molecules. In addition, there is a lack of predictive biomarkers for the selection of patients that may respond to the targeted drugs that are currently available (*KRAS* status remains as the only clinical predictor of cetuximab response). The discovery of new molecular targets of colorectal carcinogenesis as well as novel predictors of treatment response will not only improve the management of patients but also consolidate the idea of personalized therapy.

Therapeutic approaches targeting genomic loss

Targeted anti-cancer therapies have so far focused on the genetic gain-of-function arising in oncogenes that drive disease progression. Examples include cetuximab, panitumumab, erlotinib and gefitinib, agents that target the constitutively activated form of EGFR present in colorectal, pancreatic and non-small cell lung cancers[87]. Similarly, trastuzumab and lapatinib target *ERBB2* amplification driving progression in HER2-positive breast cancers. Currently available drugs neutralize ligands initiating a particular signaling pathway, occupy receptor-binding sites to prevent ligand binding, block receptor signaling inside the cell or interfere with downstream signaling molecules[87]. To deliver therapy, monoclonal antibodies and small molecule inhibitors are frequently used as agents. Since antibodies are water soluble and large (~150 kDa), they are limited to targeting extracellular components that trigger aberrant signaling. In contrast, small molecule inhibitors are able to penetrate cell membranes, being preferentially used to block receptor signaling and interfere with molecules that are downstream in the signaling pathway[88].

Whereas targeting of oncogenes has proved itself a clinically relevant strategy for the treatment of tumors, no FDA-approved therapy focusing on inactivating tumor suppressor genes has been successful. One reason for this is that it is much more feasible to block an activated oncogenic protein in the cell than to restore the function of a TSG. However, restoration of tumor suppressor function has a particular clinical relevance, as the pro-apoptotic response of TSGs triggered in the presence of DNA damaging agents could sensitize tumors to standard chemotherapy regimens. For this reason, functional loss remains a focus of targeted therapy.

Genomic loss of tumor suppressor genes

Inactivating mutations in TSGs occur in virtually all forms of cancer. To restore tumor suppressor function, gene therapy approaches have been attempted. An example is *TP53*, where gene therapy efforts include the transfer of vectors encoding wild-type protein[89-91] and T-cell immunotherapy[92]. Despite some promising results in the clinic, reduced vector efficiencies *in vivo* and limitations in the manufacture of genetically modified T-cells re-

main challenges to overcome. Alternative approaches to restore tumor suppressor function include the use of small-molecule inhibitors and screens to identify synthetic lethal partners of the lost tumor suppressor.

Pharmacological restoration of tumor suppressor function has been attempted in a limited number of TSGs by inhibiting proteins that act as their regulators. One example is the retinoblastoma 1 protein (RB1), inactivated through phosphorylation by CDK4 and CDK6 for which inhibition has been attempted. One CDK4/6-specific inhibitor, palbociclib, is currently in Phase I/II clinical trials for the treatment of breast cancer[93]. Rescue of p53 function in cancer has also been attempted, as the TSG plays a critical role in the maintenance of genomic stability. Degradation of p53 is normally mediated by the ubiquitine ligase MDM2, which suppresses its transcriptional activity by promoting proteasomal degradation. To disrupt p53/MDM2 association, small p53 peptides and small-molecule MDM2 antagonists have been tested. Some of these drugs, like MI-219, Nutlin-3, and RG7112 are undergoing Phase I/II clinical trials[94,95].

Inactivating mutations affecting TSGs create genetic vulnerabilities that are tumor-specific and could therefore have a therapeutic potential. One strategy to find such vulnerabilities is to identify genes that are synthetic lethal partners of the TSG that is lost in the tumor. A clinical translation of this approach is tumors with loss-of-function mutations in the *BRCA1* and *BRCA2* genes, which are sensitive to treatment with the poly ADP ribose polymerase (PARP) inhibitor olaparib[96]. Screens for synthetic lethal partners of TSGs have been performed with siRNA/shRNA libraries or small-molecule inhibitors. Cells with loss of *TP53* have been found to be sensitive to inhibition of the small nucleolar ribunocleoprotein (snoRNP) assembly, in particular the chaperone gene *NOLC1*[97]. Similarly, co-deletion of *RB* and the mTOR complex activator *TSC2* causes death by cellular stress[98]. Restoration of tumor suppressor function and synthetic lethality are promising concepts for therapy, but the difficulty of drugging inactivated TSGs limits the pace at which new targets emerge. This has prompted the analysis of other genes lost in cancer, questioning if they can also be of therapeutic interest.

Genomic loss of passenger genes

Due to the increased levels of aneuploidy in CIN tumors, TSGs are frequently inactivated by genomic deletion. Since the loss of tumor suppressor function is usually required to drive cancer progression, these deletions tend to be positively selected for and expand as a signature of that particular tumor. However, deletion of chromosomal regions not only affects the specific locus where the TSG is located but extends to other genes situated in close

proximity. Consequently, these "bystander" genes are also lost even though they do not directly confer any selective growth advantage to the tumor[99]. Deletions in passenger genes have been traditionally regarded as irrelevant for therapeutic purposes, but recent studies revealed that many passenger genes are actually involved in several metabolic and housekeeping functions that may render cancer cells vulnerable to therapy.

Three main approaches involving the deletion of passenger genes have been suggested for targeted therapy: collateral lethality, haploinsufficiency and allele-specific inhibition (Figure 4).

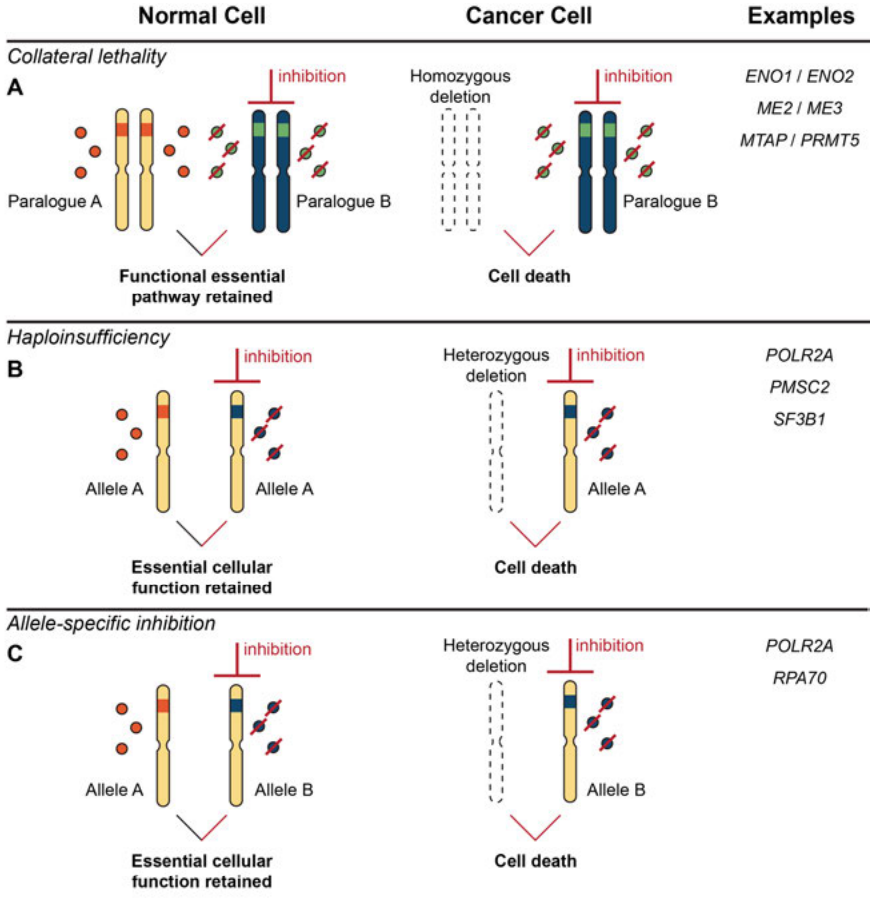

Figure 4. Therapeutic approaches targeting the deletion of passenger genes. Homozygous passenger deletions are exploited in **(A)** collateral lethality strategies. Different chromosomes are shown in yellow and blue, whereas different alleles and their encoded proteins are shown in orange and green. Heterozygous deletions are targets of **(B)** haploinsufficiency and **(C)** allele-specific inhibition. Different alleles and encoded proteins are shown in orange and blue.

Collateral lethality

This approach aims to identify synthetic lethal partners of bystander passenger genes that are homozygously deleted in cancer. In contrast to TSGs, several passenger genes have housekeeping functions and are essential for cell survival. After genomic deletion, cancer cells are still viable due to the expression of redundant paralogues that can maintain cellular function. Inhibition of the remaining paralogue would then result in tumor death. As normal cells express both paralogues, inhibition of one would still leave one member available to carry out cellular function[100].

