Molecular genetic changes in colorectal and small bowel cancer

Ulrika Hänninen

Faculty of Medicine
Department of Medical and Clinical Genetics, Medicum
Applied Tumor Genomics Research Program
Doctoral Programme in Biomedicine (DPBM)
University of Helsinki
Finland

DOCTORAL DISSERTATION

To be presented for public discussion with the permission of the Faculty of Medicine of the University of Helsinki, in Metsätalo, Lecture Hall 6, Unioninkatu 40, Helsinki, on the 16th of October, 2020 at 13 o'clock.

Helsinki 2020

Supervised by Academy Professor Lauri A. Aaltonen, M.D., Ph.D.
Department of Medical and Clinical Genetics, Medicum
Applied Tumor Genomics Research Program
Faculty of Medicine
University of Helsinki, Helsinki, Finland

Netta Mäkinen, Ph.D.
Department of Medical and Clinical Genetics, Medicum
Applied Tumor Genomics Research Program
Faculty of Medicine
University of Helsinki, Helsinki, Finland

Sari Tuupanen, Ph.D.
Department of Medical and Clinical Genetics, Medicum
Applied Tumor Genomics Research Program
Faculty of Medicine
University of Helsinki, Helsinki, Finland

Reviewed by Professor Arto Mannermaa, Ph.D.
Pathology and Forensic Medicine, Institute of Clinical Medicine
Translational Cancer Research Area
University of Eastern Finland, Kuopio, Finland
Biobank of Eastern Finland

Docent Tero Hiekkalinna, Ph.D.
Genomics and Biobank Unit
Finnish Institute for Health and Welfare (THL)
Biomedicum Helsinki, Helsinki, Finland

Official Opponent Professor Mef Nilbert, M.D., Ph.D.
The Danish Cancer Society Research Center, Copenhagen, Denmark
Clinical Research Center
Hvidovre Hospital
University of Copenhagen, Copenhagen, Denmark

Doctoral Programme in Biomedicine
ISBN 978-951-51-6487-2 (paperback)
ISBN 978-951-51-6488-9 (PDF)
Unigrafia Oy
Helsinki 2020
The Faculty of Medicine uses the Urkund system (plagiarism recognition) to examine all doctoral dissertations.

To my family

ABSTRACT

Colorectal cancer (CRC) is the third most common cancer and the second main cause of cancer mortality worldwide. The colorectum and the small bowel are part of a continuous passageway called the gastrointestinal (GI) tract, however, the cancer incidence varies greatly between these two organs. Although the small bowel constitutes three quarters of the length of the GI tract, only 3% of GI cancers are located in the small bowel. These cancers develop due to both somatic and inherited germline mutations. Thus, characterizing the genetic events that drive tumorigenesis is crucial to provide ways to improve prevention and clinical management of the disease. The general aim of this thesis was to gain new insights into the molecular genetic backgrounds of CRC and small bowel adenocarcinoma (SBA).

The first aim of the thesis was to characterize the somatic mutation patterns of the AT-Rich Interaction Domain (ARID) family genes in CRCs with microsatellite instability (MSI). Approximately 15% of CRCs display MSI which arises due to a defective DNA mismatch repair system. These tumors accumulate a high number of mutations, especially small insertions and deletions in repetitive genomic areas called microsatellites. The ARID gene family comprises 15 members, including a known tumor suppressor gene *ARID1A*. We utilized exome sequencing data of 25 MSI CRCs and their corresponding normal tissues and identified 12 of the ARID genes to display mutations with a frequency of 4-52%. Four genes were selected for further analysis in 21 additional MSI CRCs. We found that, in addition to *ARID1A*, also *ARID1B*, *ARID2*, and *ARID4A* were frequently mutated and might play a role in MSI CRC. However, additional studies are warranted to further scrutinize the function of these mutations in MSI CRC genesis.

The second aim of the thesis was to identify novel oncogenes in MSI CRC. These tumors represent a sensitive system for studying the generation and selection of oncogenic mutations. In contrast to many reported MSI target genes, few oncogenes are known in MSI CRC and they often display specific mutation hotspots. Thus, we used the exome sequencing data of 25 MSI CRCs and their corresponding normal tissues to search for genes with recurrent somatic missense mutations. We identified 33 novel candidate oncogenes of which the following fourteen genes displayed hotspot mutations also in the validation set of 254 MSI CRCs: *ANTXR1*, *CEP135*, *CRYBB1*, *MORC2*, *SLC36A1*, *GALNT9*, *PI15*, *KRT82*, *CNTF*, *GLDC*, *MBTPS1*, *OR9Q2*, *R3HDM1*, and *TTPAL*. This work revealed a variety of novel recurrent candidate oncogene mutations that might potentially be used to develop personalized therapies. Further research is still needed to confirm their pathogenic role and detailed function in tumorigenesis.

The third aim of the thesis was to study the genetic overlap within synchronous CRCs (SCRCs). Approximately 4% of CRC patients display multiple simultaneous primary cancers in the colorectum. Understanding whether SCRCs within a patient are genetically similar or distinct is essential when designing personalized treatments. Exome sequencing data of 23 SCRC pairs and their corresponding healthy tissues revealed that the paired tumors shared a maximum of only a few somatic mutations. This indicated that the tumors have independent origins. Furthermore, paired tumors favored different somatic mutations in known CRC genes and signaling pathways. Variation was observed among clinically relevant genes, such as the discordant *KRAS* mutation status in a quarter of patients. Tumors within pairs also displayed variation in their mutational signature content suggesting that, regardless

of the shared environment, some pairs might have undergone different mutational processes. Finally, by analyzing immune cell counts, we observed that the intratumor immune response varied within most tumor pairs. This was not explained by mutation burden or clinicopathological variables. Overall, this work revealed major diversity within SCRCs and highlights the need to evaluate all synchronous lesions within an individual for optimized therapeutic approach. Additional studies are still required to further elucidate the reasons underlying tumor multiplicity.

The fourth aim of the thesis was to characterize the somatic mutation content in SBA. Due to its rarity, knowledge on the genetic background of SBA has remained somewhat elusive. We conducted the first large exome sequencing effort of 106 population-based SBAs representing all three small bowel segments. This work revealed significantly mutated genes previously associated with SBA (*TP53*, *KRAS*, *APC*, *SMAD4*, and *BRAF*), as well as novel candidate drivers, such as *ACVR2A*, *ACVR1B*, *BRCA2*, and *SMARCA4*. We identified clear mutation hotspot patterns in *ERBB2* and *BRAF*. Interestingly, we observed no V600E mutations, the most common *BRAF* mutation hotspot in CRC. Other clinically relevant aspects included mutations in ERBB family genes in over a quarter of SBAs as well as mutations in multiple genes that could predict anti-EGFR treatment resistance. We performed the first comprehensive mutation signature analysis on SBA that highlighted four signatures: 1A, 6, 17, and U2. Comparison of the three small bowel segments unveiled some variation in tumor characteristics. Further studies are needed to robustly clarify these differences and their clinical relevance. This comprehensive characterization provided further evidence that SBA is a distinct tumor type and singled out many potential therapeutic targets that could be utilized in SBA treatment development.

TABLE OF CONTENTS

LIST OF ORIGINAL PUBLICATIONS

This thesis is based on the following publications:

I. Cajuso T*, **Hänninen UA***, Kondelin J, Gylfe AE, Tanskanen T, Katainen R, Pitkänen E, Ristolainen H, Kaasinen E, Taipale M, Taipale J, Böhm J, Renkonen-Sinisalo L, Mecklin JP, Järvinen H, Tuupanen S, Kilpivaara O, Vahteristo P. Exome sequencing reveals frequent inactivating mutations in *ARID1A*, *ARID1B*, *ARID2* and *ARID4A* in microsatellite unstable colorectal cancer. *Int J Cancer*. 2014 135:611-23.

II. Tuupanen S, **Hänninen UA**, Kondelin J, von Nandelstadh P, Cajuso T, Gylfe AE, Katainen R, Tanskanen T, Ristolainen H, Böhm J, Mecklin JP, Järvinen H, Renkonen-Sinisalo L, Andersen CL, Taipale M, Taipale J, Vahteristo P, Lehti K, Pitkänen E, Aaltonen LA. Identification of 33 candidate oncogenes by screening for base-specific mutations. *Br J Cancer*. 2014 111:1657-62.

III. **Hänninen UA**, Wirta E-V, Katainen R, Tanskanen T, Hamberg J, Taipale M, Böhm J, Renkonen-Sinisalo L, Lepistö A, Forsström LM, Pitkänen E, Palin K, Seppälä TT, Mäkinen N, Mecklin J-P, Aaltonen LA. Exome and immune cell score analyses reveal great variation within synchronous primary colorectal cancers. *Br J Cancer*. 2019 120:922-930.

IV. **Hänninen UA**, Katainen R, Tanskanen T, Plaketti RM, Laine R, Hamberg J, Ristimäki A, Pukkala E, Taipale M, Mecklin JP, Forsström LM, Pitkänen E, Palin K, Välimäki N, Mäkinen N, Aaltonen LA. Exome-wide somatic mutation characterization of small bowel adenocarcinoma. *PLoS Genet*. 2018 14:e1007200.

* Equal contribution

The publications referred to in the text by their Roman numerals. The publications are reproduced with the permission of the copyright holders.

ABBREVIATIONS

ACVR2A	Activin A Receptor Type 2A
AI	allelic imbalance
ANTXR1	ANTXR Cell Adhesion Molecule 1
APC	APC Regulator Of WNT Signaling Pathway
ARID	AT-Rich Interaction Domain
ATM	ATM Serine/Threonine Kinase
BRAF	B-Raf Proto-Oncogene
BRCA2	BRCA2 DNA Repair Associated
CIN	chromosomal instability
CI	confidence interval
CRC	colorectal cancer
CTNNB1	Catenin Beta 1
DNA	deoxyribonucleic acid
EGFR	Epidermal Growth Factor Receptor
ERBB	Erb-B2 Receptor Tyrosine Kinase
ERK	extracellular signal-regulated kinase
FAP	familial adenomatous polyposis
FFPE	formalin-fixed paraffin-embedded
GATK	Genome Analysis Toolkit
gnomAD	Genome Aggregation Database
GI	gastrointestinal
HER2	human epidermal growth receptor 2
IBD	inflammatory bowel disease
IHC	immunohistochemistry
IS	immune cell score
JPS	juveline polyposis syndrome
KRAS	KRAS Proto-Oncogene
LS	Lynch syndrome
MAF	minor allele frequency
MAP	MUTYH-associated polyposis
MAPK	mitogen-activated protein kinase
MLH1	MutL Homolog 1
MSH2	MutS Homolog 2
MSH6	MutS Homolog 6
MMR	mismatch repair
MORC2	MORC Family CW-Type Zinc Finger 2
MSI	microsatellite instability
MSS	microsatellite stable
MUTYH	MutY DNA Glycosylase
mut/Mb	mutations per megabase
MYC	MYC Proto-Oncogene
NGS	next generation sequencing
OR	odds ratio

PCR	polymerase chain reaction
PIK3CA	Phosphatidylinositol-4,5-Bisphosphate 3-Kinase Catalytic Subunit Alpha
PI3K	phosphoinositide-3-kinase
PJS	Peutz-Jeghers syndrome
PMS2	PMS1 Homolog 2
POLD1	DNA Polymerase Delta 1
POLE	DNA Polymerase Epsilon
PTEN	Phosphatase And Tensin Homolog
SBA	small bowel adenocarcinoma
SCRC	synchronous colorectal cancer
SMAD4	SMAD Family Member 4
SMARCA4	SWI/SNF Related, Matrix Associated, Actin Dependent Regulator Of Chromatin, Subfamily A, Member 4
SNP	single nucleotide polymorphism
SNV	single nucleotide variant
SWI/SNF	Switch/Sucrose Non-Fermentable
TGFβ	transforming growth factor beta
TP53	Tumor Protein P53
WGS	whole genome sequencing

REVIEW OF THE LITERATURE

1. Etiology of cancer

Cancer refers to a large group of diseases that share the ability for abnormal growth and proliferation of cells, regardless of normal cell control mechanisms. It can originate from most cell types and organs. Adenocarcinoma is the most common cancer type and it arises from columnar epithelial cells found e.g. in the lining of the gastrointestinal (GI) tract. Cancers are characterized by their ability to invade locally, disseminate from the primary site, and metastasize to distant organs. This separates them from benign tumors that, by definition, are localized and non-invasive. Benign tumors can, however, also become clinically relevant if they grow enough to create pressure to the surrounding tissues or produce abnormal quantities of hormones.

Cancer is one of the leading causes for morbidity and mortality worldwide. In Finland, over 34,000 cancer cases were diagnosed in 2017 and slightly under 12,800 patients died of it (Finnish Cancer Registry, https://syoparekisteri.fi/tilastot/syopa-2017-raportti/). Most cancer related deaths are due to metastatic disease[1]. Today, nearly two thirds of the cancer patients are cured (www.cancer.fi). In Finland, the most common primary sites for cancer are prostate, colorectum, and lung in males and breast, colorectum, and lung in females (www.cancerregistry.fi/statistics/cancer-statistics/).

Cancer is thought to originate from a single cell. Accumulation of mutations in deoxyribonucleic acid (DNA) together with epigenetic changes provide a cell growth advantage over the adjacent cells and the ability to undergo clonal expansion[2,3]. Thus, cancer is considered as a genetic disease and tumorigenesis an evolutionary process comparable to Darwinian natural selection. During the expansion of neoplastic cell population, individual cells obtain different mutations during each cell division leading to branched evolution where some subclones proliferate faster than other lesser fit clones. This produces genetic heterogeneity within the tumor as well as metastasis, between different metastatic lesions within a patient, and between tumors of each individual[4]. Tumor growth and its spread to distant sites is further advanced by complex interplay between neoplastic cells and their surrounding microenvironment. Cancer can be considered as a complex dynamic environment that contains heterogeneous populations of cancer cells in contact with the surrounding stroma and its many types of cells, such as immune cells, fibroblasts, endothelial cells, pericytes, and smooth muscle cells[5].

Features that a cell needs to obtain to evade the mechanisms that control cell growth and to reach a malignant state are described as the hallmarks of cancer[5,6]. These characteristics include sustaining proliferative signaling, evasion of growth suppressors, resisting cell death, enabling replicative immortality, inducing angiogenesis, and activating invasion and metastasis. There are two additional hallmarks that enable the above-mentioned hallmarks, namely genome instability and tumor-promoting inflammation. Finally, two emerging hallmarks have been added to the list: reprogramming energy metabolism and evasion of immune destruction.