The first study exploiting collateral lethality mined the TCGA data set for genes involved in essential cell activities which become homozygously deleted in glioblastomas. This screen identified the enolase gene *ENO1*, located in the 1p36 tumor suppressor locus which is lost in ~5% of glioblastomas. Cells with homozygous deletion of *ENO1* were highly sensitive to inhibition of the functional paralogue *ENO2* using the specific inhibitor phosphonoacetohydroxamate (PHAH)[99]. Another example of collateral lethality is the interaction between malic enzyme ME2 and its functional paralogue ME3. Mining of a cell line compendium identified losses of *ME2* to be present in over 25% of pancreatic cancers. This enzyme has a housekeeping function, as it catalyzes the conversion from malate to pyruvate that is required for NADPH regeneration in the cell. Cells with homozygous deletion of *ME2* showed sensitivity to shRNA-mediated depletion of *ME3*[101].

Collateral lethality partners are not restricted to functional paralogues but may also include genes involved in processes with biochemical redundancy. An example is *MTAP*, a gene involved in the metabolism of methionine that is frequently lost in several cancer types. Genome-scale shRNA screening data from Project Achilles identified the methylosome member *PRMT5* as a synthetic lethal partner of *MTAP*. PRMT5 catalyzes the transfer of methyl groups to arginine side chains of histones and other target proteins. In these studies, *MTAP*-null cells were shown to be sensitive to suppression of PRMT5 with the inhibitor EPZ015666[102,103].

Although more work is needed to validate drugs that inhibit functional essential pathways in a tumor-specific manner, collateral lethality is an up-and-coming technique to identify novel targets of therapeutic value in tumor types (e.g. pancreatic and gastric cancers) with limited treatment options.

Haploinsufficiency

Heterozygous deletions occur throughout the genome with higher frequency than homozygous deletions. Haploinsufficiency approaches aim to identify essential genes with heterozygous deletions (i.e. copy losses) that render cancer cells sensitive to suppression of the remaining allele. An example of a haploinsufficiency target is *POLR2A*, encoding the catalytic subunit of the RNA polymerase II complex. Due to its proximity to *TP53*, deletion of *POLR2A* occurs in ~53% of CRCs. Complete suppression of polymerase activity has been attempted *in vitro* and *in vivo* by the use of the inhibiting toxin α-amanitin, and CRC cells containing hemizygous deletion of *POLR2A* have shown 8-fold higher sensitivity to treatment when compared to *POLR2A*-intact cells[104].

For some genes with heterozygous deletions, suppression of the remaining allele leads to high levels of cellular toxicity. This group of genes has been termed CYCLOPS (copy number alterations yielding cancer liabilities owing to partial loss) and is usually enriched for members of the proteasome and spliceosome complexes. One CYCLOP gene is *PSMC2*, a member of the 19S proteasome. Cancer cells with heterozygous deletion of *PSMC2* are sensitive to shRNA-mediated inhibition of *PSMC2* when compared to neutral cells[105]. Genome-scale viability screens have identified the pre-mRNA splicing factor *SF3B1* as an additional target. Similar to *PSMC2*, human cancer cells with partial deletion of *SF3B1* are sensitive to shRNA-mediated suppression of the remaining *SF3B1* allele when compared to cells without copy number loss[43].

The discovery of bystander genes with haploinsufficiency is also relevant for therapeutic purposes, as it expands the knowledge on which pathways are required for the progression of a particular tumor type and suggests targets for inhibition.

Allele-specific inhibition

Allele-specific inhibition (ASI) is a strategy aiming to target vulnerabilities created in cancer cells that undergo LOH. Due to the high rate of SNPs present across the genome (1 per 300 bp)[106], many loci are polymorphic. If one of the alleles for a particular gene is inactivated by deletion in cancer, LOH will occur in that genomic locus. As a consequence, within the same individual, cancer cells will be hemizygous for a gene that is heterozygous in normal cells. If the affected gene is essential for cell survival, drugs inhibiting the expression of the retained allele in tumor cells will cause cell death.

In contrast, normal cells will survive as the remaining allele is capable of maintaining cellular function.

Initially, polymorphisms were targeted at the mRNA level by antisense oligonucleotides, RNA interference (RNAi) and oligonucleotide-based drugs[107]. The feasibility of ASI was first demonstrated by targeting variants of the replication protein RPA70. The *RPA70* gene is located on 17p13.3, a region undergoing frequent heterozygous deletions in cancers as it harbors *TP53*. Oligonucleotides were designed to target the two different *RPA70* alleles present in normal cells, and selective killing of cancer cells was demonstrated *in vitro* by blocking the expression of the retained allele in tumors[107]. Similarly, antisense oligonucleotides, siRNAs and EpCAM antibodies have been designed to target cancers with *POLR2A* hemizygous deletion[108-110].

Initial ASI approaches were limited by the selection of essential genes that could serve as putative targets and by the design of short nucleotide sequences that would not be degraded by cellular nucleases[107]. The emergence of large-scale sequencing studies as well as genome-wide RNAi and CRISPR/Cas9 screens has expanded the repertoire of genes that provide essential cellular functions. In parallel, catalogues of genetic variation have improved the characterization of polymorphisms that are frequent in the human population. With these new tools for analysis, it is likely that other ASI targets emerge in the future.

Together with ASI, collateral lethality and haploinsufficiency approaches have proven that deleterious deletion of passenger genes may also influence tumor progression, expanding the number of actionable targets for the treatment of cancer.

Present investigations

Aims

The work presented in this thesis explores the consequences of functional loss affecting tumor suppressor genes (**Paper I**) and passenger genes (**Papers II-IV**) in CRC, with the aim of identifying vulnerabilities that can be exploited for therapy. A complementary characterization of colorectal tumors is presented (**Paper V**) as an attempt to understand the genetic causes leading to genomic instability in cancer. The specific aims of each paper are as follows:

I Model the c.4467delA mutation inactivating *PRDM2* in colorectal tumors by genome editing and characterize the contribution of such alterations to tumorigenesis.

II Determine whether the reduced allelic diversity present in cancer cells undergoing LOH relative to the remaining healthy tissues of the body can be exploited for therapy. Identify putative targets for such an approach by mapping prevalent nsSNVs frequently lost in cancer.

III Define the cohort of eligible patients that could benefit from NAT2 allele-selective chemotherapy in the context of colorectal cancer.

IV Extend the search of therapeutic LOH targets by identifying prevalent loss-of-function (LoF) alleles retained in tumors undergoing LOH.

V Characterize a set of 56 colorectal cancers by whole genome sequencing, focusing on mutation frequency and mutational signatures present in these tumors.

Loss of PRDM2 and its contribution to colorectal carcinogenesis

Almost 15% of colorectal cancers show microsatellite instability (MSI) due to deficiencies in mismatch repair genes. A direct consequence of defective DNA repair is the accumulation of alterations throughout the genome, especially in short tandem repeats. However, the identification of which of these alterations contributes to cancer cell growth remains challenging. In **Paper I** we have focused on a particular frameshift deletion mutation (c.4467delA) occurring in the (A)9 repeat of exon 8 in *PRDM2*. This mutation is observed in ~30% of MSI CRCs and leads to functional inactivation of the tumor suppressor gene.

To determine the effects of *PRDM2* loss in MSI CRCs, we engineered isogenic cell systems where one mutant allele of *PRDM2* was restored in HCT116 cells (homozygous for c.4467delA) by rAAV-mediated genome editing. To test if PRDM2 activity was successfully reestablished after gene editing, we measured the levels of histone 3 lysine (H3K9) methylation across the genome. By ChIP-seq, we observed that levels of H3K9 dimethylation were indeed enriched by 3-fold in the genic regions of restored clones, and in particular at the start of transcriptional sites.

To assess the impact of restored PRDM2 function on cancer phenotypes, we compared the migration and anchorage-independent growth properties of parental HCT116 cells and the edited clones. Cells with PRDM2 activity showed impaired growth under reduced serum conditions and a 3-fold reduction in anchorage-independent growth and migration capabilities when compared to parental cells. When the effects of *PRDM2* restoration were tested i*n vivo*, a 2-fold decrease in growth rate was observed for tumors with corrected PRDM2 activity when compared to regular tumors. These observations point to loss of one *PRDM2* allele in HCT116 cells being enough to promote cell growth and invasiveness.

To understand the effects of *PRDM2* loss in CRC, we performed RNA sequencing of HCT116 cells and clones with restored tumor suppressor activity. Transcriptome analysis revealed a limited number of genes that become deregulated after PRDM2 restoration, and in particular enrichment for genes involved in epithelial-to-mesenchymal transition (EMT).

Paper I confirmed the role of *PRDM2* as a tumor suppressor gene and provided direct evidence for the c.4467delA mutation being a driver event in colorectal cancer.

Targeting allelic loss of bystander genes in cancer

Throughout disease progression, tumors are prone to lose genomic material and undergo loss of heterozygosity (LOH). In the context of tumors that express one functional and one deficient allele for a particular enzyme, loss of the functional allele may render cancer cells sensitive to therapy with cytotoxic substrates that cannot longer be processed (Figure 5). **Paper II** and **Paper IV** in this thesis cover heterozygous deletions affecting bystander drug-metabolic genes.

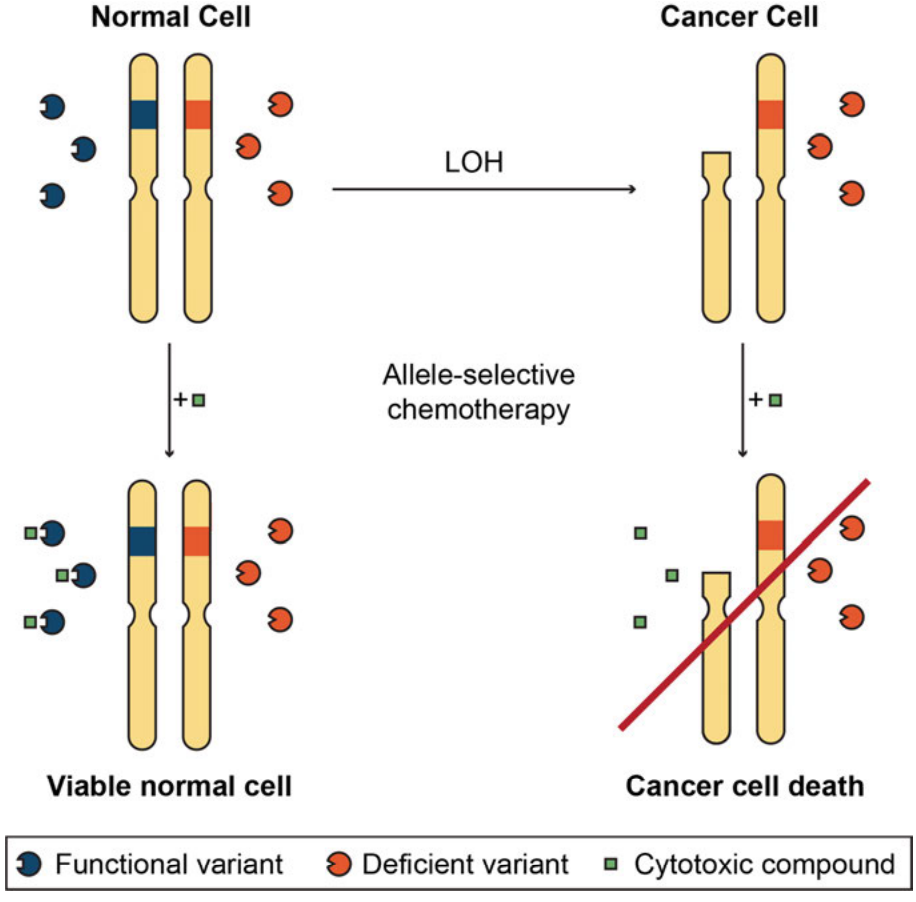

Figure 5. Exploiting loss of heterozygosity for allele-selective chemotherapy. Eligible patients are heterozygous for an enzymatic gene, encoding a functional variant (blue) and a variant deficient in activity (orange). During cancer progression, cancer cells may undergo LOH and lose the functional allele (blue). Treatment with a cytotoxic compound (green) that can only be processed by the functional enzymatic variant will selectively kill tumor cells.

Allelic loss of NAT2

In **Paper II** we developed a bioinformatic strategy to identify prevalent nsSNVs found in the human population and located in regions frequently lost in common cancers. We identified 45 nsSNVs near the catalytic sites of 17 enzymes, and for proof-of-concept selected rs1799930 in the metabolic enzyme NAT2 as a target for CRC. We then developed cell models of the rapid (*NAT2**13; wild-type) and slow (*NAT2*6*) NAT2 activity variants resulting from the chosen nsSNV, as such cell systems could be used for screening compounds that selectively kill NAT2-deficient tumors. To achieve this, stable clones in the CRC cell lines RKO and DLD-1 were generated by stable transfection of overexpression vectors encoding the *NAT2*13* and *NAT2*6* variants. Detection of recombinant protein and catalytic activity in the generated clones validated the use of the system in a phenotypic screening assay.

For screening purposes, 176 potential NAT2 substrates were filtered from a library of small-molecules based on their chemical structure. Cells were incubated with different concentrations of each compound and cell viability was measured after 72 h in an MTT assay. From this set, one molecule called APA showed selective cytotoxicity towards tumor cells with reduced NAT2 activity. Validation of the hit compound *in vitro* was performed by establishing full-dose response curves in clones with high and low levels of NAT2 activity, assessing binding specificity to human NAT recombinant proteins and performing ADME profiling. To identify putative targets of APA, a kinome screening was performed assessing the binding capability of APA and its acetylated form NAPA to a panel of 468 protein kinases. The top candidate from the screen was Aurora kinase A (AURKA), a regulator of cell cycle progression that is required for correct spindle formation and segregation. Upon APA treatment, DLD-1 cells with slow NAT2 activity showed a ~40% reduction in phosphorylated AURKA (pAURKA) and increased levels of the mitotic marker histone H3 pSer 10. About 15% of these cells were in mitosis, where the formation of multipolar spindles could be seen. In contrast, cells with rapid NAT2 activity did not show alteration of pAURKA levels and the fraction of mitotic cells remained <5%.

The anti-tumor effects of APA were assessed *in vivo* in BALB/c nude mice with xenografts of tumor cells expressing rapid and slow NAT2 activity variants. For enhanced tumor targeting, APA was encapsulated in liposomes prior delivery via intravenous injection. After 5 days, tumor growth in NAT2-deficient tumors was reduced by 50% as compared to tumors with NAT2 wild-type activity. Further analyses in patient material revealed that

response to APA treatment is dependent on NAT2 status in the patient's primary tumor.

The findings in **Paper II** proved the existence of chemotherapies targeting loss of heterozygosity and demonstrated that bystander mutations occurring in drug-metabolic genes can modulate tumor response.

Defining a cohort eligible for NAT2 allele-selective chemotherapy

Paper II demonstrated that LOH-based targeting of NAT2 represents a new avenue for colorectal cancer therapy. Eligible patients for APA treatment are initially heterozygous for *NAT2*, carrying one rapid acetylator allele and one slow acetylator allele (i.e. intermediate acetylators). In addition, their tumors lose the functional *NAT2* allele due to LOH and become deficient in enzymatic activity. As *NAT2* is a highly polymorphic gene, there are several allele groups (*NAT2*5*, *NAT2*6*, *NAT2*7* and *NAT2*14*) that encode enzymatic variants with slow acetylator phenotypes[111]. So far, the effect of APA treatment has only been evaluated in tumors that retain the *NAT2*6* allele. In **Paper III** we assessed the response of the remaining slow acetylator variants to APA treatment, aiming to define the cohort of patients that can benefit from targeting LOH in NAT2.

Initial mining of the 1000 Genomes dataset revealed that 35.8% of the world population is heterozygous for *NAT2*, encoding one rapid and one slow activity variant. The haplotypes for these intermediate acetylators are divided in 15.1% (*Rapid*/*NAT2*5*), 13.4% (*Rapid*/*NAT2*6*), 5.7% (*Rapid*/*NAT2*7*) and 1.6% (*Rapid*/*NAT2*14*). Given the frequency at which colorectal tumors undergo LOH (~21%[112]) and the yearly incidence of the disease (~1.2 million cases[1]), we estimated that over 98,000 individuals could theoretically be eligible for allele-selective chemotherapy every year.

To confirm the proportion of intermediate acetylators present in the CRC patient population and the frequency at which LOH events occur in the *NAT2* locus, we first genotyped two independent CRC patient cohorts and matched normal samples (T/N pairs) using genomic arrays and next-generation sequencing. The identified NAT2 haplotypes and LOH events are summarized in Table 4.

Table 4. NAT2 slow acetylator frequencies and LOH events detected in two independent CRC cohorts.