The development of a solid tumor, from a benign local lesion into a malignant cancer, has been estimated to take up to several decades[7]. Due to accumulation of mutations over the course of time, age is considered as a key risk factor for cancer. Cancer risk is also influenced by lifestyle and

environmental factors which vary between different cancers. The most common preventable lifestyle risk factors include smoking, obesity, diet, and alcohol consumption. Environmental risk factors entail ultraviolet and ionizing radiation and certain infectious agents, such as human papillomaviruses that predispose to many cancers, most commonly cervical cancer, and Helicobacter pylori that predisposes to gastric cancer[8,9]. Hereditary factors, that is genomic changes in the germline of an individual, are also recognized as important contributors to many common cancers.

2. Genetic and epigenetic features in cancer

Mutations in the DNA are either inherited or acquired. Inherited mutations exist in the fertilized egg and therefore found in all cells of an individual. Some of these germline mutations increase the risk of developing cancer. This contribution in risk varies greatly between mutations. Hereditary cancer refers to cases that carry a germline mutation that confers increased susceptibility to cancer, whereas cancers that arise without any known germline contribution are called sporadic. Contribution of predisposing mutations can vary between different cancers. As an example, familial adenomatous polyposis (FAP) patients with a germline mutation in the gene *APC Regulator Of WNT Signaling Pathway* (*APC*) present more often cancers in the colorectum than in the small bowel[10].

Acquired, also called somatic, mutations are alterations that gather during one's lifetime and can occur in all dividing cells, both normal and malignant[11]. The median somatic mutation frequency in a normal cell has been estimated to be 2.8×10^{-7} per base pair[12]. These mutations arise either from exposure to endogenous and exogenous mutagens or in DNA replication during mitotic cell division[11]. The changes include base substitutions, deletions or insertions of bases, copy number alterations, and rearrangements[3]. Base substitutions can be further divided into nonsynonymous and synonymous, depending on whether they result in an amino acid change or not. In a normal cell, most mutations are corrected by DNA repair mechanisms. Mutations that are not repaired are passed onto the following daughter cells (Figure 1). Due to the accumulation of mutations in the course of time, the risk of developing cancer increases with age. It has been estimated that approximately 29% of mutations in cancers are due to environmental factors, 66% are attributable to DNA replication errors, and 5% are due to hereditary causes[13].

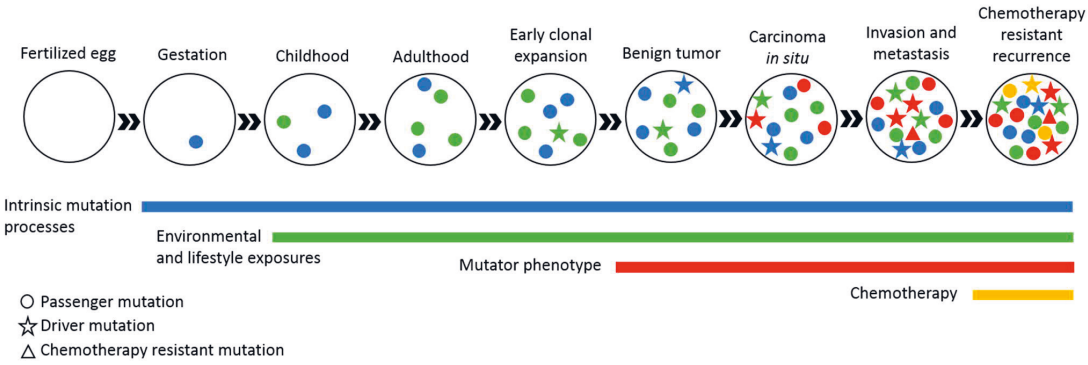

Figure 1. The development of a single cell lineage to a malignant tumor. A cancer cell develops via a lineage of mitotic cell divisions from the fertilized egg. Somatic mutations accumulate over a lifetime due to intrinsic and environmental factors. Most of these mutations are passenger events, whereas a small subset are driver mutations that confer selective growth advantage. Chemotherapy resistant mutations often predate the cancer treatment and can cause relapse afterwards. Adapted by permission from Macmillan Publishers Ltd: Nature (Stratton *et al*. 2009), copyright (2009).

Mutations that contribute to obtaining cancer hallmarks, bring selective advantage to the cell, and thus drive tumorigenesis are called driver mutations[4]. It has been estimated that only two to eight such driver events are required for cancer development[4,14]. Mutations accumulate through cell division and once several driver genes are defective, cancer will develop. Of note, some mutations are only needed at later stages of tumor evolution to achieve resistance to chemotherapy. Whereas, mutations that do not confer selective growth advantage and have not been subject to selection are called passenger mutations[11]. Most of the mutations found in cancer are passenger events that have occurred by chance in a progenitor cell that underwent clonal expansion.

Cancer driver genes are classically divided into oncogenes and tumor suppressors. Oncogenes can promote tumor growth through activating mutations. Tumor suppressor genes, on the other hand, achieve their tumorigenic effect by mutations that inactivate the gene function. Some cancer driver genes, such as *Tumor Protein P53* (*TP53*) and *KRAS Proto-Oncogene* (*KRAS*), are frequently mutated in a variety of cancers, whereas other genes are specific to certain tumor types[15]. It has been suggested that tumorigenesis is driven by 12 main signaling pathways that regulate cell survival, cell fate, and genome integrity[4]. Known cancer genes can be associated with one or more of these pathways. Classification of genes into pathways provides better understanding why certain driver mutations are essential to cancer, how mutations in different genes can have similar effect on tumorigenesis, and helps direct cancer therapy development.

2.1. Oncogenes

Proto-oncogene is the normal equivalent of an oncogene. When proto-oncogenes are altered, they are transformed into oncogenes that can drive tumorigenesis. Usually these changes act in a dominant manner, that is, mutation in one allele is enough to provide selective growth advantage to the cell. Oncogenes encode proteins that regulate cell proliferation and survival. These include growth factors and their receptors, transcription factors, chromatin remodelers, signaling molecules, and apoptosis

regulators[16]. Gain-of-function defects in oncogenes are typically point mutations that reoccur at one or few amino acid positions, called mutation hotspots. In addition, activation can happen through gene amplification and chromosomal rearrangements. Causative germline mutations are only rarely found in oncogenes due to their embryonic lethality.

First somatic point mutation identified in cancer was *HRAS* harboring recurrent missense mutations in codon 12[17]. *HRAS*, together with *KRAS* and *NRAS*, are members of the RAS family that encode for small GTPases which are involved in transducing signals within the cell through mitogen-activated protein kinase (MAPK) signaling. When mutated at codons 12, 13, or 61, resulting proteins remain in an active state inducing continuous cell growth[16]. *KRAS* is considered as the most commonly mutated oncogene[18]. Another well-known oncogene displaying activating hotspot mutations is *B-Raf Proto-Oncogene* (*BRAF*), usually harboring a mutation at codon V600. BRAF acts downstream of RAS in the MAPK/extracellular signal-regulated kinase (ERK) pathway and when aberrantly activated, phosphorylates downstream targets constitutively leading to aberrant cell growth[19]. *BRAF* mutations are found in a variety of cancers, such as melanoma, non-small cell lung cancer, thyroid, hepatocellular, and colorectal cancer (CRC).

Chromosomal rearrangements activating oncogenes are usually translocations that combine proto-oncogenes with an enhancer or form a chimeric protein product. The first evidence that cancer is caused by somatic genetic changes appeared from Burkitt lymphoma studies where *MYC Proto-Oncogene* (*MYC*) was observed to translocate to three alternative loci that harbored an immunoglobulin gene with an enhancer element, thus activating the juxtaposed *MYC*[20–22]. Whereas, the fusion of ALL1 with a variety of proteins is a well-known example of a chimeric protein formation underlying acute myelogenous leukemia and acute lymphoblastic leukemia[16]. Translocations are also encountered in solid tumors, such as the *TMPRSS2-ERG* gene fusion commonly found in prostate cancer[23]. Amplification that results in an increased expression of an oncogene is often seen in MYC and RAS family genes in various cancers. Other targets for amplification are Erb-B2 Receptor Tyrosine Kinase (ERBB) family genes, such as *Epidermal Growth Factor Receptor* (*EGFR*, also known as *ERBB1*) and *ERBB2* (also called *Human Epidermal Growth Receptor 2* (*HER2*)). *EGFR* amplifications are encountered e.g. in glioblastoma and head and neck cancers and *ERBB2* in breast cancer.

Oncogenic proteins can be targeted by small molecules or monoclonal antibodies[16]. They are feasible therapy targets since cancer cells are dependent on the oncogene products, making them more sensitive to therapy compared to normal cells. One example of such a drug is a monoclonal antibody against HER2, trastuzumab, that is effectively used in breast cancer overexpressing *HER2*.

2.2. Tumor suppressor genes

In normal cells, tumor suppressors function to keep the cell differentiation stable by, for example, restraining growth, proliferation, and passage through the cell cycle[24]. Tumor suppressors are traditionally divided into two categories, gatekeeper and caretaker genes[25]. Gatekeepers limit tumor initiation directly by inhibiting growth or promoting death, whereas caretakers maintain genome integrity. Tumor suppressor genes are often inactivated by protein-truncating mutations coupled with allelic loss of the second allele or by promoter hypermethylation[26]. As opposed to oncogenes,

mutations in tumor suppressor genes are typically scattered along the gene. Inactivation of tumor suppressor genes can lead to either reduced protein activity or a complete loss of function.

Tumor suppressor genes are recessive in nature. This means that usually both alleles must be inactivated to result in selective growth advantage for the cell. Patients carrying a germline mutation in a tumor suppressor gene, have an increased risk of cancer since only one additional somatic mutation is needed for biallelic loss of function. This two-hit hypothesis was first proposed by Alfred Knudson in early-onset retinoblastoma, a tumor of the retina[27]. Fifteen years later, *RB1* - the gene responsible for the disease, was discovered[28]. *RB1* is a negative regulator of the cell cycle and represents a classic example of a gatekeeper gene. When both alleles of the gene are targeted, it drives cell progression in a direct manner[25].

Exceptions to the classical two-hit hypothesis also exist. In some cases, the inactivation of one allele of a tumor suppressor is enough to promote tumorigenesis, termed "haploinsufficiency"[29]. Sometimes this single-copy loss may even be preferred in tumor evolution. One example of this is a haploinsufficient loss of *Phosphatase And Tensin Homolog* (*PTEN*) which, while providing growth advantage, avoids triggering p53-dependent senescence that complete loss of *PTEN* would promote[30,31]. Another exception is called a dominant-negative effect, that is, when one mutated gene copy interferes with the function of the normal protein produced by the wild-type allele. For example, in *TP53*, the best-known tumor suppressor, mutation in one allele may contribute to the oligomerization of the mutant and wild-type protein[32].

Caretakers, also called "stability genes", include genes that encode for mitotic checkpoint and DNA repair proteins, such as *BRCA1*, *BRCA2 DNA Repair Associated* (*BRCA2*), *ATM Serine/Threonine Kinase* (*ATM*), and *MutL Homologue 1* (*MLH1*) together with other mismatch repair (MMR) genes[25,33]. Their inactivation promotes tumorigenesis indirectly by leading to increased mutation rate across the genome including genes that provide selective growth advantage. Germline mutations in the DNA repair genes *BRCA1* and *BRCA2* are found in hereditary breast cancer and *ATM* mutations in ataxia telangiectasia. An example of epigenetic inactivation of such genes is biallelic promoter hypermethylation of *MLH1* seen in sporadic CRC[34].

Some tumor suppressors, such as *TP53*, might harbor features of both classes[35]. *TP53* is categorized as a gatekeeper due to its direct involvement in cell cycle regulation and cell proliferation. It also possesses caretaker features as it functions in the DNA damage response. Another example is *AT-Rich Interaction Domain 1A* (*ARID1A*), recognized as one of the most commonly mutated genes in cancer[15]. ARID1A is a chromatin remodeler and involved in cell cycle regulation[36]. It may also present features of a caretaker in the prevention of genomic instability[37]. ARID1A is one of 15 ARID domain-containing proteins that bind to DNA either specifically at AT-rich sites or in a non-specific manner[38]. In general, the ARID proteins partake in many biological processes, such as regulation of development and gene expression.

A third category of tumor suppressors, the landscapers, has also been introduced. This includes genes that do not have a direct impact on cell growth. Instead, they promote neoplastic transformation of cells by creating an abnormal stromal environment[39]. Compared to oncogenes, developing therapies which target tumor suppressors is much more difficult since the replacement of the defective protein

would be required to restore its normal activity. However, another promising approach is to target the signaling pathways controlled by these genes[40].

2.3. Genomic instability

Genomic instability occurs in most cancers and refers to chromosomal instability (CIN) or increased mutation frequency in the DNA[41]. CIN is the most common type of genomic instability in cancer and is characterized by an increased rate of structural and numerical changes in chromosomes. Structural variations include insertions, deletions, duplications, inversions, translocations, amplifications, and insertions of transposable elements. CIN can also cause aneuploidy through loss or gain of whole chromosomes. Such changes can contribute to driving tumorigenesis by activating oncogenes and inactivating tumor suppressors. Though the rate of chromosomal changes in tumors is elevated, the point mutation rate is estimated to be similar to that of normal cells[4]. Processes underlying CIN are not fully understood but it is thought to originate from various mechanisms, such as impaired mitotic checkpoint function or cell-cycle regulation, that lead to missegregation of chromosomes[42]. CIN is thought to occur early and to be a pivotal step in tumorigenesis[43]. Given that CIN influences various steps in the initiation and development of a tumor, and may contribute to the development of metastases, therapeutics targeting CIN could have a major impact in the improvement of clinical outcomes[44].

The increased rate of mutations can be due to a defective DNA MMR system. This is caused by biallelic inactivation of one of the MMR genes (*MLH1*, *MutS Homolog 2* (*MSH2*)*, MutS Homolog 6* (*MSH6*), or *PMS1 Homolog 2* (*PMS2*)). When defective, MMR cannot repair the slippage errors made by the DNA polymerase during DNA replication. This results in the accumulation of mutations, mainly insertions and deletions in short repetitive DNA tracks called microsatellites. This phenomenon is called microsatellite instability (MSI) and is encountered e.g. in CRC, gastric, and endometrial cancer. The point mutation rate is also elevated in these tumors[24]. Another mechanism to drive malignant transformation emerges from defective DNA proofreading enzymes, DNA Polymerase Epsilon (POLE) and DNA Polymerase Delta 1 (POLD1), which leads to increased frequency of point mutations in the DNA[45,46]. This ultra-mutated phenotype has been found in CRC and endometrial cancer.