Cohort size and ethnicity	Frequency of slow acetylator alleles	Intermediate acetylators	LOH events detected	Tumors losing the wild-type *NAT2* allele
1,511 T/N pairs (Finnish)	*NAT2*5* (45.9%) *NAT2*6* (23.5%) *NAT2*7* (4.1%) *NAT2*12* (<0.1%)	37%	30.9%	24% (of total)
56 T/N pairs (Swedish)	*NAT2*5* (44.6%) *NAT2*6* (31.3%) *NAT2*7* (4.5%)	26%	8.9%	1.8% (of total)

Given the frequency at which other slow acetylator variants (apart from *NAT2*6*) contribute to the fraction of NAT2 intermediate acetylators, we sought to evaluate whereas tumors retaining these low activity alleles after LOH are responsive to APA treatment. Stable clones expressing the NAT2 slow variants *NAT2*5*, *NAT2*7* and *NAT2*14* were generated by transduction of RKO cells and validated by immunoblotting. We first assessed the NAT2 catalytic activity of each clone towards APA, and compared it to RKO cells engineered to express the wild-type *NAT2*4* allele. Impaired affinity towards APA was observed for *NAT*5* and *NAT2*14* variants (K_m= 4.38 and 16.11), where NAPA formation was 10-fold less compared to the wild-type clone. In contrast, clones expressing the *NAT2*7* allele had an increased affinity towards APA (K_m= 0.42) and showed substrate inhibition kinetics assimilating those of the NAT2 wild-type clone. The differential levels of APA acetylation had an impact on cell viability, where the growth of *NAT2*5* and *NAT2*14* clones was 4-fold less than wild-type cells upon treatment (IC_{50}= 0.86 μM and 0.15 μM). The growth of *NAT2*7* cells was almost unaltered during the course of the experiment (IC_{50}= 3.46 μM). These results suggested that intermediate acetylators with the haplotypes *Rapid/NAT2*5* and *Rapid/NAT2*14* would also benefit from NAT2 allele-selective chemotherapy, modifying the estimation of eligible patients to ~79,000 individuals per year.

We finally developed a clinically applicable method for detecting *NAT2* haplotypes in a multiplexed fashion. The assay was tested in DNA from different cancer cell lines, where the *NAT2* region was amplified with primer pairs containing unique barcodes. The resulting PCR products were quantified and pooled in equimolar ratios for multiplexed single molecule real-time (SMRT) sequencing. In our analysis, 99.8% of the sequencing reads were correctly aligned to the *NAT2* reference sequence and enough to resolve the haplotype of each sample.

In **Paper III** we have determined the cohort of CRC patients that can benefit from NAT2 allele-selective chemotherapy and developed a method for the clinical detection of *NAT2* allelic loss.

Allelic loss of CYP2D6

In **Paper IV,** we expanded the search for putative LOH targets by including the set of LoF variants retained in tumors after allelic loss. These included SNVs causing protein truncation through base substitution, indels and splice site disruptions. By mining the 1000 Genomes dataset, we were able to discard variants located in non-coding genomic regions and with allele frequencies under 0.5%. Our initial set included 449 SNVs causing premature stop codons, 565 indels causing frameshift or truncation and 384 mutations disrupting splice sites. In comparison to previous genome-wide analysis of LoF variants[113], we identified 60.5% of deleterious variants in our set that had not been previously described. After additional filters considering heterozygosity frequency, gene expression and LOH occurrence of each variant, we identified a final set of 60 LoF variants located in 60 candidate genes. Taking into consideration the role of xenobiotic enzymes in cancer epidemiology, we chose the splice site variant rs3892097 in the drug-metabolic enzyme CYP2D6 as a priority target. CYP2D6 is a highly polymorphic enzyme known to metabolize ~25% of clinically used drugs[114]. The chosen SNV has a heterozygous prevalence of 16% in the human population and disrupts the acceptor splice site located at the start of exon 4 in the gene.

As in **Paper II**, we aimed to generate cell models expressing the presence of functional and mutated CYP2D6, to reflect what happens in the normal and tumor cells of a patient. However, a challenge in this study was to predict the effect of the chosen splice site disruption and generate cells expressing an accurate representation of the *CYP2D6*4* transcript. We hypothesized that the SNV creates a cryptic acceptor site at the start of exon 4, causing a frameshift that affects *CYP2D6* transcription and splicing. The resulting aberrant transcript lacks 1 bp in exon 4 and is translated into a truncated protein of 181 amino acids in length.

Generation of stable clones with ectopic expression of wild-type CYP2D6 (*CYP2D6*1*) and mutant *CYP2D6*4* (resulting from rs3892097) was pursued by transduction of human HepG2 cells with lentiviral particles. Validation of the obtained clones included detection of recombinant CYP2D6 by immunoblotting and quantification of catalytic activity by monitoring dextromethorphan metabolism in cells with functional and inactive enzyme.

In **Paper IV** we have extended the list of prevalent LoF alleles present in the population and generated cell model systems to validate CYP2D6 as a target for LOH therapy.

Characterization of colorectal cancers by whole genome sequencing

The mutational landscape of CRCs is composed of a few genes that are frequently mutated and a rather larger fraction of genes that become mutated at lower frequencies. Despite the advances of genome-scale studies, the contribution of the latter group of genes to colorectal tumorigenesis remains poorly understood. In **Paper V** we have performed whole genome sequencing of 56 microsatellite-stable colorectal tumors and patient-matched normal samples, aiming to further characterize CRC in terms of mutational frequencies and signatures.

Sequencing of matched tumor and normal samples was achieved with a mean coverage of 90.1× and 88.1×. An average of 16,953 mutations was detected per sample, with a mutation rate of 5.2 mutations per megabase. Out of this set, a median of 89 mutations were non-synonymous in each tumor, out of which missense mutations had the biggest contribution. The highest mutation frequencies were found in the known driver genes *TP53* (77%), *APC* (73%), *KRAS* (39%), *BRAF* (14%) and *PIK3CA* (14%). To determine whether the somatic mutations found in these CRC genes had been previously reported, we compared our set to the Catalogue of Somatic Mutations in Cancer (COSMIC). High concordance was reported for *PTEN* (100%), *KRAS* and *TP53* (~81%) mutations, including nsSNVs and frameshift deletions. For genes such as *APC*, *BRAF* and *PIK3CA*, the percentage of matching mutations ranged 65-70%. No concordance was found for *SMAD4* mutations between our set and COSMIC.

To gain insight into the mutational processes of our tumor set, we performed context-dependent analysis of all SNVs in each sample to determine the presence of mutational signatures. Based on the cophenetic-correlation coefficient, we identified signatures 1, 3 and 17 to be present in the dataset. The relative contribution of each signature differed between samples, with Signatures 1 and 3 present in all tumors and Signature 17 dominant in 10% of the set. This was an interesting observation, as the etiology of Signature 17 is unknown and has not yet been reported in CRC[115].

In **Paper V** we have verified the mutation frequencies of known CRC genes and identified for the first time a strong contribution of an unknown base level mutational process in 10% of the analyzed tumors.

Concluding remarks and future perspectives

The earliest descriptions of cancer therapy date from 460 BC, where Hippocrates registered aberrations of the skin and selected treatment based on the patient's humor. For the next two millennia, and until the discovery of cells in the 19th century, cancer therapy relied on diet, blood-letting and laxatives. Much has been learned about the molecular mechanisms of cancer since then, and therapies have adopted the use of surgery, chemotherapy and radiotherapy for the treatment of tumors.

Cancer remains a health threat in the world, particularly in developing countries where 80% of patients have a late stage incurable disease by the time of diagnosis[116]. Detection of cancers at earlier stages therefore needs to improve. Genome-wide sequencing efforts have expanded the mutational landscape of CRCs, and advancements have been made in the detection of these alterations both in primary tumors and metastasis. An example is the use of liquid biopsies as a non-invasive way of characterizing primary tumors. Methods to sequence circulating tumor DNA (ctDNA) have progressively improved the sensitivity and specificity at which a panel of oncogenic mutations is detected in tumor cells. An exciting achievement is the development of multi-analyte blood test for the detection of common cancers with high sensitivity and specificity[117]. This tool detects very low numbers of mutant ctDNA molecules in plasma, allowing cancers to be detected at earlier stages than previously possible. In addition, the complementary analysis of circulating proteins helps predict the tissue of origin. This is particularly valuable for patients suffering of ovary, liver, stomach, pancreas and esophagus cancers, as no screening tests are available. For tumor types where early screening methods are clinically established (such as CRC), this test constitutes an improved option at a comparable cost.