2.4. Epigenetic changes

In addition to genetic alterations, gene products can be affected through epigenetic modifications[4]. Epigenetics refers to changes in gene expression occurring independent of changes in the underlying DNA sequence. The cancer epigenome is characterized by changes in DNA methylation (such as genome-wide hypomethylation or hypermethylation of gene promoters with CpG-islands), histone modification, nucleosome positioning, and micro-ribonucleic acids (miRNAs)[47]. These changes can result in silencing of tumor suppressors and activation of oncogenes and thus play an important role in cancer initiation and progression. Epigenetic alterations can lead to inactivation of tumor suppressor genes either independently or together with an inactivating genetic change. Therefore, epigenetic changes can serve as the second hit required for cancer initiation according to Knudson's "two-hit" hypothesis. Since epigenetic alterations are potentially reversible, they make appealing candidates for the development of targeted therapies.

Chromatin structure defines how DNA is organized within a cell, influencing which areas are active in a given time[47]. There are many mechanisms that allow compaction and decompaction of DNA. One of such is Switch/Sucrose Non-Fermentable (SWI/SNF) complexes that regulate gene expression by using the energy of ATP to remodel chromatin[48]. Mammalian SWI/SNF complexes contain a core ATPase, either BRG1 or BRM, involved in nucleosome remodeling and other co-factors, such as members of the ARID family, ARID1A, ARID1B, and ARID2. These complexes can be divided into two broad categories based upon the presence of either one of the mutually exclusive ARID1A or ARID1B subunits (BAF complex) or ARID2 and PBRM1 subunits (PBAF complex). The ARID domain allows the proteins to bind to DNA. SWI/SNF complexes have been estimated as the most frequently mutated epigenetic regulators in cancer[49]. These complexes are thought to have a widespread role in tumor suppression. Of the genes encoding for the subunits, *ARID1A* is the most frequently mutated in cancer[48]. Mutations in many of the other genes, such as *SWI/SNF Related, Matrix Associated, Actin Dependent Regulator Of Chromatin, Subfamily A, Member 4* (*SMARCA4*, encodes for BRG1), *PBRM1, ARID1B*, and *ARID2*, have also been found in a variety of cancers[49].

2.5. Inherited predisposition to cancer

Genetic susceptibility for cancer can be divided into three categories based on risk allele frequencies in the population and the relative cancer risk these alleles confer[50]. High-penetrance alleles are typically rare but have a strong effect (minor allele frequency (MAF) ≤0.1%, odds ratio (OR) ≥10), whereas rare, moderate-penetrant alleles are more common with weaker effects (MAF ≤2%, OR ≥2.0), and low-penetrance alleles are even more common and have even weaker effects (MAF >10%, OR <1.5).

Most of the common cancers are sporadic. However, approximately 5-10% of all cancers are thought to arise due to inherited cancer syndromes caused by rare, highly penetrant germline mutations[51]. In addition, approximately 10-15% of cancers are classified as "familial", defined as a cluster of cancers seen within a family more frequently than would be expected. These cases do not typically follow any clear genetic inheritance pattern and their molecular background remains largely unknown. Familial clustering of cancers is most likely due to combined effects of common low-penetrance alleles, gene-environment interactions, or both. The clinical phenotype and cancer risk related with different cancer predisposing alterations may also be influenced by other genetic and non-genetic factors[52].

2.5.1. Inherited cancer syndromes

Inherited cancer syndromes are usually caused by rare germline mutations that are highly penetrant, that is, they confer high relative risks of cancer in carriers. Compared to sporadic cases, inherited cancers are associated with distinct clinical features such as several individuals affected by the same tumor type in the family over multiple generations, early age of onset, and individuals harboring multiple primary cancers[51]. Most syndromes follow Mendelian inheritance, that is they are monogenic and inherited in an autosomal dominant manner.

To date, germline mutations in over a hundred genes have been implicated to confer moderately or highly increased risk of cancer[52]. Most of the cancer predisposition genes act as tumor suppressors

that conform to the two-hit model of cancer susceptibility. Examples of such genes include *RB1* in retinoblastoma and *APC* in FAP[50]. Cancer predisposition genes have a broad range of functions. In general, they are involved in fundamental cellular processes such as cell-cycle regulation and DNA repair. *BRCA2* is an example of the complexity that underlies the cancer risk conferred by cancer predisposition genes[52]. BRCA2 participates in DNA repair by homologous recombination. Germline loss-of-function mutations in *BRCA2* increase the lifetime risk for breast and ovarian cancer substantially, whereas for prostate cancer the increase is only small. The effect of *BRCA2* mutations within a particular cancer may also vary; different mutations can confer different cancer risk but also a specific mutation may provide different risks depending on the context.

2.5.2. Other forms of genetic cancer susceptibility

The inherited cancer syndromes explain only a fraction of familial risk[53]. It is recognized that much of the inherited cancer susceptibility is likely explained by polygenic inheritance. A combination of inherited genetic changes, each with a modest individual effect, can result in a wide range of risk in the population[50]. These alterations include both rare alleles with moderate penetrance and common alleles of low penetrance. The development of genome-wide association technologies has enabled the identification of hundreds of low-penetrant variants that confer a low risk of causing cancer[53]. Some of the single nucleotide variants (SNVs) associated with elevated cancer risk are cancer-specific, whereas others influence a variety of tumor types. In contrast to high-penetrance susceptibility changes, low-penetrance variants might show only little to no effect on tumor phenotype[50].

2.6. Next-generation sequencing

2.6.1. Sequencing efforts

The Human Genome Project, launched in 1990, was established to sequence the vast majority of the human genome to high accuracy. A decade later, in 2001, the first drafts of the human genome were published, followed by the first completed draft of the human genome three years later[54–56]. This effort was performed by Sanger sequencing and was thus both expensive and time-consuming. The new era of genomic analyses began when the first next generation sequencing (NGS) platform became commercially available. Compared to the Sanger method, NGS technologies provide massively parallel, high-throughput sequencing of short DNA fragments making it possible to sequence the human genome within days. The dramatic reduction in costs and run time together with the improved performance of bioinformatics tools and efficiency of data handling have enabled the generation of large-scale genomic data.

Today, it is routine to sequence multiple exomes (protein coding regions of the genome) or genomes simultaneously. The advances in sequencing technologies have also led to many collaborative efforts to sequence large study cohorts. The 1000 Genomes Project was established to comprehensively describe the common human genetic variation among multiple populations by constructing the genomes of approximately 2,500 individuals[57]. The Exome Aggregation Consortium (ExAC) collected and harmonized over 60,000 exomes from various population genetic and disease-specific studies and created a reference set for allele frequencies in different populations[58]. This was followed

by The Genome Aggregation Database (gnomAD) that includes over 125,000 exomes and over 15,000 genomes (set v2)[59].

Many efforts have also been conducted towards understanding the somatic events in disease progression, including cancer. The Cancer Genome Atlas (TCGA) has provided molecular characterization of over 11,000 cancers and matched normal tissues including exomes from 33 tumor types[60,61]. The Pan-Cancer Analysis of Whole Genomes (PCAWG) initiative has gathered genomes of over 2,600 tumors in order to define similarities and differences between genomic mutation patterns of various cancer types[62]. They detected no driver events in approximately 5% of the cancers, implicating that the catalogue of cancer drivers is still incomplete. In a clinical setting, the MSK-IMPACT sequencing panel was used to profile the somatic changes in over 10,000 metastatic cancers[18]. As a result, 37% of the patients harbored at least one clinically actionable mutation.

The increasing amount of genomic data has provided information on various tumor predisposition syndromes, advanced the understanding of tumor biology, and aided in identifying several disease mechanisms. The exploration of the tumor mutation landscapes has uncovered a repertoire of driver mutations in various cancer genes and identified mutational signatures derived from underlying biological processes that generate somatic mutations[4,63]. The revolution of NGS sequencing has also improved genetic diagnostics and provides opportunities for development of new targeted cancer treatments[18].

2.6.2. Cancer genome landscapes

The typical amount of mutations a cancer contains varies between tumor types. Common solid tumors are estimated to display mutations in an average of 33 to 66 genes[4]. Childhood tumors and leukemias contain, on average, 9.6 mutations, whereas melanomas and lung cancers harbor around 200 nonsynonymous mutations. In certain tissues where self-renewal occurs, the number of mutations is directly correlated with age[64]. Tissues with more rapid cellular turnover rate, and thus with more mitotic events, also gather more mutations. These explain partly the difference in mutation amounts between different tumors. Pediatric cancers usually develop from non-dividing tissues and those that do, such as leukemias, have not undergone as many mitotic divisions. Adult cancers have more mutations, and between them, CRCs harbor a greater amount of mutations than tumors from non-dividing tissues.

Some of the highest mutation rates are contributed by extensive exposure to well-known carcinogens, such as smoking in lung cancer and ultraviolet radiation in melanoma. This explains variation in the mutation amount between different cancers but also within a particular tumor type. Lung cancers of smokers, as an example, have ten times more mutations than those of non-smokers[65]. In addition, different exposures can have an impact on the distribution of mutation types. Lung cancers of smokers have been shown to exhibit more C:G>A:T mutations, whereas non-smokers have predominantly C:G>T:A changes. Mutation counts vary between cancer subtypes also due to key biological factors, such as high mutation rates in CRC due to MMR deficiency or defective DNA proofreading enzymes[4,24,45,46].

The mutation rate varies also within a cancer genome. As an example, open chromatin regions exhibit decreased mutation density compared to gene-poor and heterochromatic regions[66]. This is likely due to better accessibility of the DNA repair machinery to open chromatin areas. Regional mutational heterogeneity exists also between protein coding regions and has been proposed to occur e.g. due to gene expression level changes, that is, mutation rates of highly expressed genes are lower, and due to DNA replication time, since late-replicating regions have been shown to harbor higher mutation rates[67].

The genomic landscape of many tumor types consists of a few genes that are frequently mutated, so called "mountains", and a large number of other genes that are mutated infrequently, termed "hills"[68]. Since most of the genes are classified as hills, there is usually only little overlap between mutated genes of two tumors even within the same histopathologic category. In addition, the majority of the observed mutations are passenger events[4]. Due to variability of the background mutation rate along the genome and between patients, methods based on mutation frequency alone cannot reliably identify mutated driver genes. At best, they might aid in prioritizing genes for further analyses. Thus, it has been suggested that the pattern and impact of the mutations, rather than frequency, should be considered when estimating causality. To this end, many statistical approaches have been developed to prioritize which genes are most likely to promote tumorigenesis when mutated[69]. Still, distinguishing true driver mutations among the repertoire of somatic mutations continues to be a major challenge.

3. Colorectal and small bowel adenocarcinoma

3.1. Background and clinical aspects

The major organs that constitute the GI tract include the stomach, the small bowel, and the colorectum. Although it is a continuous passageway, cancer incidence varies greatly between the different organs. The small bowel makes up approximately 75% of the length of the GI tract but only approximately 3% of GI cancers are located in the small bowel[70]. Small bowel includes three segments, duodenum, jejunum, and ileum, of which duodenum is the most common location for adenocarcinomas[71]. Whereas, over 40% of GI cancers are found in the colorectum[70]. CRCs can be further divided depending on their anatomical location. Proximal (or right-sided) tumors entail tumors of the cecum, ascending colon, hepatic flexure, and two-thirds of transverse colon, whereas distal (or left-sided) tumors include tumors of the distal third of transverse colon, hepatic flexure, descending colon, sigmoid, and rectum[72]. Different molecular and histological characteristics have been associated with these groups, such as MSI which is more commonly found in right-sided tumors.

CRC is the third most common cancer type and the second main cause of cancer deaths worldwide[73]. In 2018, nearly 1.8 million new cases were diagnosed worldwide and, with estimated nearly 881,000 deaths, CRC accounted for over 9% of cancer-related deaths. Over 90% of cancers found in the colorectum are adenocarcinomas (here referred to as CRC). The mortality rates of CRC have been decreasing during the last two decades in many high-income countries[74]. This may at least partly be due to improved screening and identification of early stage disease, as well as developments in treatment regimens. Both genetic and environmental factors contribute to disease development.

Predisposing syndromes include known hereditary cancer syndromes, such as FAP and Lynch syndrome (LS, also known as hereditary non-polyposis colorectal cancer), and inflammatory bowel disease (IBD, including Crohn's disease and ulcerative colitis). Environmental and behavioral risk factors include obesity, consumption of red meat and alcohol, and lack of exercise and fiber[75]. Studies have indicated that use of aspirin might reduce CRC risk and improve survival[76,77]. These risk effects can sometimes be challenging to determine. As an example, an individual with a predisposing syndrome harbors a clearly increased genetic risk, whereas the contribution of germline variants to the risk in a sporadic CRC patient may be small compared to behavioral factors.

Approximately 4% of CRC cases present more than one simultaneous lesion in the colorectum[78–80]. These synchronous CRCs (SCRCs) have also been associated with the above-mentioned predisposing syndromes that, however, are estimated to account only slightly more than 10% of SCRC cases[79]. Another predisposing factor suggested as a potential causative mechanism is field effect, a process in which a large area of cells within histologically normal-appearing tissue is affected by molecular aberrations[81,82]. The proposed molecular basis for field effect includes massive exposure to a carcinogen, genetic predisposition through mosaicism, and epigenetic changes. Compared to solitary CRCs, some studies have associated SCRCs more often with MSI, mucinous histology, right-sided location, presence of adenomas, and older diagnosis age[83]. There is no consensus on whether prognosis differs between synchronous and solitary CRCs.

Of small bowel cancers, approximately one third are adenocarcinomas[71]. Other common histological types include carcinoids, lymphomas, and sarcomas. Small bowel adenocarcinoma (SBA) is an aggressive tumor type with an estimated five-year relative survival rate of 40%, indicating a worse prognosis compared to that of CRC. Reasons for the rarity of SBA are not fully known. SBA and CRC share common genetic and environmental risk factors listed above, in addition to which also celiac disease is known to predispose to SBA[84,85]. Small bowel as an environment is thought to protect more efficiently against carcinogens with fewer bacteria, shorter transit time, and the alkaline and liquid nature of small bowel contents[85]. The slower stem cell division rate in the small bowel compared to colorectum could also contribute to the difference in cancer incidence[10]. Regardless, the SBA incidence has been slowly rising[71]. Although detection methods have improved, SBAs are still often found at an advanced stage.

The American Joint Committee on Cancer/Union Internationale Contre le Cancer (AJCC/UICC) TNM staging system is the standard method to classify cancers and is used as a guideline for prognosis and treatment planning, see Table 1[86]. This was preceded by Dukes' classification (A-D) which is no longer recommended for clinical use. TNM consists of three descriptive measurements of the anatomical spreading of cancer: the extent of the primary tumor invasion (T), presence of cancer cells in regional lymph nodes (N), and whether the tumor has metastasized into distant organs (M). However, clinical outcomes can vary considerably among patients diagnosed with same stage tumors. While TNM measures tumor burden at the time of diagnosis, it does not entail information on the biological features of cancer.

Table 1. TNM classification of CRC and SBA. Basic principles of the staging of CRC and SBA.*

T - primary tumor			
T0	No evidence of primary tumor		
Tis	Carcinoma *in situ*		
T1	Invades the submucosa		
T2	Invades the muscularis propria		
T3	Invades through the muscularis propria		
T4	Invades the visceral peritoneum or adjacent organs		
N - regional lymph nodes			
N0	No regional lymph node metastasis		
N1	1-3 positive regional lymph nodes		
N2	4 or more positive regional lymph nodes		
M - distant metastasis			
M0	no distant metastasis		
M1	distant metastasis		
Stage	**T**	**N**	**M**
Stage 0	Tis	N0	M0
Stage I	T1-T2	N0	M0
Stage II	T3-T4	N0	M0
Stage III	T1-T4	N1-N2	M0
Stage IV	T1-T4	N0-N2	M1

*Modified from (Amin *et al*. 2018)

In addition to the traditional TNM classification, a method called Immunoscore® has been introduced in literature as a possible novel way to evaluate tumor prognosis. It was developed after the amount of immune infiltrate in CRCs were shown to correlate with better patient outcome[87,88]. Immunoscore combined with TNM classification has been suggested as a new predictor of survival and therapy response in CRC[89]. Its principle is based on determining the densities of CD3+ and CD8+ lymphocytes in the tumor centre and invasive margins. These data are categorized into scoring of 0-4 (0 reflecting a low and 4 a high density of both cell types in both regions) or, as in the newest consensus Immunoscore, from low to high.

The primary treatment for CRC, with the attempt of curative outcome, is surgical removal of the primary cancer (and/or limited metastasis) combined with chemoradiation in certain cases[90]. Patients with unresectable disease and eligible for systemic therapy can be treated with conventional chemotherapy or newer agents, such as EGFR antibodies, antiangiogenic agents, or multi-kinase inhibitors. Stage I and II tumors are locally invasive and can thus usually be cured by surgery. The majority of the tumors with regional lymph node spread (stage III) are also curable with a combination of surgery and adjuvant treatment. Tumors that have already metastasized (stage IV), however, most often cannot be cured. SBAs have been primarily treated similarly to CRC[91].

One example of molecular information translated into a clinical setting is the implementation of anti-EGFR treatment combined with the screening of *RAS* mutations to indicate who will benefit from the treatment[90]. New approaches in precision medicine are being investigated, such as immunotherapy for various tumor types. PD-1 inhibitor, pembrolizumab, has been shown to be effective in MSI tumors and the first U.S. Food and Drug Administration (FDA) approved cancer therapy towards a

specific molecular subclass of tumors[92]. Tumor mutation burden in general is a proxy for the amount of neo-antigens in the tumor cells and has become an emerging biomarker for immune checkpoint therapy response. Also, anti-HER2 therapy has shown promise in HER2 positive metastatic CRC[93].

3.2. Molecular pathogenesis

Adenocarcinomas of the colorectum and the small bowel arise from epithelial cells that are organized as a single layer lining the intestinal wall[94,95]. Intestinal epithelium is rapidly renewing and shaped into invaginations that form crypts protruding into the underlying connective tissue, and in small bowel, also luminal protrusions that form small bowel villi. At the bottom of the crypts reside multipotent stem cells that divide and are capable to give rise to all intestinal epithelial cell types. Thus, all cells within a crypt are clonal and derived from the stem cells. Division of stem cells in colorectal epithelium occurs every 24 hours giving rise to daughter cells. Stem cells of the small bowel are thought to divide somewhat less frequently[10,96]. When pushed away from the bottom of the crypts, daughter cells transform through transit-amplifying cells into mature cells, migrate up towards the colonic intercrypt plate or the adjacent small bowel villi, and finally shed off into the intestinal lumen.

Together with surrounding mesenchymal myofibroblasts, epithelial stem cells form a stem-cell niche[94]. In a normal state, myofibroblasts produce various homeostatic signals, such as Wnt signaling ligands, that drive epithelial cell proliferation[97]. The initial, gatekeeping mutation in carcinogenesis occurs in an epithelial stem cell providing the cell growth advantage to dominate the entire crypt through clonal expansion[94,98]. The mutant clones can then continue their expansion through clustering of crypts, called aberrant crypt foci, which is the earliest detectable step. The cells migrate up the crypt, fail to differentiate in a normal manner, and spread into the intestinal epithelium - a sequence which leads to the formation of a tumor.

Adenocarcinomas of both small bowel and colorectum are thought to arise from an adenomatous polyp and develop through an advanced dysplastic adenoma into an invasive adenocarcinoma[99,100]. The malignant transformation results from accumulation of genetic and epigenetic alterations. The mutation rates of some driver genes have been reported to differ in CRC and SBA[85,101]. The adenoma-carcinoma sequence is a multistep process that may take up to decades to evolve into a metastatic cancer[102]. During tumor evolution, the time needed for an advanced adenoma to grow into an advanced carcinoma is suggested to be much longer than for a carcinoma to metastasize[103]. The most common sites for metastatic lesions of intestinal cancers include liver, lung, and peritoneum[104,105]. Despite extensive efforts, the search for driver genes specific for metastatic traits has not provided clear results[4].

A key step in the intestinal tumorigenesis is the acquisition of genomic instability[43]. It is considered as an early event in tumor formation and leads to the accumulation of further genetic changes that drive carcinogenesis towards invasive adenocarcinoma[102]. Tumors can acquire genomic instability in various ways. Classically this has been divided into two distinct pathways, CIN and MSI, portrayed in Figure 2. Additionally, a small subset of CRCs is ultramutated due to mutations in the DNA proofreading domain of *POLE* or *POLD1*[45,46]. These genes encode for two of the main enzymes that replicate eukaryotic DNA. Genomic instability can also manifest through an epigenetic way.

Although one common characteristic of CRC is global hypomethylation, a subset of CRCs exhibits high methylation of CpG islands (including the promoter region of *MLH1*)[106,107]. These CpG island methylator phenotype CRCs display distinct histopathological and molecular characteristics, such as association with MSI, *BRAF* mutations, and serrated adenomas as the precursor[108].

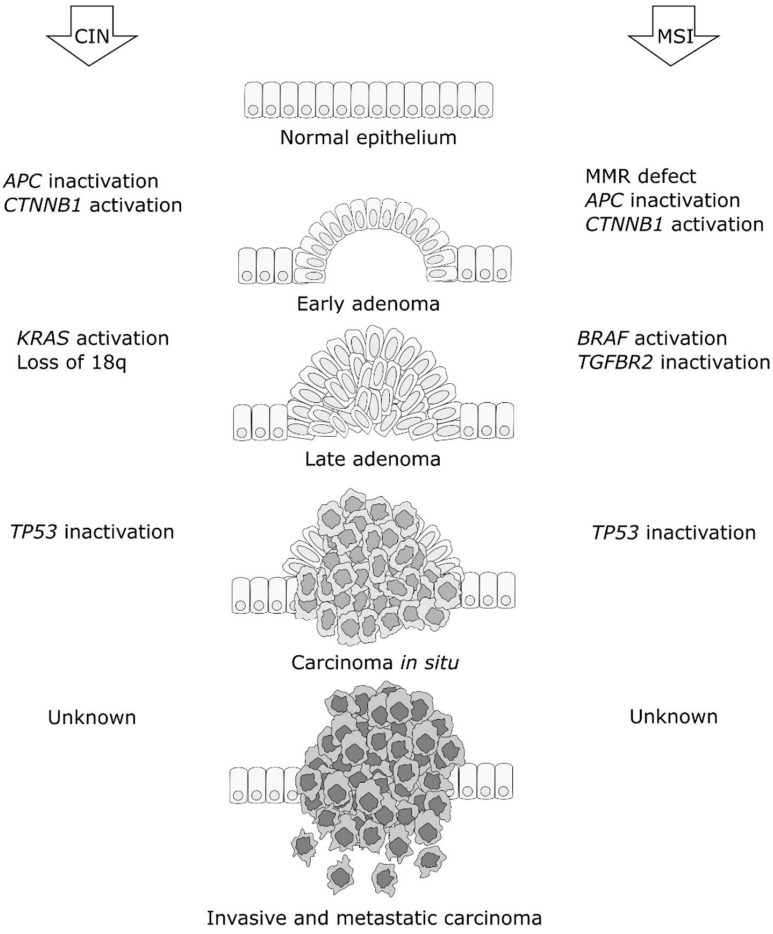

Figure 2. The stepwise progression of CRC. The main genetic alterations for CIN and MSI pathways are displayed.

3.2.1. Chromosomal instability

In 1990, Fearon and Vogelstein proposed a multistep genetic model of CRC genesis, where a cell accumulates sequential genetic changes that gives it growth advantage[32]. Although this theory has been refined throughout the years, it continues serving as the basis of the carcinogenesis model and characterizes the CIN pathway. Approximately 85% of CRCs are chromosomally unstable. Similar frequencies have also been reported in SBA[109]. CIN tumors are usually microsatellite stable (MSS) and such CRCs are often left-sided. CIN pathway consists of frequent chromosomal losses and gains, typically including losses on chromosomal arms 5q, 17p, and 18q and gains on 8q, 13, and 20q[32,100,110]. CIN has been demonstrated to be an early event in tumorigenesis and to increase during tumor progression[111].

According to the adenoma-carcinoma model, the first gatekeeping mutation happens usually in *APC* located at 5q. Biallelic loss of *APC* initiates tumorigenesis by activating Wnt signaling pathway leading to upregulation of oncogenic MYC[112,113]. Somatic inactivation of *APC* has been estimated to occur in approximately 70-80% of sporadic CRCs. In *APC* wild-type CRCs, alternative routes to constitutive Wnt signaling include activating *Catenin Beta 1* (*CTNNB1*) mutations or R-spondin fusion genes[100,114]. In SBA, *APC* mutations are thought to occur less frequently in approximately 5-30% of tumors[85,115]. *CTNNB1* mutations have also been reported to be relatively uncommon in SBA. Once Wnt signaling is incessantly activated, epithelial cells of the early adenoma grow into a physical mass that overgrows the surrounding mesenchymal tissue.

When the descendants of these cells acquire secondary advantageous changes, such as activating mutations in *KRAS*, cell number increases resulting in the growth and progress of the adenoma. *KRAS* mutations are typically found in adenomas larger than 1cm and in invasive adenocarcinomas[100,110]. Around 40% of CRCs and 50% of SBAs harbor a *KRAS* mutation, typically affecting codons 12 and 13 of exon 2 and, to a lesser extent, codon 61 of exon 3[100,115]. These mutations result in a constitutive activation of RAS downstream signaling and are mutually exclusive with downstream kinase *BRAF* V600E mutations. *BRAF* mutations are found in approximately 10% of CRCs, mainly in sporadic MSI CRC[45,116,117]. Up to one fifth of *BRAF* mutations found in CRCs are located outside the most common mutation hotspot, V600[118,119]. In SBA *BRAF* mutations occur at a relatively same frequency but the majority of the mutations have been proposed to constitute of atypical, non-V600 mutations[115,120]. Alternative mutational targets for activation of the RAF-MAPK pathway include *NRAS*, *EGFR*, and *ERBB2*[45]. Another frequently dysregulated signaling pathway is the phosphoinositide-3-kinase (PI3K) pathway, an alternative to MAPK signaling, typically affected by mutations in the tumor suppressor *PTEN* or oncogene *Phosphatidylinositol-4,5-Bisphosphate 3-Kinase Catalytic Subunit Alpha* (*PIK3CA*). MAPK signaling route is regulated by EGFR and thus cancers with mutated *KRAS*, and probably mutated *BRAF*, *PTEN*, and *PIK3CA*, do not respond to anti-EGFR antibody treatment[121–123].

A typically late event in carcinogenesis is the loss of tumor suppressor *TP53*[124]. In normal conditions, p53 coordinates cellular responses to stress, including DNA damage, aberrant proliferation signals, and oxidative stress. *TP53* is mutated in 60-70% of CRCs and in approximately 40-50% of SBAs[100,115]. It is usually targeted by chromosomal loss at 17p coupled with a somatic mutation in the remaining allele[125]. A great majority of the somatic mutations in *TP53* are missense changes[100]. The transforming growth factor beta (TGFβ) signaling pathway is also frequently dysregulated[45]. TGFβ signaling is usually affected by mutations in *TGFBR2* and *Activin A Receptor Type 2A* (*ACVR2A)* in MSI tumors, and in downstream mediators SMAD family members *SMAD2*, *SMAD3*, and *SMAD4* in CIN tumors.

The molecular basis for CIN is incompletely understood. It is not known whether the cause is genetic, epigenetic, or related to abnormal architectural features of a cancer cell[102]. One theory behind CIN is mutations or abnormal expression patterns of spindle-checkpoint genes, such as *BUB1* and *MAD2*, that regulate the mitotic spindle formation and the proper alignment and segregation of chromosomes during mitosis[102,126,127]. In addition, inactivation of *APC* has been thought to promote CIN through its role in maintaining chromosomal stability during mitosis[128]. Other factors suggested to contribute to

CIN include telomerase dysfunction, abnormal number of centrosomes, and disruption of genes involved in the regulation of DNA damage and replication checkpoints[111].

3.2.2. Microsatellite instability

In 1993, a subset of CRCs was recognized to harbor a distinct molecular phenotype with a large number of replication errors located frequently at the repetitive DNA tracks[129–131]. This, later named MSI, is exhibited by approximately 15% of CRCs[24]. Of MSI CRCs, 80% arise sporadically and 20% due to LS. In SBA, 5-35% of tumors are estimated to display MSI[109].

MSI tumors arise due to a faulty DNA MMR machinery. LS patients carry a defective allele of one of the MMR genes *MLH1*, *MSH2, MSH6*, or *PMS2*. Germline mutations in *MLH1* and *MSH2* account for a majority of these cases[100]. Second inactivating hit occurs sporadically leading to MSI. MSI in sporadic tumors is most often caused by hypermethylation of both alleles of the *MLH1* promoter that results in the loss of MLH1 expression[132]. MSI tumors usually have near-diploid karyotype[24]. They have been associated with proximal location in colon, poor differentiation, mucinous histology, infiltration of lymphocytes, better prognosis, low metastatic potential, and they are shown to be more resistant to 5-fluorouracil-based adjuvant chemotherapy than MSS tumors[24,133,134]. Sporadic MSI tumors also occur generally at an older age and have sessile-serrated adenoma as the precursor[100,135].