We now understand that different genetic makeups determine the mechanisms through which disease progression is fueled and that a handful of genes involved in these pathways can be targeted for therapy. Despite the advancements in tumor characterization, the road from genes to therapy is limited by the pace at which drug discovery efforts can identify molecules to target these genetic alterations. Although a vast number of oncogenes driving cancer progression are known, a substantial group of these have alterations that cannot be easily targeted by any available drug. An example is

KRAS, the most frequently mutated oncogene in colorectal, pancreatic and primary lung tumors[10,118,119]. The RAS family of GTPases is not suitable for direct therapeutic inhibition, as they do not contain accessible active sites of pockets that can be blocked with small molecules[120]. Consequently, new therapeutic approaches focus on the inhibition of downstream signaling effectors instead. For the fraction of CRCs where treatment is available, many patients still die from the intrinsic and acquired mechanisms of resistance that resume tumor progression and metastasis. This clinical reality reflects the acute need for additional targets in the treatment of CRC.

This thesis aimed to identify cancer vulnerabilities that can be exploited for CRC therapy. We first focused on characterizing the functional effects of tumor suppressor loss in MSI tumors, a fraction of CRCs that are therapeutically interesting for their better prognosis in comparison to CIN tumors[121]. We identified *PRDM2* loss as a driver event of MSI CRCs and showed that deleterious mutations in this TSG enable tumor growth and increased migration capabilities. The approaches we have developed can be extrapolated to study the effects of frameshift mutations occurring in >65% repeats mutated in MSI cancers. Doing so may enable the identification of other genes involved in colorectal carcinogenesis.

Designing drugs with both sensitivity and specificity for tumor cells remains a challenge, and is particularly difficult in the setting of restoring tumor suppressor function. We therefore analyzed the effects of allelic loss targeting passenger genes that get co-deleted with TSGs in CIN tumors. This strategy is conceptually similar to the emerging collateral lethality and haploinsufficiency strategies targeting loss of bystander genes. In this thesis we have proved for the first time that cancer chemotherapies targeting LOH are theoretically feasible, further demonstrating that the therapeutic index of drugs can be actually modified by mutations occurring in bystander metabolic genes. The xenobiotic enzymes NAT2 and CYP2D6 were chosen as priority targets in this work, but we have additionally identified variants in over 70 candidate genes that are frequent in heterozygous state and commonly lost in cancer. Considering the incidence of common cancer types, this translates into hundreds of thousands of patients benefiting from treatment every year. In the particular case of *NAT2* allelic loss, we have demonstrated (by genotyping efforts and functional studies) that ~79,000 CRC patients are eligible for allele-selective chemotherapy. These studies are particularly relevant as they target a population of CRC patients that currently might not benefit from any other treatment option.

Genome-scale sequencing efforts can accelerate drug discovery by identifying other genes and pathways disrupted in cancer. In the context of colorectal tumors, the contribution of genes with low mutation frequency is poorly

understood and might open novel avenues for targeted therapy. To further characterize the mutational landscape of CRC, we have performed deep sequencing of CIN tumors and identified a strong contribution of an unknown mutational process in 10% of cases. This observation may pave the road to identify key promoters of genomic instability in this cancer type.

Acknowledgements

The work presented in this thesis was carried out at the Department of Immunology, Genetics and Pathology, Uppsala University. I would like to express my gratitude to some people who have made this PhD journey possible:

First of all, I would like to thank my supervisor **Tobias Sjöblom** for giving me the opportunity to start my graduate studies in his group. Tobias, I appreciate the trust that you have put in me during these years and I am particularly grateful for the tools that you have given me to become a more independent and critical scientist along the way. I also thank you for the encouragement to communicate my research ideas both in the academic and industrial world, it has been such a rewarding experience!

I would like to thank my co-supervisor **Tanja** for the valuable input to this thesis. Tanja, I admire how you explain the most complicated aspects of cancer with such ease and I am grateful that you have shared this knowledge with me. Also, thank you for teaching me how to work with cells during my first year! I also want to thank my other co-supervisor **Lena** for her support, the input to this thesis work, and for encouraging me to go to Singapore to do experiments for one of the papers!

To the members of the Molcan group: **Artur**, **Chatarina**, **Erik**, **Ivaylo**, **Klara**, **Luis**, **Natallia**, **Snehangshu**, **Teymur**, **Tom**, **all students** and **former colleagues**. You have no idea of how much I have learned from all of you, I feel so lucky to have worked in such a multicultural and stimulating environment! Thank you for the good talks and tasty fikas. I would like to thank in particular the colleagues I have worked closely with: **Ivaylo**, thank you for training me during the first years of my PhD and for becoming such a caring colleague afterwards. I will for sure remember some of our not-so-short conversations. **Lucy**, I am so grateful for what you have done for me during these years. Thank you for all the time you have spent proofreading and improving this thesis, for your always-sincere advice, for being a constant source of support and for making sure that my stomach and back also made it in one piece! **Natallia**, your kindness and the effort you put into everything you do have made these last months of thesis hysteria much more manageable. **Snehangshu**, your technical skills have saved some of our pro-

jects (and even more grey hairs from popping up)! Thank you for all your hard work and for being such a fun office mate. **Viktor**, thank you for the late-night talks and for the constant support. I am so glad we got to co-author a paper and discuss 'science' together!

Thank you to all the collaborators in Sweden, Finland and Singapore that have enabled this work, and in particular to **Ian** for being such an attentive host and for providing me with first-hand in vivo training. I would also like to thank the Uppsala Genome Center and the SNP&SEQ facility for providing their services.

A big thank you to **Christina Magnusson** and **Helene Norlin** from the IGP Administration and to **Karin Forsberg Nilsson** for all the support they provided me and all of the PhD students at the Department. Thank you to the members of the PhD Council: **Anja M**, **Bene**, **Dijana**, **Emma Y**, **Johanna M**, **Leonor**, **Linnea**, **Loora**, **Mohan**, **Marcus**, **Sanna**, **Sara** and **Svea**. It has been a pleasure working with you organizing conferences, seminars and other less scientific activities. To the teaching crew: **Ross**, **Leonor** and **Yang**. Who would have thought our PhD included a "minor" in forensics?

Thank you **Chiara**, **Leonor** and **Svea** for being my sisters away from home. You are fantastic friends and scientists! To all my wonderful colleagues at Rudbeck and BMC, in particular **Andrew**, **Alba**, **Anja N**, **Anna B**, **Carmen**, **Diego**, **Dominic**, **Elin A**, **Elin S**, **Elisabeth**, **Emma G**, **Emma Y**, **Frank**, **Inés**, **Karin**, **Leire**, **Luuk**, **Maike**, **Marco**, **Maria P**, **Maria R**, **Mark**, **Matko**, **Miguel**, **Ross**, **Sara B**, **Suomi**, **Tor**, **Tsong** and **Yanara** for all the fun in and (especially) outside work! To **Aleksi**, **Erik O**, **Elin**, **Fiona**, **Johanna L** and **Malin** for reminding me that there is also a world outside the lab. To my Venezuelans in Sweden –**Elio**, **Gabriel**, **Harold**, **Mariana**, **Oscar** and **Rodrigo**– for keeping me close to the traditions I miss the most from my country. To my **doñas**, who have proven that not even 8,691 kilometers can end our friendship. To **Alessandra**, **Eugenia**, **Katty**, **Mariana** and **Pedro**, for being supportive of my PhD throughout these years.

To my dearest **Eric**, you and **your family** warm up my life in Sweden. When I first started my PhD I never imagined that a colleague would turn into my best friend and then into the most loyal companion one could ever wish for. This PhD would not have been possible without your encouragement and teamwork!

To my family –**Mamá**, **Papá**, **Elisa** and **Nacho**–, for being a constant source of motivation. Life is tough without you by my side but the values you have instilled in me are a daily reminder of our enduring bond. Thank you for always believing in me and for enabling me to follow my dreams.