DNA MMR system recognizes and orchestrates the repair of nucleotide mismatches after DNA replication[136]. This system is highly conserved from bacteria to humans. The main components of this system are MutSα (a heterodimer of MSH2 and MSH6), MutSβ (a heterodimer of MSH2 and MSH3), and MutLα (a heterodimer of MLH1 and PMS2). Either MutSα or MutSβ detects a mismatch or an insertion-deletion loop and creates a sliding clamp around the DNA. This is then bound by MutLα which activates the process of repairing the mistake.

A defective MMR system results in the accumulation of mutations in the DNA, especially in areas called microsatellites. These short repetitive stretches of DNA can be composed of mono-, di-, tri-, or tetranucleotide repeats. They are found both in coding and non-coding regions and are prone to expansion or shortening (i.e. instability). Insertions and deletions in microsatellites of the coding sequence typically cause frameshifts that can lead to truncated, functionally inactive proteins. Also, the point mutation rate is elevated[24]. The overall mutation rate has been estimated to be a 100-fold higher in MSI compared to MSS cancers[45,137]. Ultimately, clones that acquire mutations in key driver genes due to faulty MMR, gain growth advantage that promotes tumorigenesis[26].

Since a majority of the mutations in MSI cancers are passenger events, it has been challenging to distinguish drivers that contribute to tumorigenesis. Many criteria have been proposed for defining true MSI target genes, such as high frequency of biallelic inactivating changes, association to a growth suppressive pathway, the same pathway inactivated in MSS tumors, and functional evidence[138]. Many candidate MSI target genes have been reported, however, only few genes have robust functional evidence. These well-established MSI target genes include *TGFBR2* and *ACVR2A*[139,140]. Additionally, only a few oncogenic changes have been confirmed to have a role in MSI carcinogenesis. The best characterized is a hotspot mutation *BRAF* V600E, occurring in over 40% of sporadic MSI CRCs[26,45]. Other genes harboring causative point mutation hotspots include *KRAS*,

CTNNB1, and *PIK3CA*[100]. Compared to MSS CRC, *APC*, *KRAS*, and *TP53* generally exhibit less frequent mutations in MSI tumors[45,141].

The standard method of detecting MSI is the polymerase chain reaction (PCR)-based Bethesda panel that consists of five markers: two mononucleotide repeats (BAT25 and BAT26) and three dinucleotide repeats (D5S346, D2S123, and D17S250)[142]. According to the guidelines, a sample is considered MSI-high when two or more markers show instability[138]. If only one marker is mutated, the tumor is referred to as MSI-low. Discontinuation of this category has been proposed, however, since MSI-low tumors display biological similarities with MSS tumors[143]. Another approach is to perform immunohistochemical (IHC) staining on the MMR proteins. A tumor is diagnosed with MMR deficiency if expression of one or more of these proteins is abnormal. This method is often utilized in the clinical setting and its sensitivity and specificity corresponds to that of the Bethesda panel[24]. MMR genes with abnormal expression may be sequenced to further determine the causative change. Additionally, NGS has provided opportunities to detect MSI status from large-scale sequencing data with high accuracy[144]. This has led to the development of multiple NGS-based MSI classifiers[145–147]. However, no consensus on the calling parameters have thus far been determined. Also, normal tissue contamination in DNA-based testing or interpretation of antibody-based tests can sometimes create challenges for obtaining accurate results.

3.2.3. Advances in characterizing the genetic landscape

In recent years, large-scale NGS efforts have been conducted to improve our understanding of the CRC mutation landscape. Knowledge on the general amount and distribution of mutations has revealed CRCs to harbor only a handful of commonly mutated genes[68]. Each tumor contains a unique set of mutations of which the majority are passengers. Hence, statistical methods have been developed to evaluate the significance of mutations and this has led to the discovery of new driver genes associated with various biological processes[45,68,148]. Examples of novel driver genes include *ARID1A*, involved in chromatin remodeling, and *POLE*, where somatic mutations in the proofreading domain cause the ultramutated phenotype found in approximately 1% of CRCs[45]. Also novel non-coding mutational patterns have been discovered in CRC[149,150].

Large-scale sequencing data has also enabled the detection of mutational signatures[45,63]. Different mutational processes, such as mistakes occurring during DNA replication or mutagen exposures, generate unique combinations of mutation types, referred to as mutational signatures. Mutational signatures (and their proposed etiology) found in CRC include 1A (aging), 6 (MSI), 10 (*POLE* mutations), and 17 (cause unknown)[63,149]. CRCs have also been classified into four distinct groups by gene expression studies[151]. These consensus molecular subtypes are CMS1 (MSI immune), CMS2 (canonical), CMS3 (metabolic), and CMS4 (mesenchymal).

Such large-scale studies have not been conducted on SBA. Therefore, less is known about the mutation landscape specific to this tumor type. Studies have mainly focused on known mutational hotspots or a set of known cancer genes[115,120,152]. There have been a few exome-wide studies on a limited number of duodenal tumors[153,154]. These have shown that the most frequently affected genes include similar driver genes as in CRC, such as *TP53*, *KRAS*, *SMAD4*, and *APC*[85,115]. DNA copy number changes in SBA might be more similar to those observed in CRC than in gastric cancer[155].

3.3. Hereditary predisposition

Approximately one-third of all CRCs are thought to arise due to familial predisposition[156]. Mendelian cancer syndromes resulting from a high penetrance germline mutation have been estimated to account for approximately one fifth of these cases[157]. The majority of SBAs are also assumed sporadic[101]. The most common hereditary syndromes predisposing to CRC and SBA are LS and FAP (Table 2.)[101,157]. The next most common syndromes include *MUTYH*-associated polyposis (MAP), juvenile polyposis syndrome (JPS), and Peutz-Jeghers syndrome (PJS). LS, FAP, and MAP are characterized by the development of adenomatous polyps, whereas in JPS and PJS the primary lesions are hamartomatous polyps.

Table 2. Summary of the most common hereditary syndromes for CRC and SBA.

Syndrome	Gene(s)	Gene Function	Inheritance
LS	*MLH1, MSH2, MSH6, PMS2*	DNA mismatch repair	autosomal dominant
FAP	*APC*	Wnt signalling	autosomal dominant
MAP	*MUTYH*	DNA base excision repair	autosomal recessive
JPS	*SMAD4, BMPR1A*	TGFβ signalling	autosomal dominant
PJS	*STK11*	Activation of AMPK-related kinases	autosomal dominant

Other rarer syndromes predisposing to CRC also exist, such as serrated polyposis and polymerase proofreading–associated polyposis. Serrated polyposis is characterized by several serrated polyps in the colon and, on average, a five-fold increased risk for CRC[158]. Genetic background of the disease remains largely unknown. Polymerase proofreading-associated polyposis is an autosomal dominant disease caused by germline mutations in the exonuclease domains of *POLE* and *POLD1*[46]. This leads to hypermutated tumors characterized mainly by increased rate of base substitutions. Patients seem to exhibit variable phenotypes that include several colorectal adenomas and carcinomas[159].

Much of the heritable risk for CRC remains molecularly unexplained. It is thought to derive from alleles with lower penetrance, some of which have been identified[160]. One well-established moderate-penetrance variant is *APC* I1307K that is found in ~6% of Ashkenazi Jews and increases the risk for CRC two-fold[161]. Genome-wide association studies have identified many loci with low-penetrance alleles associated with CRC[162–165]. One of the most interesting regions found in these studies is located at 8q24, with the polymorphism rs6983267[166]. Later studies have identified a cancer-specific enhancer element at this region that has been proposed to control the expression of *MYC* oncogene[167,168]. In general, the distinction between familial (excluding the cases due to a well-known predisposing variant) and sporadic CRCs can be difficult since a large number of predisposition alleles with varying risk levels and prevalence in the population are likely to collectively contribute to the genetic susceptibility to CRC in both cases[160].

3.3.1. Lynch syndrome

LS, the most common genetic predisposition syndrome for CRC and SBA, accounts for approximately 5% of the cases in both tumor types[101,160]. Among mutation carriers, lifetime risk for developing MSI SBA is 2-8% and MSI CRC 39–70%[85]. LS is inherited in an autosomal dominant

manner and predisposes to various types of MSI tumors, especially colon and endometrial[51,169]. Other cancers seen in LS include stomach, small bowel, ovary, hepatobiliary system, and upper urinary tract. Patients with LS display also more synchronous and metachronous MSI CRCs. Compared to sporadic cases, LS-associated CRCs are usually diagnosed at an earlier age, with the average age of onset approximately 45 years, and have a better prognosis[51,170]. LS-associated tumors rarely harbor somatic *BRAF* V600E mutation which is found in approximately 40% of sporadic MSI CRCs[116,171].

In 1993, a linkage to 2p15-16 was observed in many LS families providing the first susceptibility locus for the syndrome [172]. Subsequently, *MSH2* was identified as the predisposing gene[173,174]. At the same time, LS tumors were shown to display variation in microsatellite lengths, linking the syndrome to defective MMR[129]. This was followed by the discovery of *MLH1* germline mutations at 3p21[175–177]. Later, germline mutations in two other key MMR genes, *MSH6* and *PMS2*, and deletion of the 3' end of *EPCAM* have also been linked to LS[178–181]. Together, *MLH1* and *MSH2* germline mutations explain approximately 90% of LS cases[182]. In Finland, two founder mutations in *MLH1* account for approximately 60% of inherited LS mutations[183].

Several international diagnostic criteria have been introduced for identifying LS patients. These include Amsterdam criteria I & II and the Bethesda guidelines, see Table 3[142,184–186]. Since there are no clinically distinct features associated with LS, the diagnosis is confirmed only by a gene test. Furthermore, since the clinical criteria are not sensitive enough to catch all LS cases and MSI status affects prognosis and possibly the adjuvant treatment, many international clinical guidelines recommend MSI screening of all newly diagnosed CRC cases[187]. For mutation carriers, surveillance colonoscopy is recommended every 1-3 years from the age 20-25 onwards[188–190]. The surveillance recommendations for LS-associated extracolonic cancers vary.

Table 3. Diagnostic criteria for identification of LS patients and the guidelines for testing colorectal tumors for MSI.

The revised ICG-HNPCC Criteria (Amsterdam criteria II) *
There should be at least 3 relatives with an LS-associated cancer (CRC, cancer of the endometrium, small bowel, ureter, or renal pelvis)
One should be a first-degree relative of the other 2
At least 2 successive generations should be affected
At least 1 should be diagnosed before age 50
Familial adenomatous polyposis should be excluded in the CRC case(s) if any
Tumors should be verified by pathological examination
The revised Bethesda guidelines **
Tumors from individuals should be tested for MSI in the following situations:
1) CRC diagnosed in a patient who is less than 50 years of age
2) Presence of synchronous, metachronous colorectal, or other HNPCC associated tumors, regardless of age.
3) CRC with the MSI histology (Presence of tumor infiltrating lymphocytes, Crohn's-like lymphocytic reaction, mucinous/signet-ring differentiation, or medullary growth pattern) diagnosed in a patient who is less than 60 years of age.
4) CRC diagnosed in one or more first-degree relatives with an HNPCC-related tumor, with one of the cancers being diagnosed under age 50 years.
5) CRC diagnosed in two or more first- or second-degree relatives with HNPCC-related tumors, regardless of age.

*Reprinted from Gastroenterology 116/6, Vasen *et al*. New clinical criteria for hereditary nonpolyposis colorectal cancer (HNPCC, Lynch syndrome) proposed by the International Collaborative group on HNPCC, p.1455, Copyright (1999), with permission from Elsevier; **Umar et al. Revised Bethesda Guidelines for hereditary nonpolyposis colorectal cancer (Lynch syndrome) and microsatellite instability, J Natl Cancer Inst. 2004, 96/4, p.15, by permission of Oxford University Press.

3.3.2. Familial adenomatous polyposis

Approximately 0.5% of CRCs are due to FAP, making it the second most common inherited CRC syndrome[191]. In SBA, FAP is estimated to be present in less than 5% of cases[101]. FAP is characterized by hundreds to thousands of adenomas in the colorectum, typically beginning to develop during adolescence. In case prophylactic surgery is not performed, CRC will develop in nearly all affected individuals by the age of 50[157]. Extracolonic manifestations include hypertrophy of retinal pigment epithelium, gastric fundic gland polyps, and duodenal adenomas[51]. FAP patients have also an increased risk for extracolonic malignancies. Of these, duodenal cancer is the most common with a lifetime risk up to 10%[190]. After the removal of CRC risk by total colectomy, duodenal cancer remains as one of the main causes for mortality[192].

This autosomal dominant condition is caused by inactivating germline mutations in *APC*. Protein truncating mutations make up over 95% of the inherited mutations[160]. Up to 25% of FAP patients, however, have a negative family history and present a *de novo* germline mutation in *APC*[193]. A clinical diagnosis of FAP can be made when over a hundred colorectal adenomas are detected[51]. A gene test is recommended to confirm the diagnosis. Genetic testing is also recommended for individuals with 10-20 colorectal adenomas, typical extraintestinal manifestations, or if there is a known *APC* mutation in the family. Annual endoscopy surveillance should be offered for all mutation carriers[190]. The primary prophylactic measure for FAP is colectomy or proctocolectomy.

Attenuated FAP is a less severe form of the disease in which the amount of polyps varies between 10 and 100 and has on average a later age of onset[194]. Another variation of FAP is called Gardner syndrome, where in addition to FAP, patients can present desmoid tumors, epidermoid cysts, osteomas, and dental anomalies[51].

3.3.3. Other syndromes

MAP is caused by biallelic mutations in the gene *MUTYH* and is inherited in an autosomal recessive manner[195]. *MUTYH* is a base-excision repair gene and when impaired, tumors gather an excess amount of G:C>T:A transversions. The disease manifests usually by the time patients reach their sixties[196]. The clinical phenotype resembles that of attenuated FAP and thus genetic testing of *MUTYH* is warranted in patients with more than ten adenomas and no detectable *APC* mutation[157]. Unlike in attenuated FAP, however, MAP patients present hyperplastic polyps in addition to adenomatous polyps. The risk for duodenal cancers is slightly elevated[197].

PJS is an autosomal dominant syndrome characterized by hamartomatous polyps along the GI tract, located most often in the small bowel[157]. PJS predisposition has been linked to germline mutations in *STK11* (also known as *LKB1*) that encodes for a serine/threonine kinase[198]. The syndrome has almost complete penetrance. GI symptoms, such as obstruction and bleeding, usually occur already during early teens[157]. The most typical extracolonic manifestation is mucocutaneous pigmentation. Mutation carriers have an increased risk for multiple malignancies, such as female breast, colon, pancreas, stomach, ovary, lung, and small bowel cancers[199]. Although PJS polyps occur most frequently in the small bowel, the most common site for GI cancer is the colon.

JPS is also a hamartomatous polyposis syndrome with autosomal dominant inheritance and almost complete penetrance[157]. Germline mutations in TGFβ signaling pathway genes *SMAD4* or *BMPR1A* are known to predispose to JPS[200,201]. Unlike PJS, JPS does not typically have clinical manifestations to facilitate diagnosis[157]. Juvenile polyps are most commonly found in the colorectum but also in the small bowel and the stomach. Mutation carriers harbor an increased risk for CRC, gastric, small bowel, and pancreatic cancer. Congenital defects are present in approximately 15% of JPS patients.

AIMS OF THE STUDY

The main aim of this thesis work was to provide further insight into the genetic background of CRC and SBA utilizing exome sequencing. The specific aims are listed below.

I. To study the mutation pattern of ARID gene family in MSI CRC

II. To identify novel oncogenes with somatic mutation hotspots in MSI CRC

III. To examine the genetic overlap within synchronous CRC pairs

IV. To characterize the somatic mutational landscape of SBA

MATERIALS AND METHODS

The materials and methods used in this thesis work are presented briefly below. More detailed descriptions can be found in the original publications (referred to by their Roman numerals).

1. Ethical approval

The studies were conducted according to the Declaration of Helsinki and approved by the Ethics Committee of the Hospital District of Helsinki and Uusimaa, Finland (408/13/03/03/2009). Either a signed informed consent or authorization from the National Supervisory for Welfare and Health was obtained for all the samples, as determined in the National legislation.

2. Sample material

2.1. Microsatellite unstable colorectal cancers (I-II)

The discovery set of 25 sporadic MSI CRCs and their corresponding normal samples for studies I-II was derived from a population-based series of 1,042 Finnish CRCs collected 1994-1998 (the C series)[202,203]. Collection was continued with a subsequent, ongoing cohort of Finnish CRCs (the S series) collected in two Finnish central hospitals since 1998[204]. These collections consist of fresh-frozen tumors and their corresponding fresh-frozen healthy colorectal tissue samples or blood. Tumor samples have been examined for MSI by radioactive labeling techniques, fluorescence-based PCR, or fragment analysis[202–204]. Detailed clinicopathological information was available for all patients.

In study I, an independent validation set consisting of 21 sporadic MSI CRCs was obtained from the two above-mentioned cohorts. In study II, the independent validation set was composed of 167 Finnish MSI CRCs (42 LS and 125 sporadic cases) from the two above-mentioned cohorts and 87 Danish MSI CRCs. Corresponding normal samples were available for somatic validation of the findings.

2.2. Synchronous colorectal cancers (III)

The sample set in study III consisted of 23 SCRC pairs and their corresponding normal tissue samples. Of these, 20 patients were obtained from a population-based series of 1,088 CRCs from Jyväskylä, Finland and three patients from the above-mentioned C and S series. We included all patients with available tumor material from the paired CRC tumors. The set included 59 formalin-fixed paraffin-embedded (FFPE) samples and 11 fresh-frozen samples. From one patient we had three tumors to study and this patient belonged to a LS family. Other cases were assumed sporadic. All relevant clinicopathological information was available. For Jyväskylä samples, MMR status was determined by IHC staining.

2.3. Small bowel adenocarcinomas (IV)

We obtained information on all patients diagnosed with SBA in Finland during years 2003-2011 from the Finnish Cancer registry. Cases reported without histopathological confirmation of small bowel primary tumor or only by autopsy were excluded. Also, tumors of the papillary region were excluded as these may have derived from the pancreas or the biliary tract. We examined the FFPE tissue blocks of the remaining 162 patients and selected all cases with available tumor material and a minimum of 50% tumor content. Of these, one case was later omitted due to low sequencing depth. Thus, the final set was composed of 106/162 (65%) confirmed SBA cases (excluding autopsies). The set included tumors from all three segments of the small bowel (26 duodenal, 52 jejunal, and 18 ileal tumors) together with ten tumors without a specified location. Six patients had a hereditary syndrome (four LS and two FAP), the rest were assumed sporadic. All relevant medical records for the cases were available for the study.

2.4. Cell lines (II)

Genomic DNA from ten CRC cell lines (VACO5, CCL231, GP5D, HCA7, HCT-116, LOVO, LS174T, RKO, SNUC2B, and DLD1) were utilized for further validation of mutation hotspots. The cell lines were obtained from the American Type Culture Collection, the European Collection of Cell Culture, or provided by Professor Ian Tomlinson.

3. Genetic analyses

3.1. DNA extraction (I-IV)

The extraction of genomic DNA for fresh-frozen tissue was done with a non-enzymatic DNA extraction protocol (I-III), and for FFPE blocks with a standard phenol-chloroform isolation method (III-IV) or GeneRead FFPE-kit (QIAGEN, Hilden, Germany) (IV).

3.2. Exome sequencing (I-IV)

Genomic libraries were prepared with NEBNext DNA Sample Prep Reagent Set 1 Kit (New England Biolabs, Ipswich, MA, USA) (I-II) or KAPA Hyper Prep Kit (Kapa Biosystems, Wilmington, MA, USA) (III-IV). Exomic regions were enriched with the Agilent SureSelect All Exon Kit v1 (Agilent, Santa Clara, CA) (I-II) or the NimbleGen SeqCap EZ Exome Library v3 Kit (Roche NimbleGen, Madison, WI, USA) (III-IV). For I-II, paired-end short read sequencing was performed with Illumina Genome Analyzer II (Illumina, San Diego, CA, USA) at the Karolinska Institute, Sweden or the Finnish Institute for Molecular Medicine (FIMM) Genome and Technology Center, Finland. For III and IV, the paired-end sequencing was carried out with Illumina HiSeq 2000 or 4000 (Illumina) at the Karolinska Institute.

The raw sequencing data was run through an in-house analysis pipeline constructed for exome sequencing data. This included the following tools and programs: FastQC (http://www.bioinformatics.bbsrc.ac.uk/projects/fastqc/) (I-IV), QualiMap[205] (III-IV), Trim Galore!

(http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/) (III-IV), Burrows-Wheeler Aligner[206] (I-IV), BamUtil ClipOverlap (http://genome.sph.umich.edu/wiki/BamUtil#Releases) (III-IV), Picard Tools MarkDuplicate (http://picard.sourceforge.net) (I-II) or Samtools rmdup (III-IV)[207], the Genome Analysis ToolKit (GATK) IndelRealigner[208] (I-IV), GATK BaseRecalibrator (I-IV), Samtools mpileup (I-II)[207], and the GATK UnifiedGenotyper (I-II) or GATK HaplotypeCaller (III-IV). Instead of Trim Galore!, an in-house script was used in studies I-II to remove 3' read ends with high adapter similarity. Samtools rmdup was used to remove duplicate reads on both paired-end and single-end reads to correct for e.g. FFPE-derived sequencing artefacts. The version GRCh37/Hg19 was used as the human reference sequence.

In studies I-III, somatic mutations were extracted by filtering the tumor exome data against the exome data from the corresponding normal tissues. The remaining variant calls were refined by filtering them either against 93 Finnish samples from the 1000 Genomes Project (www.1000genomes.org/), single nucleotide polymorphisms (SNPs) available at the dbSNP database (Build 132; www.ncbi.nml.nih.gov/SNP/), and 69 population matched exomes from other projects in the laboratory (I-II) or a pooled set of whole-genome sequencing data from ten blood samples (III). For study III, germline mutations were obtained by filtering the normal tissue exome data against the whole gnomAD (http://gnomad.broadinstitute.org/). Germline mutations were excluded if they were present with a MAF >0.001 in the whole gnomAD or in the Finnish gnomAD exomes, or >0.01 in the Finnish genome set. In study IV, somatic mutations were called from the tumor exome data using similar methods to those reported in Hiltemann *et al.*[209] together with additional filtering steps and larger population and sequencing pipeline specific datasets (in-house whole genome sequenced (WGS) normal samples (n = 183), a pooled set of ten blood sample genomes, and gnomAD set of exome and WGS normal samples (n = 138,632)). BasePlayer (earlier version called RikuRator) was used to visualize and analyze the exome data (I-IV)[210]. The specific criteria to call a mutation varied between studies: GATK quality score at least 20 (I, III) or 50 (II), minimum total coverage 4 (III) or 5 (I-II), and the minimum mutant allele fraction 10% (III) or 20% (I-II). For study IV, recommended GATK hard filters were used (see Supplementary Table S8 in the original publication). Filtering steps, including mapping quality, were used to filter out false calls that derive e.g. from FFPE material. In addition, 1000 Genomes phase 1 pilot-style callability region mask was used to filter out possible false calls, such as deamination artefacts due to FFPE, located in regions with poor callability (III).

3.3. Sanger sequencing (I-IV)

Sequencing primers were designed with ExonPrimer (http://ihg.gsf.de/ihg/ExonPrimer.html) (I-II) and Primer3Plus (I-IV)[211]. For FFPE samples, PCR reactions were performed in triplicates to ensure consistency of the observations. Sequencing reactions were performed with the Big Dye Terminator v.3.1 kit (Applied Biosystems, Foster City, CA, USA) on an ABI3730 Automatic DNA Sequencer (FIMM Technology Center and DNA sequencing and Genomics laboratory, Institute of Biotechnology, Helsinki, Finland). The sequence graphs were analyzed both with the Mutation Surveyor software (Softgenetics, State College, PA) and manually.

Sanger sequencing of genomic DNA was used for validation of selected findings, screening of additional tumors, and verification of the somatic nature of these findings (I-IV). To study the expression of *ANTXR Cell Adhesion Molecule 1* (*ANTXR1*) mutant alleles, cDNA sequencing was

applied to seven *ANTXR1* mutated tumors from which RNA isolated from fresh-frozen tumor tissue was available (II).

4. Computational methods

4.1. In silico mutation effect prediction (I-II)

Ensembl Variant Effect Predictor (VEP) that includes SIFT (http://sift.jcvi.org/) and PolyPhen-2 (http://genetics.bwh.harvard.edu/pph2/) was ran to predict the functional consequences of the mutations (I-II). ClustalW2 (http://www.ebi.ac.uk/Tools/msa/clustalw2/) was used to perform multispecies protein sequence alignments (II).

4.2. Statistical analyses (III-IV)

R was used to perform statistical analyses in studies III (v.3.5.1) and IV (v.3.4.1) (R Foundation for Statistical Computing, Vienna, Austria; https://www.R-project.org/). In study III, Spearman's rank correlation coefficient was applied to evaluate the correlation of IS within tumor pairs and generalized estimating equations (GEE) with ordinal response were used to examine the association between IS and clinicopathological variables[212]. For study IV, Fisher's exact test was used to test for independence of the categorical variables. For continuous variables, Mann-Whitney U test was applied. Cox proportional hazards regression with Firth's penalized likelihood was used to analyze the disease-specific survival. This model included MMR status, sex, tumor stage, and age at diagnosis as covariates. Negative binomial regression was used to estimate per-tumor mutation counts attributable to mutational signatures in MSS tumors and whether they associate to any clinical variables. *P*-values were two-sided, unadjusted for multiple comparison, and $P < 0.05$ was regarded as statistically significant.

4.3. Mutation signature analysis (III-IV)

In study III, an in-house method was applied to detect mutational signatures from the exome data. Mixed linear model was used to compare the variation of signature contents between tumors. A detailed description of these models can be found in the original publication. Comparison of the modeled signatures to the published signatures was performed using Maximum-A-Posteriori probability signatures[63].

In study IV, using all mutations within exome target regions, mutation signature analysis was performed by non-negative matrix factorization of six substitution types. The obtained signatures were compared to the published signatures[213]. Fifteen tumors displayed MSI signature (signature 6). These MSI tumors were grouped separately and the mutation signature analysis was subsequently performed for the 91 MSS samples.

4.4. Ingenuity Pathway Analysis (III-IV)

Ingenuity Pathway Analysis (IPA) was used to determine the frequency of which known cancer pathways (ATM, p53, Wnt/β-catenin, ERBB, PTEN, PI3K/AKT, MAPK/ERK, and TGFβ) were affected within tumor pairs (III) and in MSS SBAs (IV). We used this program to define genes linked to these pathways. The analysis considered all genes with at least one non-synonymous mutation.

4.5. OncodriveFML (IV)

OncodriveFML was utilized to identify genes displaying statistical evidence of positive selection for mutations in the 91 MSS SBAs. This permutation-based method used the somatic mutations within the coding area and compared the region's mean functional impact score to its null distribution by randomizing observed mutations within the coding sequence. Default settings were used. The performed analysis focused on genes mutated in at least four tumors.

4.6. Allelic imbalance analysis (IV)

In study IV, allelic imbalance (AI) regions were called in the 106 tumors using SNVs that filled the following criteria: rs-coded, minimum of 10 coverage at the variant call locus, not defined as somatic in the study, location within exome target region, and did not overlap with regions prone to false AI calls. The AI regions were called with B allele frequency segmentation algorithm[214].

5. Immune cell score determination

In study III, the immune cell score (IS) was determined using the same basic principles as in Immunoscore®[88]. The IS consists of CD3 and CD8 lymphocyte counts from tumor centre and invasive margin. CD3 (Novocastra, NCL-L-CD3, clone PS1) and CD8 (Thermo Scientific, RM-9116, clone SP16) antibodies were used for IHC staining of whole tissue sections. Slides of the tissue sections were then digitally scanned with NanoZoomer-XR (Hamamatsu Photonics K.K., Hamamatsu City, Japan) and QuPath image analysis software was used to calculate the amount of positively stained T-cells[215]. To form IS, similar cut-off values were used as in a previous study with a larger study population[216].

RESULTS

1. Mutation profiles of ARID genes in MSI CRC (I)

1.1. Screening of the 15 ARID domain containing genes

We studied the somatic mutation pattern of all 15 ARID domain containing genes in 25 MSI CRCs. The following 12 ARID genes harbored at least one mutation: *ARID1A*, *ARID1B*, *ARID2*, *ARID3A*, *ARID4A*, *ARID4B*, *ARID5A*, *JARID2*, *KDM5A*, *KDM5B*, *KDM5C*, and *KDM5D*. Their mutation frequency varied between 4-52% (1-13 mutations in the 25 tumors). No mutations were found in *ARID3B*, *ARID3C*, or *ARID5B*. Twenty out of the 25 tumors (80%) presented at least one nonsynonymous mutation in one of the 12 mutated genes.

1.2. Four ARID genes frequently mutated

Four genes, *ARID1A*, *ARID1B*, *ARID2*, and *ARID4A*, were chosen for further analysis since they fulfilled the criteria of being mutated in at least four tumors and included at least one splice-site or nonsense mutation. Mutations encountered in these four genes in exome data were validated and their somatic origin was confirmed by Sanger sequencing. Furthermore, the coding regions of these genes were screened with Sanger sequencing in an additional set of 21 MSI CRCs.