References

1. Torre, L. A., Siegel, R. L., Ward, E. M. & Jemal, A. Global Cancer Incidence and Mortality Rates and Trends--An Update. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology* **25**, 16-27, doi:10.1158/1055-9965.Epi-15-0578 (2016).
2. Vogelstein, B. & Kinzler, K. W. Cancer genes and the pathways they control. *Nature medicine* **10**, 789-799, doi:10.1038/nm1087 (2004).
3. Weinberg, R. A. How cancer arises. *Scientific American* **275**, 62-70 (1996).
4. Stratton, M. R., Campbell, P. J. & Futreal, P. A. The cancer genome. *Nature* **458**, 719-724, doi:10.1038/nature07943 (2009).
5. Vogelstein, B. *et al.* Cancer genome landscapes. *Science (New York, N.Y.)* **339**, 1546-1558, doi:10.1126/science.1235122 (2013).
6. Giordano, A. & Lee, W. H. Preface. *Oncogene* **25**, 5189-5189 (2006).
7. Knudson, A. G., Jr. Mutation and cancer: statistical study of retinoblastoma. *Proceedings of the National Academy of Sciences of the United States of America* **68**, 820-823 (1971).
8. Berger, A. H., Knudson, A. G. & Pandolfi, P. P. A continuum model for tumour suppression. *Nature* **476**, 163-169, doi:10.1038/nature10275 (2011).
9. Kinzler, K. W. & Vogelstein, B. Cancer-susceptibility genes. Gatekeepers and caretakers. *Nature* **386**, 761, 763, doi:10.1038/386761a0 (1997).
10. Wood, L. D. *et al.* The genomic landscapes of human breast and colorectal cancers. *Science (New York, N.Y.)* **318**, 1108-1113, doi:10.1126/science. 1145720 (2007).
11. Stephens, P. J. *et al.* Complex landscapes of somatic rearrangement in human breast cancer genomes. *Nature* **462**, 1005-1010, doi:10.1038/nature08645 (2009).
12. Comprehensive molecular portraits of human breast tumours. *Nature* **490**, 61-70, doi:10.1038/nature11412 (2012).
13. Banerji, S. *et al.* Sequence analysis of mutations and translocations across breast cancer subtypes. *Nature* **486**, 405-409, doi:10.1038/nature11154 (2012).
14. Stephens, P. J. *et al.* The landscape of cancer genes and mutational processes in breast cancer. *Nature* **486**, 400-404, doi:10.1038/nature11017 (2012).
15. Comprehensive molecular characterization of human colon and rectal cancer. *Nature* **487**, 330-337, doi:10.1038/nature11252 (2012).
16. Seshagiri, S. *et al.* Recurrent R-spondin fusions in colon cancer. *Nature* **488**, 660-664, doi:10.1038/nature11282 (2012).
17. Yaeger, R. *et al.* Clinical Sequencing Defines the Genomic Landscape of Metastatic Colorectal Cancer. *Cancer cell* **33**, 125-136.e123, doi:10.1016/ j.ccell.2017.12.004 (2018).
18. Comprehensive genomic characterization of squamous cell lung cancers. *Nature* **489**, 519-525, doi:10.1038/nature11404 (2012).

19. Integrated genomic analyses of ovarian carcinoma. *Nature* **474**, 609-615, doi:10.1038/nature10166 (2011).
20. Berger, M. F. *et al.* Melanoma genome sequencing reveals frequent PREX2 mutations. *Nature* **485**, 502-506, doi:10.1038/nature11071 (2012).
21. Agrawal, N. *et al.* Exome sequencing of head and neck squamous cell carcinoma reveals inactivating mutations in NOTCH1. *Science (New York, N.Y.)* **333**, 1154-1157, doi:10.1126/science.1206923 (2011).
22. Ding, L. *et al.* Clonal evolution in relapsed acute myeloid leukaemia revealed by whole-genome sequencing. *Nature* **481**, 506-510, doi:10.1038/nature10738 (2012).
23. Welch, J. S. *et al.* The origin and evolution of mutations in acute myeloid leukemia. *Cell* **150**, 264-278, doi:10.1016/j.cell.2012.06.023 (2012).
24. Nik-Zainal, S. *et al.* Mutational processes molding the genomes of 21 breast cancers. *Cell* **149**, 979-993, doi:10.1016/j.cell.2012.04.024 (2012).
25. Alexandrov, L. B. *et al.* Signatures of mutational processes in human cancer. *Nature* **500**, 415-421, doi:10.1038/nature12477 (2013).
26. Lindahl, T. Instability and decay of the primary structure of DNA. *Nature* **362**, 709-715, doi:10.1038/362709a0 (1993).
27. Helleday, T., Eshtad, S. & Nik-Zainal, S. Mechanisms underlying mutational signatures in human cancers. *Nature reviews. Genetics* **15**, 585-598, doi:10.1038/nrg3729 (2014).
28. Oikawa, S., Tada-Oikawa, S. & Kawanishi, S. Site-specific DNA damage at the GGG sequence by UVA involves acceleration of telomere shortening. *Biochemistry* **40**, 4763-4768 (2001).
29. Hendriks, G. *et al.* Transcription-dependent cytosine deamination is a novel mechanism in ultraviolet light-induced mutagenesis. *Current biology : CB* **20**, 170-175, doi:10.1016/j.cub.2009.11.061 (2010).
30. Pfeifer, G. P. *et al.* Tobacco smoke carcinogens, DNA damage and p53 mutations in smoking-associated cancers. *Oncogene* **21**, 7435-7451, doi:10.1038/sj.onc.1205803 (2002).
31. Feldmeyer, N. *et al.* Further studies with a cell immortalization assay to investigate the mutation signature of aristolochic acid in human p53 sequences. *Mutation research* **608**, 163-168, doi:10.1016/j.mrgentox.2006.02.017 (2006).
32. Lawes, D. A., SenGupta, S. & Boulos, P. B. The clinical importance and prognostic implications of microsatellite instability in sporadic cancer. *European journal of surgical oncology : the journal of the European Society of Surgical Oncology and the British Association of Surgical Oncology* **29**, 201-212 (2003).
33. Lengauer, C., Kinzler, K. W. & Vogelstein, B. Genetic instability in colorectal cancers. *Nature* **386**, 623-627, doi:10.1038/386623a0 (1997).
34. Lengauer, C., Kinzler, K. W. & Vogelstein, B. Genetic instabilities in human cancers. *Nature* **396**, 643-649, doi:10.1038/25292 (1998).
35. Al-Tassan, N. *et al.* Inherited variants of MYH associated with somatic G:C-->T:A mutations in colorectal tumors. *Nature genetics* **30**, 227-232, doi:10.1038/ng828 (2002).
36. Leach, F. S. *et al.* Mutations of a mutS homolog in hereditary nonpolyposis colorectal cancer. *Cell* **75**, 1215-1225 (1993).
37. Ripperger, T., Gadzicki, D., Meindl, A. & Schlegelberger, B. Breast cancer susceptibility: current knowledge and implications for genetic counselling. *European journal of human genetics : EJHG* **17**, 722-731, doi:10.1038/ejhg.2008.212 (2009).

38. Cahill, D. P. *et al.* Mutations of mitotic checkpoint genes in human cancers. *Nature* **392**, 300-303, doi:10.1038/32688 (1998).
39. Cahill, D. P. *et al.* Characterization of MAD2B and other mitotic spindle checkpoint genes. *Genomics* **58**, 181-187, doi:10.1006/geno.1999.5831 (1999).
40. Halazonetis, T. D., Gorgoulis, V. G. & Bartek, J. An oncogene-induced DNA damage model for cancer development. *Science (New York, N.Y.)* **319**, 1352-1355, doi:10.1126/science.1140735 (2008).
41. Thiagalingam, S. *et al.* Mechanisms underlying losses of heterozygosity in human colorectal cancers. *Proceedings of the National Academy of Sciences* **98**, 2698-2702, doi:10.1073/pnas.051625398 (2001).
42. Thiagalingam, S. *et al.* Mechanisms underlying losses of heterozygosity in human colorectal cancers. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 2698-2702, doi:10.1073/pnas.051625398 (2001).
43. Paolella, B. R. *et al.* Copy-number and gene dependency analysis reveals partial copy loss of wild-type SF3B1 as a novel cancer vulnerability. *eLife* **6**, doi:10.7554/eLife.23268 (2017).
44. Network, C. G. A. Comprehensive molecular characterization of human colon and rectal cancer. *Nature* **487**, 330-337, doi:10.1038/nature11252 (2012).
45. Brenner, H., Kloor, M. & Pox, C. P. Colorectal cancer. *Lancet* **383**, 1490-1502, doi:10.1016/s0140-6736(13)61649-9 (2014).
46. PF, D. Cancer Staging using anatomical markers. *Bull Inst Nat Hyg* **5**, 81-86 (1950).
47. Institute, N. C. Cancer Staging. (2015). <https://www.cancer.gov/about-cancer/diagnosis-staging/staging>.
48. Hisamuddin, I. M. & Yang, V. W. Genetics of colorectal cancer. *MedGenMed : Medscape general medicine* **6**, 13 (2004).
49. Theodoratou, E. *et al.* A large-scale meta-analysis to refine colorectal cancer risk estimates associated with MUTYH variants. *British journal of cancer* **103**, 1875-1884, doi:10.1038/sj.bjc.6605966 (2010).
50. Palles, C. *et al.* Germline mutations affecting the proofreading domains of POLE and POLD1 predispose to colorectal adenomas and carcinomas. *Nature genetics* **45**, 136-144, doi:10.1038/ng.2503 (2013).
51. Briggs, S. & Tomlinson, I. Germline and somatic polymerase epsilon and delta mutations define a new class of hypermutated colorectal and endometrial cancers. *The Journal of pathology* **230**, 148-153, doi:10.1002/path.4185 (2013).
52. Half, E., Bercovich, D. & Rozen, P. Familial adenomatous polyposis. *Orphanet journal of rare diseases* **4**, 22, doi:10.1186/1750-1172-4-22 (2009).
53. Fearon, E. R. Molecular genetics of colorectal cancer. *Annual review of pathology* **6**, 479-507, doi:10.1146/annurev-pathol-011110-130235 (2011).
54. Jasperson, K. W., Tuohy, T. M., Neklason, D. W. & Burt, R. W. Hereditary and familial colon cancer. *Gastroenterology* **138**, 2044-2058, doi:10.1053/j.gastro.2010.01.054 (2010).
55. Fearon, E. R. & Vogelstein, B. A genetic model for colorectal tumorigenesis. *Cell* **61**, 759-767 (1990).
56. Davies, R. J., Miller, R. & Coleman, N. Colorectal cancer screening: prospects for molecular stool analysis. *Nat Rev Cancer* **5**, 199-209, doi:10.1038/nrc1545 (2005).
57. Jones, S. *et al.* Comparative lesion sequencing provides insights into tumor evolution. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 4283-4288, doi:10.1073/pnas.0712345105 (2008).