For all four genes the relatively high frequency of nonsynonymous mutations was confirmed. We observed 6-18 nonsynonymous mutations per gene in the 46 tumors (13-39%), see Figure 3. A majority of the tumors (28/46, 61%) harbored a nonsynonymous mutation in at least one of the four genes, and nine tumors contained mutations in two or more of the four genes. The mutation spectrum included 47 different mutations: 18 frameshift (38%), 18 missense (38%), seven synonymous (15%), three nonsense (6%), and one splice-site (2%) mutations. Besides a mutation cluster in a known mutation hotspot of a mononucleotide repeat in *ARID1A*, p.D1850fs, mutations were scattered along the genes.

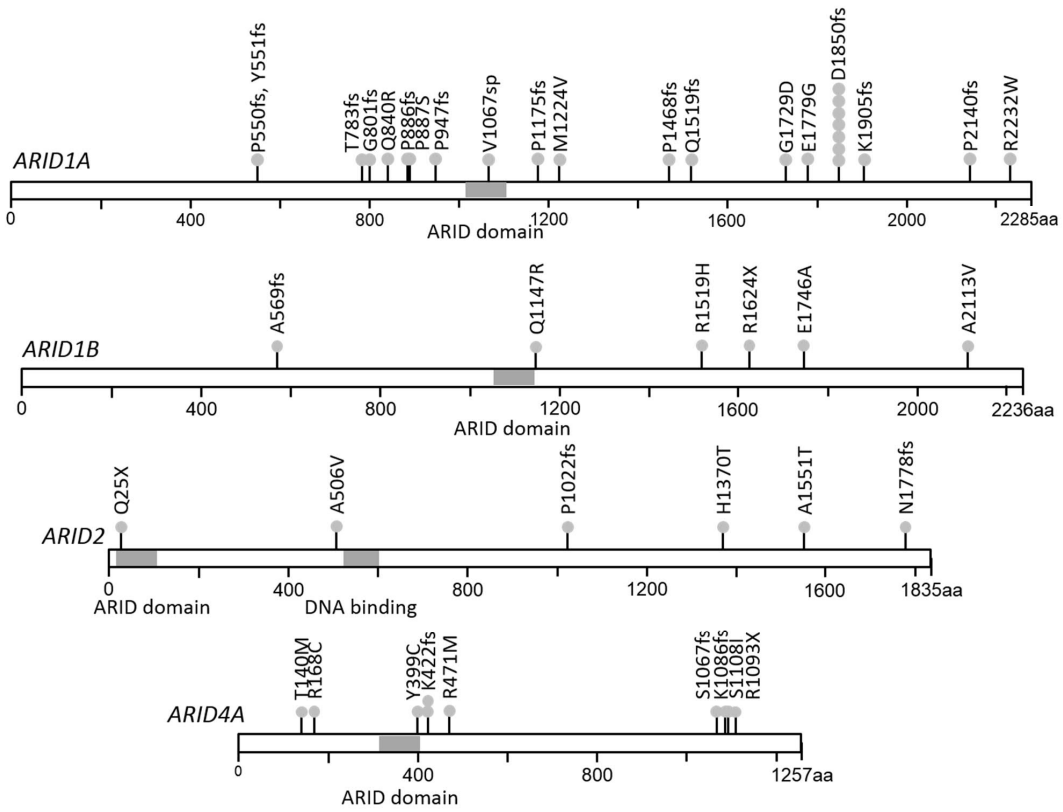

Figure 3. Mutations in *ARID1A*, *ARID1B*, *ARID2*, and *ARID4A*. Distribution of somatic, non-synonymous mutations detected by sequencing in the coding regions of these genes are presented. Each dot represents one mutation. Only some domains are displayed for clarity. Copyright (2014) Wiley. Reproduced with permission from Cajuso *et al*, 2014, IJC, Wiley.

ARID1A, the most frequently mutated of the four genes, harbored a nonsynonymous mutation in 18 out of 46 tumors (39%). Seven of these tumors contained two *ARID1A* mutations. Altogether 19 different mutations were observed. The majority of these were frameshift mutations (12/19, 63%). The above-mentioned known hotspot mutation (p.D1850fs) was identified in seven tumors. Four of these mutations were observed only when visualizing exome data of each sample individually to detect all possible mutations at this locus. There were also one splice-site and six missense mutations. Three of the missense mutations were predicted to be damaging by Ensembl Variant Effect Predictor.

ARID4A was the second most frequently mutated gene with altogether ten mutations in nine tumors (9/46, 20%). Half of these mutations were missense changes, all of which were predicted to be damaging. In addition, we observed one nonsense and three frameshift mutations.

ARID1B mutations were encountered in six tumors (6/46, 13%). These mutations included four missense, one nonsense, and one frameshift change. Three of the four missense mutations were predicted to be damaging. Three tumors harbored a concurrent *ARID1A* and *ARID1B* mutation. *ARID2* was mutated also in six tumors. Here, the mutation spectrum consisted of three missense, one

nonsense, and two frameshift mutations. Two out of the three missense mutations were predicted as damaging.

Finally, we studied whether mutations in these four ARID-encoding genes co-occurred with mutations in known cancer genes *KRAS*, *BRAF*, or *TP53*. Twenty out of 46 tumors (43%) harbored a *BRAF* V600E mutation, two (4.3%) contained a mutation in *KRAS* exon 2 and nine (20%) in *TP53* exons 5-8, Figure 4. There seemed to be no clear correlation between the mutations in any of the four ARID genes and *KRAS*, *BRAF*, or *TP53*. This could, however, be due to the relatively small sample set.

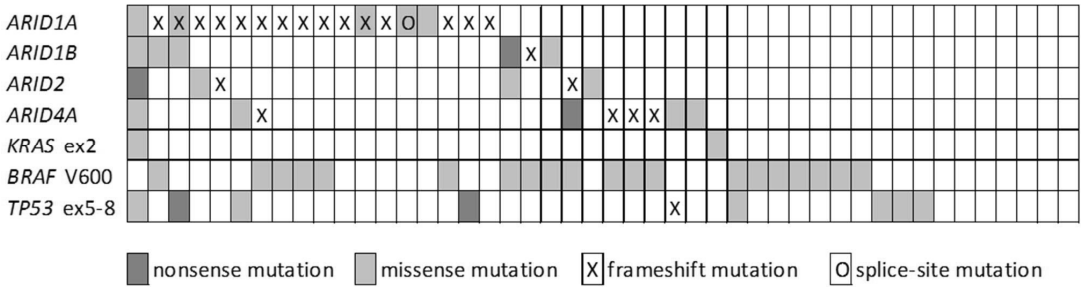

Figure 4. Nonsynonymous somatic mutations detected in the four candidate ARID genes, *KRAS* (exon 2), *BRAF* (V600E), and *TP53* (exons 5-8). Each column represents one tumor and each row a gene. No exclusivity between any of the ARID genes and *KRAS*, *BRAF*, or *TP53* was observed. Copyright (2014) Wiley. Reproduced with permission from Cajuso *et al*, 2014, IJC, Wiley.

2. Novel candidate oncogenes in MSI CRC (II)

2.1. Discovery of potential oncogenes with a mutation hotspot

We searched for novel oncogenes systematically in a set of 25 exome sequenced sporadic MSI CRCs. Our effort focused on identifying recurrent missense mutations hitting the exact same base in at least two tumors. Other than identical base-specific hotspot mutations had been investigated in this set by a previous study[217]. Our search resulted in a list of 42 genes that had possible heterozygous somatic hotspot mutations. Of these, 36 were confirmed real and somatic by Sanger sequencing. This included three known oncogenes *PIK3CA*, *CTNNB1*, and *BRAF* together with 33 candidate driver genes.

2.2. Fourteen genes harbor additional hotspot mutations

The 33 recurrent somatic missense mutations were then screened by Sanger sequencing in a validation set of 254 MSI CRCs. The following 14 candidate genes displayed additional hotspot mutations: *ANTXR1*, *CEP135*, *CRYBB1*, *MORC2*, *SLC36A1*, *GALNT9*, *PI15*, *KRT82*, *CNTF*, *GLDC*, *MBTPS1*, *OR9Q2*, *R3HDM1*, and *TTPAL* (Table 4). Here, all mutations residing in the same codon as the mutation in the discovery set were considered as a hotspot mutation. Altogether 28 (11%) tumors in the validation set, and 40 (14%) tumors in both sets combined, harbored at least one of the novel hotspot mutations. The most prominent gene was *ANTXR1*, harboring a hotspot mutation R438C altogether in seven (2.8%) validation set MSI CRCs. This hotspot was located in the anthrax-binding

domain. However, no expression of the mutant allele of *ANTRX1* was detected in the five tumors available for the analysis. The mutation pattern of *ANTXR1* is displayed in Figure 5. The next frequent hotspots in the validation set were *CEP135* R1115C and *CRYBB1* A171T, both mutated in 1.2% of the samples. The hotspot mutation pattern in *MORC Family CW-Type Zinc Finger 2* (*MORC2*) varied somewhat compared to other genes. In addition to harboring one discovery hotspot mutation (S25L) in the validation set, *MORC2* contained another hotspot nearby, found in one discovery set and one validation set tumor (K27R/M). Known hotspots in oncogenes *BRAF*, *KRAS*, *PIK3CA*, and *CTNNB1* were also screened and they displayed mutations in the validation set with the following frequencies: 42%, 15%, 5.5-11%, and 6.9%, respectively.

Table 4. Fourteen novel candidate oncogenes with validated somatic hotspots. Adapted from Tuupanen *et al*. (2014), BJC, Springer Nature under the Creative Commons Attribution-NonCommercial-Share Alike 3.0 Unported License.

Gene	Description	Hotspot mutation	Frequency Discovery set n=25	Frequency Validation set n=254
ANTXR1	ANTXR Cell Adhesion Molecule 1	R438C	3 (12%)	7/251 (2.8%)
CEP135	Centrosomal Protein 135	R1115C	2 (8%)	3/245 (1.2%)
CRYBB1	Crystallin Beta B1	A171T	2 (8%)	3/248 (1.2%)
MORC2	MORC Family CW-Type Zinc Finger 2	S25L / K27M	3* (12%)	2**/244 (0.8%)
SLC36A1	Solute Carrier Family 36 Member 1	A136T	3 (12%)	2/253 (0.8%)
GALNT9	Polypeptide N-Acetylgalactosaminyltransferase 9	R8C	2*** (8%)	2/238 (0.8%)
PI15	Peptidase Inhibitor 15	A105V	2**** (8%)	2/247 (0.8%)
KRT82	Keratin 82	R496W	2 (8%)	1/252 (0.4%)
CNTF	Ciliary Neurotrophic Factor	T35M	2 (8%)	1/251 (0.4%)
GLDC	Glycine Decarboxylase	R962Q	2 (8%)	1/249 (0.4%)
MBTPS1	Membrane Bound Transcription Factor Peptidase, Site 1	A368T	2 (8%)	1/253 (0.4%)
OR9Q2	Olfactory Receptor Family 9 Subfamily Q Member 2	A249T	2 (8%)	1/251 (0.4%)
R3HDM1	R3H Domain Containing 1	R296Q	2 (8%)	1/251 (0.4%)
TTPAL	Alpha Tocopherol Transfer Protein Like	R147C	2 (8%)	1/252 (0.4%)

*2x S25L, 1x K27M; **1x K27R, 1x S25L; ***1x R8S, 1x R8H; ****1x A105T, 1x A105V

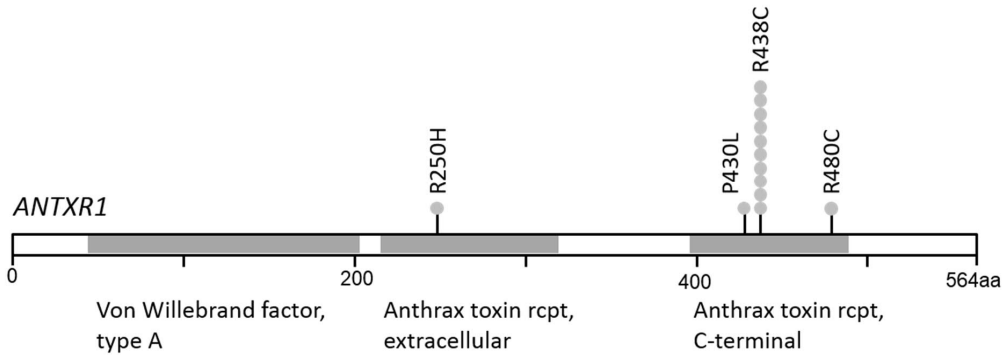

Figure 5. The spectrum of validated *ANTRX1* mutations. The hotspot mutation (R438C) was discovered altogether in ten tumors (three in discovery set and seven in validation set). The other validated mutations observed in this study were R250H and P430L (in MSS CRCs) and R480C (discovery set MSI CRC). rcpt = receptor. Reproduced from Tuupanen *et al*. (2014), BJC, Springer Nature under the Creative Commons Attribution-NonCommercial-Share Alike 3.0 Unported License.

In most of the genes (10/14, 71%), the targeted hotspot amino acids showed high conservation in multispecies protein alignments. These genes were *ANTXR1*, *MORC2*, *CEP135*, *GALNT9*, *PI15*, *KRT82*, *GLDC*, *MBTPS1*, *R3HDM1* and *TTPAL*. The amino acid substitutions in these ten genes together with *SLC36A1* were also predicted to have a damaging or deleterious effect.

Next, the 14 validated novel hotspots were examined in in-house data of ten CRC cell lines and 91 WGS MSS CRCs. In the cell lines, two mutations were found: *MORC2* S25A in HCA7 and *OR9Q2* A249T in DLD1. In MSS CRCs, no hotspot mutations were detected. However, ten of the 14 genes harbored other coding mutations (see Supplementary Table S4 in the original publication). One of them was *ANTXR1* with two missense mutations (R250H and P430L), Figure 5. Since *ANTXR1* was our most prominent hit, these two mutations were also validated and confirmed somatic with Sanger sequencing. However, the mutant allele was again absent in the cDNA level in both cases.

Finally, cBioPortal[218], Catalogue of Somatic Mutations in Cancer (COSMIC)[219], and International Cancer Genome Consortium Data Portal (ICGC)[220] were used to study the 33 hotspot mutations across different tumor types and cell lines. In CRC, we found identical hotspot mutations in five genes: *ANTRX1*, *OR9Q2*, *CTTNBP2*, *HEXIM2*, and *PALD1*. Although not identical, *KCNB1* and *NLRP10* harbored mutations in the same codon as the discovery set hotspot mutations. Also, *MORC2* mutations K27R and K27T were encountered. Two MSI CRC cell lines showed hotspot mutations: HCT-15 in *OR9Q2* and HCT-116 in *HEXIM2*. The search revealed also hotspot mutations in extracolonic cancers in following genes: *ANTXR1*, *CRYBB1*, *PALD1*, *SLC36A1*, *KCNB1*, *SETD9*, *TMEM53*, *PTPRS*, *KRT82*, *PI15*, and *GALNT9* (see Supplementary Table S6 in the original publication). As an example, among 64 MSI stomach cancers, we encountered a hotspot mutation in *ANTXR1*, *CRYBB1*, and *KNCB1*. Although not frequent events, many of the hotspots occurred in a wide spectrum of cancers.