58. De Filippo, C. *et al.* Mutations of the APC gene in human sporadic colorectal cancers. *Scandinavian journal of gastroenterology* **37**, 1048-1053 (2002).
59. Samuels, Y. *et al.* High frequency of mutations of the PIK3CA gene in human cancers. *Science (New York, N.Y.)* **304**, 554, doi:10.1126/science.1096502 (2004).
60. Leary, R. J. *et al.* Integrated analysis of homozygous deletions, focal amplifications, and sequence alterations in breast and colorectal cancers. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 16224-16229, doi:10.1073/pnas.0808041105 (2008).
61. Akhoondi, S. *et al.* FBXW7/hCDC4 is a general tumor suppressor in human cancer. *Cancer research* **67**, 9006-9012, doi:10.1158/0008-5472.Can-07-1320 (2007).
62. Bass, A. J. *et al.* Genomic sequencing of colorectal adenocarcinomas identifies a recurrent VTI1A-TCF7L2 fusion. *Nature genetics* **43**, 964-968, doi:10.1038/ng.936 (2011).
63. Duval, A. & Hamelin, R. Mutations at coding repeat sequences in mismatch repair-deficient human cancers: toward a new concept of target genes for instability. *Cancer research* **62**, 2447-2454 (2002).
64. Davies, H. *et al.* Mutations of the BRAF gene in human cancer. *Nature* **417**, 949-954, doi:10.1038/nature00766 (2002).
65. Tie, J. *et al.* Optimizing targeted therapeutic development: analysis of a colorectal cancer patient population with the BRAF(V600E) mutation. *International journal of cancer* **128**, 2075-2084, doi:10.1002/ijc.25555 (2011).
66. Bogaert, J. & Prenen, H. Molecular genetics of colorectal cancer. *Annals of gastroenterology* **27**, 9-14 (2014).
67. Salovaara, R. *et al.* Frequent loss of SMAD4/DPC4 protein in colorectal cancers. *Gut* **51**, 56-59 (2002).
68. Weisenberger, D. J. *et al.* CpG island methylator phenotype underlies sporadic microsatellite instability and is tightly associated with BRAF mutation in colorectal cancer. *Nature genetics* **38**, 787-793, doi:10.1038/ng1834 (2006).
69. Bettington, M. *et al.* The serrated pathway to colorectal carcinoma: current concepts and challenges. *Histopathology* **62**, 367-386, doi:10.1111/his.12055 (2013).
70. Muller, M. F., Ibrahim, A. E. & Arends, M. J. Molecular pathological classification of colorectal cancer. *Virchows Archiv : an international journal of pathology* **469**, 125-134, doi:10.1007/s00428-016-1956-3 (2016).
71. Guinney, J. *et al.* The consensus molecular subtypes of colorectal cancer. *Nature medicine* **21**, 1350-1356, doi:10.1038/nm.3967 (2015).
72. Basso, M. *et al.* Conversion Chemotherapy for Technically Unresectable Colorectal Liver Metastases: A Retrospective, STROBE-Compliant, Single-Center Study Comparing Chemotherapy Alone and Combination Chemotherapy With Cetuximab or Bevacizumab. *Medicine* **95**, e3722, doi:10.1097/md.0000000000003722 (2016).
73. Hurwitz, H. *et al.* Bevacizumab plus irinotecan, fluorouracil, and leucovorin for metastatic colorectal cancer. *The New England journal of medicine* **350**, 2335-2342, doi:10.1056/NEJMoa032691 (2004).
74. Cunningham, D. *et al.* Cetuximab monotherapy and cetuximab plus irinotecan in irinotecan-refractory metastatic colorectal cancer. *The New England journal of medicine* **351**, 337-345, doi:10.1056/NEJMoa033025 (2004).
75. Van Cutsem, E. *et al.* Open-label phase III trial of panitumumab plus best supportive care compared with best supportive care alone in patients with chemotherapy-refractory metastatic colorectal cancer. *Journal of clinical*

oncology : official journal of the American Society of Clinical Oncology **25**, 1658-1664, doi:10.1200/jco.2006.08.1620 (2007).
76. Tabernero, J. *et al.* Ramucirumab versus placebo in combination with second-line FOLFIRI in patients with metastatic colorectal carcinoma that progressed during or after first-line therapy with bevacizumab, oxaliplatin, and a fluoropyrimidine (RAISE): a randomised, double-blind, multicentre, phase 3 study. *The Lancet. Oncology* **16**, 499-508, doi:10.1016/s1470-2045(15)70127-0 (2015).
77. Van Cutsem, E. *et al.* Aflibercept Plus FOLFIRI vs. Placebo Plus FOLFIRI in Second-Line Metastatic Colorectal Cancer: a Post Hoc Analysis of Survival from the Phase III VELOUR Study Subsequent to Exclusion of Patients who had Recurrence During or Within 6 Months of Completing Adjuvant Oxaliplatin-Based Therapy. *Targeted oncology* **11**, 383-400, doi:10.1007/s11523-015-0402-9 (2016).
78. Grothey, A. *et al.* Regorafenib monotherapy for previously treated metastatic colorectal cancer (CORRECT): an international, multicentre, randomised, placebo-controlled, phase 3 trial. *Lancet* **381**, 303-312, doi:10.1016/s0140-6736(12)61900-x (2013).
79. Overman, M. J. *et al.* Durable Clinical Benefit With Nivolumab Plus Ipilimumab in DNA Mismatch Repair-Deficient/Microsatellite Instability-High Metastatic Colorectal Cancer. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*, Jco2017769901, doi:10.1200/jco.2017.76.9901 (2018).
80. Casanovas, O., Hicklin, D. J., Bergers, G. & Hanahan, D. Drug resistance by evasion of antiangiogenic targeting of VEGF signaling in late-stage pancreatic islet tumors. *Cancer cell* **8**, 299-309, doi:10.1016/j.ccr.2005.09.005 (2005).
81. Karapetis, C. S. *et al.* K-ras mutations and benefit from cetuximab in advanced colorectal cancer. *The New England journal of medicine* **359**, 1757-1765, doi:10.1056/NEJMoa0804385 (2008).
82. Amado, R. G. *et al.* Wild-type KRAS is required for panitumumab efficacy in patients with metastatic colorectal cancer. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* **26**, 1626-1634, doi:10.1200/jco.2007.14.7116 (2008).
83. De Roock, W. *et al.* Effects of KRAS, BRAF, NRAS, and PIK3CA mutations on the efficacy of cetuximab plus chemotherapy in chemotherapy-refractory metastatic colorectal cancer: a retrospective consortium analysis. *The Lancet. Oncology* **11**, 753-762, doi:10.1016/s1470-2045(10)70130-3 (2010).
84. Hsu, H. C. *et al.* Mutations of KRAS/NRAS/BRAF predict cetuximab resistance in metastatic colorectal cancer patients. *Oncotarget* **7**, 22257-22270, doi:10.18632/oncotarget.8076 (2016).
85. Diaz, L. A., Jr. *et al.* The molecular evolution of acquired resistance to targeted EGFR blockade in colorectal cancers. *Nature* **486**, 537-540, doi:10.1038/nature11219 (2012).
86. Konieczkowski, D. J. & Garraway, L. A. Resistance to EGFR blockade in colorectal cancer: liquid biopsies and latent subclones. *Cell research* **23**, 13-14, doi:10.1038/cr.2012.115 (2013).
87. Gerber, D. E. Targeted therapies: a new generation of cancer treatments. *American family physician* **77**, 311-319 (2008).
88. Sawyers, C. Targeted cancer therapy. *Nature* **432**, 294-297, doi:10.1038/nature03095 (2004).
89. Clayman, G. L. *et al.* Adenovirus-mediated p53 gene transfer in patients with advanced recurrent head and neck squamous cell carcinoma. *Journal of clinical*

oncology : official journal of the American Society of Clinical Oncology **16**, 2221-2232, doi:10.1200/jco.1998.16.6.2221 (1998).
90. Nemunaitis, J. *et al.* Adenovirus-mediated p53 gene transfer in sequence with cisplatin to tumors of patients with non-small-cell lung cancer. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* **18**, 609-622, doi:10.1200/jco.2000.18.3.609 (2000).
91. Schuler, M. *et al.* Adenovirus-mediated wild-type p53 gene transfer in patients receiving chemotherapy for advanced non-small-cell lung cancer: results of a multicenter phase II study. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* **19**, 1750-1758, doi:10.1200/jco.2001.19.6.1750 (2001).
92. Kaiser, A. D. *et al.* Towards a commercial process for the manufacture of genetically modified T cells for therapy. *Cancer gene therapy* **22**, 72-78, doi:10.1038/cgt.2014.78 (2015).
93. Garber, K. The cancer drug that almost wasn't. *Science (New York, N.Y.)* **345**, 865-867, doi:10.1126/science.345.6199.865 (2014).
94. Khoo, K. H., Verma, C. S. & Lane, D. P. Drugging the p53 pathway: understanding the route to clinical efficacy. *Nature reviews. Drug discovery* **13**, 217-236, doi:10.1038/nrd4236 (2014).
95. Bykov, V. J. *et al.* Targeting of Mutant p53 and the Cellular Redox Balance by APR-246 as a Strategy for Efficient Cancer Therapy. *Frontiers in oncology* **6**, 21, doi:10.3389/fonc.2016.00021 (2016).
96. Fong, P. C. *et al.* Inhibition of poly(ADP-ribose) polymerase in tumors from BRCA mutation carriers. *The New England journal of medicine* **361**, 123-134, doi:10.1056/NEJMoa0900212 (2009).
97. Krastev, D. B. *et al.* A systematic RNAi synthetic interaction screen reveals a link between p53 and snoRNP assembly. *Nature cell biology* **13**, 809-818, doi:10.1038/ncb2264 (2011).
98. Gordon, G. M. & Du, W. Targeting Rb inactivation in cancers by synthetic lethality. *American journal of cancer research* **1**, 773-786 (2011).
99. Muller, F. L. *et al.* Passenger deletions generate therapeutic vulnerabilities in cancer. *Nature* **488**, 337-342, doi:10.1038/nature11331 (2012).
100. Muller, F. L., Aquilanti, E. A. & DePinho, R. A. Collateral Lethality: A new therapeutic strategy in oncology. *Trends in cancer* **1**, 161-173, doi:10.1016/j.trecan.2015.10.002 (2015).
101. Dey, P. *et al.* Genomic deletion of malic enzyme 2 confers collateral lethality in pancreatic cancer. *Nature* **542**, 119-123, doi:10.1038/nature21052 (2017).
102. Mavrakis, K. J. *et al.* Disordered methionine metabolism in MTAP/CDKN2A-deleted cancers leads to dependence on PRMT5. *Science (New York, N.Y.)* **351**, 1208-1213, doi:10.1126/science.aad5944 (2016).
103. Kryukov, G. V. *et al.* MTAP deletion confers enhanced dependency on the PRMT5 arginine methyltransferase in cancer cells. *Science (New York, N.Y.)* **351**, 1214-1218, doi:10.1126/science.aad5214 (2016).
104. Liu, Y. *et al.* TP53 loss creates therapeutic vulnerability in colorectal cancer. *Nature* **520**, 697-701, doi:10.1038/nature14418 (2015).
105. Nijhawan, D. *et al.* Cancer vulnerabilities unveiled by genomic loss. *Cell* **150**, 842-854, doi:10.1016/j.cell.2012.07.023 (2012).
106. Seeb, J. E. *et al.* Single-nucleotide polymorphism (SNP) discovery and applications of SNP genotyping in nonmodel organisms. *Molecular ecology resources* **11 Suppl 1**, 1-8, doi:10.1111/j.1755-0998.2010.02979.x (2011).
107. Fluiter, K., Housman, D., Ten Asbroek, A. L. & Baas, F. Killing cancer by targeting genes that cancer cells have lost: allele-specific inhibition, a novel

approach to the treatment of genetic disorders. *Cellular and molecular life sciences : CMLS* **60**, 834-843, doi:10.1007/s00018-003-3043-6 (2003).
108. Basilion, J. P. *et al.* Selective killing of cancer cells based on loss of heterozygosity and normal variation in the human genome: a new paradigm for anticancer drug therapy. *Molecular pharmacology* **56**, 359-369 (1999).
109. ten Asbroek, A. L., Fluiter, K., van Groenigen, M., Nooij, M. & Baas, F. Polymorphisms in the large subunit of human RNA polymerase II as target for allele-specific inhibition. *Nucleic acids research* **28**, 1133-1138 (2000).
110. Mook, O. R., Baas, F., de Wissel, M. B. & Fluiter, K. Allele-specific cancer cell killing in vitro and in vivo targeting a single-nucleotide polymorphism in POLR2A. *Cancer gene therapy* **16**, 532-538, doi:10.1038/cgt.2008.104 (2009).
111. Hein, D. W. Molecular genetics and function of NAT1 and NAT2: role in aromatic amine metabolism and carcinogenesis. *Mutation research* **506-507**, 65-77 (2002).
112. Mertens, F., Johansson, B., Hoglund, M. & Mitelman, F. Chromosomal imbalance maps of malignant solid tumors: a cytogenetic survey of 3185 neoplasms. *Cancer research* **57**, 2765-2780 (1997).
113. MacArthur, D. G. *et al.* A systematic survey of loss-of-function variants in human protein-coding genes. *Science (New York, N.Y.)* **335**, 823-828, doi:10.1126/science.1215040 (2012).
114. Zhou, S. F., Liu, J. P. & Lai, X. S. Substrate specificity, inhibitors and regulation of human cytochrome P450 2D6 and implications in drug development. *Current medicinal chemistry* **16**, 2661-2805 (2009).
115. Alexandrov, L. B. & Stratton, M. R. Mutational signatures: the patterns of somatic mutations hidden in cancer genomes. *Current opinion in genetics & development* **24**, 52-60, doi:10.1016/j.gde.2013.11.014 (2014).
116. Pal, S. K. & Mittal, B. Fight against cancer in countries with limited resources: the post-genomic era scenario. *Asian Pacific journal of cancer prevention : APJCP* **5**, 328-333 (2004).
117. Cohen, J. D. *et al.* Detection and localization of surgically resectable cancers with a multi-analyte blood test. *Science (New York, N.Y.)* **359**, 926-930, doi:10.1126/science.aar3247 (2018).
118. Jones, S. *et al.* Core signaling pathways in human pancreatic cancers revealed by global genomic analyses. *Science (New York, N.Y.)* **321**, 1801-1806, doi:10.1126/science.1164368 (2008).
119. Ding, L. *et al.* Somatic mutations affect key pathways in lung adenocarcinoma. *Nature* **455**, 1069-1075, doi:10.1038/nature07423 (2008).
120. Baines, A. T., Xu, D. & Der, C. J. Inhibition of Ras for cancer treatment: the search continues. *Future medicinal chemistry* **3**, 1787-1808, doi:10.4155/fmc.11.121 (2011).
121. Vilar, E. & Gruber, S. B. Microsatellite instability in colorectal cancer-the stable evidence. *Nature reviews. Clinical oncology* **7**, 153-162, doi:10.1038/nrclinonc.2009.237 (2010).

Acta Universitatis Upsaliensis

Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine 1443

Editor: The Dean of the Faculty of Medicine

A doctoral dissertation from the Faculty of Medicine, Uppsala University, is usually a summary of a number of papers. A few copies of the complete dissertation are kept at major Swedish research libraries, while the summary alone is distributed internationally through the series Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine. (Prior to January, 2005, the series was published under the title "Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine".)

Distribution: publications.uu.se
urn:nbn:se:uu:diva-345970

ACTA
UNIVERSITATIS
UPSALIENSIS
UPPSALA
2018