3. Genetic and immune cell characterization of synchronous CRCs (III)

By exome sequencing 23 SCRC pairs with their corresponding normal tissues, we examined the extent of genetic overlap within the pairs to study their similarities and differences and to assess whether they share an origin or have grown independently. In addition, we carried out the first IS analysis of SCRCs to study whether immune cell levels vary between synchronous tumors within a patient.

3.1. Cohort characteristics

The majority (18/23, 78%) of the studied tumor pairs were MSS-MSS. In addition, there were three MSI-MSI pairs and two cases where the MSI status differed within the pair. Most pairs (18/23, 78%) resided on the same side of the colorectum: 13 pairs were left-sided and five right-sided. Besides one patient from a LS family, other cases were assumed sporadic. In addition to the previously known *MLH1* mutation of a LS case, we found no known predisposing factors or clear candidate susceptibility genes when studying the germline of the 23 SCRC cases.

3.2. Lack of genetic overlap in paired tumors

By comparing the somatic mutation contents within the paired tumors, we discovered that most of the mutations were unique, see Figure 6. Out of 23 pairs, only five (22%) shared at least one identical mutation within the paired lesions. Two pairs were MSS-MSS and each pair shared one mutation. The other three pairs were MSI-MSI that shared 12-39 mutations, mainly consisting of frameshift mutations. In total, these shared mutations accounted for 0.4-1.1% of the mutations within the five pairs.

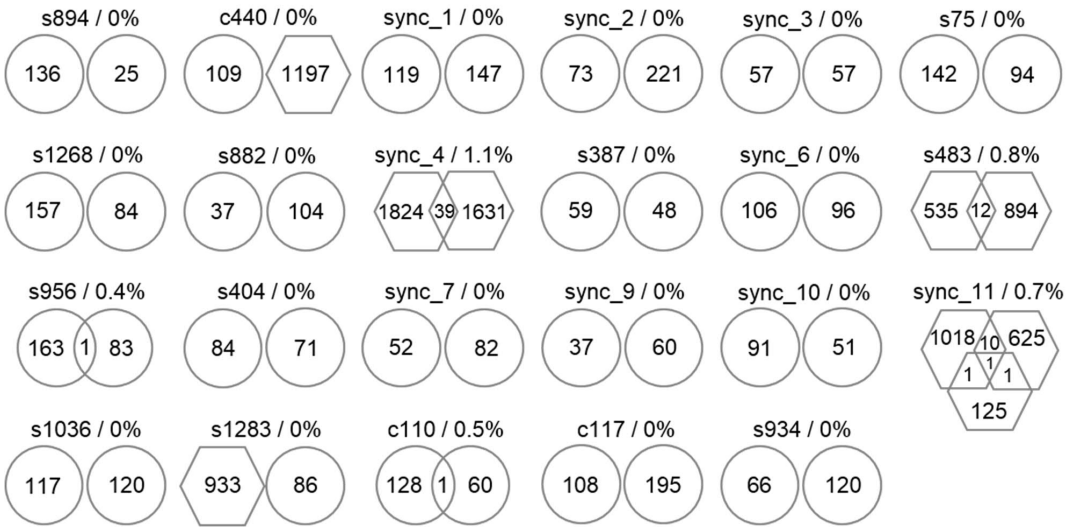

Figure 6. The amount of shared nonsynonymous mutations in SCRC pairs. Nearly all mutations were unique within the tumor pairs. Patient identifiers and the fraction of shared mutations are depicted above each tumor pair. Inside each tumor is marked the number of nonsynonymous mutations they harbor. Hexagon = MSI tumor, circle = MSS tumor. Reproduced from Hänninen *et al*. (2019), BJC, Springer Nature under the Creative Commons Attribution 4.0 International (CC BY 4.0) licence.

Differences were seen also in other aspects, such as the number of nonsynonymous mutations. The median difference was 44 mutations (interquartile range 15.5-71.8) per MSS-MSS tumor pair. Within the three MSI-MSI pairs, the mutation count differed by 193, 359, and 902 nonsynonymous mutations.

The mutation status of known CRC genes varied in many pairs (Figure 7). *KRAS* was mutated in only one of the tumors in six (26%) pairs. Similar discordance in the mutation status was seen in *BRAF*, another clinically relevant gene, that harbored a mutation in only one of the paired tumors in five cases (22%). Such variation was noted also in other genes with potential clinical relevance, such as *PIK3CA* (6/23, 26%) and *PTEN* (3/23, 13%). None of the studied cancer genes harbored similar mutation patterns (mutated in one or both tumors within a pair) across all patients.

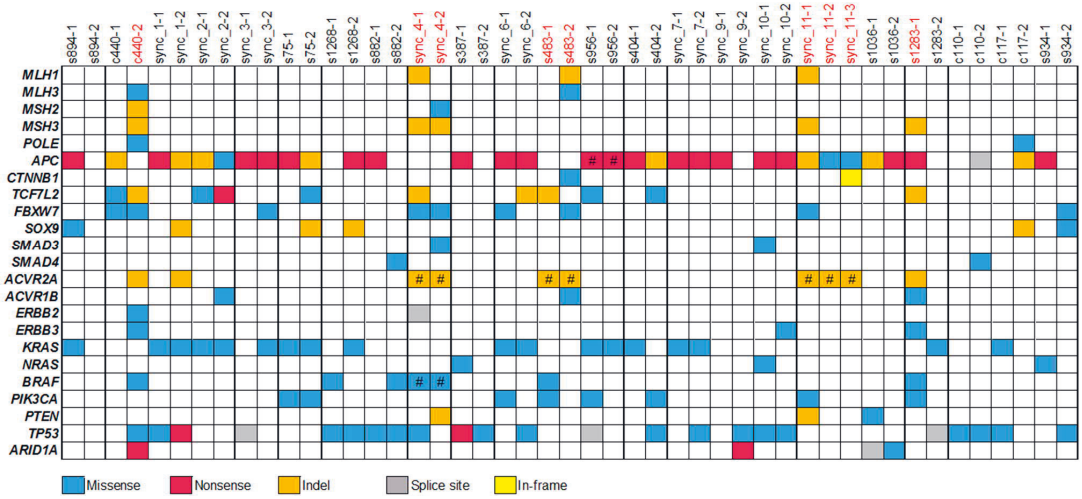

Figure 7. Mutations observed in known cancer genes. Among the most frequently mutated known cancer genes in the whole set were *APC* (33/47, 70%), *TP53* (23/47, 49%), *KRAS* (18/47, 38%), *TCF7L2* (11/47, 23%), and *ACVR2A* (10/47, 21%). MSI tumors are marked with red sample identifier. Hashtag refers to exact same changes within tumor pair. Reprinted from Hänninen *et al.* (2019), BJC, Springer Nature under the Creative Commons Attribution 4.0 International (CC BY 4.0) licence.

We then examined whether the tumors would share other mutated genes, concentrating on genes with a truncating mutation in both tumors of a pair. Excluding the exact same changes and known cancer genes listed in Figure 7, shared genes were found only in four pairs. One MSS-MSI pair had one gene (*FAM133A*) in common, and the three MSI-MSI pairs shared 3-34 genes with truncating mutations, consisting mainly of frameshift mutations. Variation was also detected in the number of overlapping mutated known cancer signaling pathways (ATM, p53, Wnt/β-catenin, ERBB, PTEN, PI3K/AKT, MAPK/ERK, and TGFβ). The number of overlapping altered pathways varied between one and eight in a pair. Six pairs harbored mutations in all eight of the studied pathways in both tumors. The number of mutated genes associated with the pathways, however, varied within these paired tumors.

3.3. Variation in the mutational signatures

Mutational signature analysis revealed four signatures: one specific to the patient, one related to MSI, and two tumor-specific signatures (ts1 and ts2). The patient-specific signature resembled the known age-related signature (signature 1) and the MSI signature the known MSI signature (signature 6)[213]. The tumor-specific signatures did not clearly resemble any of the known signatures. The paired tumors contained similar numbers of mutations contributing to patient-specific signature, indicating that the tumors within a patient were of similar molecular age.

Tumors within pairs showed also variability in their signature contents. Regarding signature fractions, the order of most contributing signatures differed within seven MSS-MSS pairs (7/18, 39%). Tumor location did not seem to explain the differences since the majority (6/7, 86%) of the tumor pairs with different signature order resided on the same side. Within 11 MSS-MSS pairs (61%) the signature order was the same, with individual signature as the largest contributor in nine pairs and ts1 in two pairs.

To compare whether paired tumors resembled each other more than other tumors in the set in general, the proportion of variation of the two tumor-specific signatures within pairs were compared to their variation observed in the whole set. Ts2 was completely independent of the patient and in ts1 patient identity explained only 8% of the variance.

3.4. Immune cell counts vary within paired tumors

Finally, we studied the immune cell counts by assessing IS in 19 SCRC pairs. IS4, the highest of the IS category, was detected most frequently (IS0, 15%; IS1, 10%; IS2, 15%; IS3, 18%; and IS4, 41%). Equal scores were seen within three (16%) tumor pairs. Therefore, the majority of pairs harbored different IS. The score differed by one point in six pairs (32%) and by at least two points in 10 pairs (53%). IS was not significantly correlated within the paired tumors (P = 0.837; Spearman's rank correlation coefficient). Clinicopathological variables, such as mutation count (P = 0.441; GEE with ordinal response), diagnosis age (P = 0.339; GEE with ordinal response), or gender (P = 0.397; GEE with ordinal response) did not seem to explain the differences within the tumor pairs. IS was not associated with mutational signatures either (individual signature, P = 0.288; ts1, P = 0.151; ts2, P = 0.463; GEE with ordinal response). In general, immune response seemed to be stronger in MSI than MSS tumors, although the difference was not statistically significant (P = 0.103; GEE with ordinal response).

4. Exome-wide characterization of small bowel adenocarcinoma (IV)

4.1. Clinicopathological characteristics

106 SBAs were exome sequenced to comprehensively study the coding regions of these tumors for somatic variation. The set included 26 (25%) duodenal, 52 (49%) jejunal, and 18 (17%) ileal tumors, and 10 (9.4%) tumors had an unspecified location. The median age at diagnosis was 62 years. This was lowest for patients with jejunal tumors (59.5 years versus 71 years for duodenal and 63 years for ileal tumors; P = 0.00108, Kruskal-Wallis test). Four of the patients belonged to a LS family, two patients had been diagnosed with FAP, ten with celiac disease, and six with an IBD. Based on the exome data, 91 tumors were designated as MSS and 15 as MSI. MSI was associated with a better disease-specific survival after adjustment for sex, age at diagnosis, and tumor stage (hazard ratio, 0.111; 95% confidence interval (CI), 0.0292–0.419; $P = 1.20 \times 10^{-3}$; Cox proportional hazards model). Half of the tumors from celiac patients (five out of ten) were MSI and therefore celiac disease was associated with MSI (OR, 8.31; 95% CI, 1.62–43.6; $P = 4.83 \times 10^{-3}$; Fisher's exact test), in accordance with previous studies.

4.2. Profile of key mutated genes

The average mutation burden was 4.30 mutations per megabase (mut/Mb) in MSS and 63.6 mut/Mb in MSI tumors. Most frequently mutated known cancer genes in MSS SBAs were *TP53* (44/91, 48%), *KRAS* (43/91, 47%), *APC* (20/91, 22%), *SMAD4* (14/91, 15%), *SOX9* (11/91, 12%), *BRAF* (10/91, 11%), and *ERBB2* (10/91, 11%). Of the known cancer signaling pathways, the most frequently affected pathway in MSS SBAs was PI3K/AKT, having at least one gene mutated in most of the

tumors (77/91, 84.6%). The next most frequently mutated pathways were ERBB (73/91, 80.2%), MAPK/ERK (72/91, 79.1%), and Wnt/β-catenin (70/91, 76.9%). In MSI tumors, the most frequently mutated known cancer genes included *ACVR2A* (13/15, 87%), *BMPR2* (9/15, 60%), *KRAS* (8/15, 53%), and *APC* (7/15, 47%). Of note, *TP53* and *KRAS* mutations in the whole set consisted mainly of missense mutations (38/52, 73% and 50/52, 96%, respectively). Of these, most of the *TP53* and all *KRAS* missense mutations clustered in typical mutation hotspots. Most of the *APC* mutations (37/42, 88%) were protein-truncating. Two out of five (40%) patients with IBD harbored an *APC* nonsense mutation.

Exome data allowed us to detect genes that showed statistical evidence of positive selection for mutations in SBA. OncodriveFML identified 44 nominally significantly mutated genes ($P < 0.05$). The following seven genes remained above the significance level after multiple testing correction (false discovery rate, q-value < 0.1): *TP53*, *KRAS*, *APC*, *SOX9*, *SMAD4*, *BRAF*, and *ACVR2A*. All genes with $P < 0.05$ were, however, considered potentially interesting. The list of 25 highest-ranking genes is displayed in Table 5. In addition to the known cancer genes in SBA, the highest-ranking driver candidates included recently reported (*ATM* and *ARID2*) and novel candidate driver genes (e.g. *ACVR1B*, *BRCA2*, and *SMARCA4*).

Table 5. The 25 highest-ranking genes in MSS SBAs. OncodriveFML denoted the following genes as the most relevant genes in the 91 MSS SBAs. Of these, *TP53*, *KRAS*, *APC*, *SOX9*, *SMAD4*, *BRAF*, and *ACVR2A* remained significant also after correction for multiple testing. Reprinted under the Creative Commons Attribution (CC BY) license from Hänninen *et al*. (2018), PLoS Genet, PLoS.

Gene	Description	*P*-value	No. of samples	No. of muts	SNP	INDELS
TP53	Tumor Protein P53	0.000001	42	43	38	5
KRAS	KRAS Proto-Oncogene	0.000001	43	43	42	1
APC	APC Regulator Of WNT Signaling Pathway	0.000001	21	29	21	8
SOX9	SRY-Box Transcription Factor 9	0.000001	11	12	4	8
SMAD4	SMAD Family Member 4	0.000034	14	17	15	2
BRAF	B-Raf Proto-Oncogene	0.000467	10	10	10	0
ACVR2A	Activin A Receptor Type 2A	0.00051	6	7	3	4
PIK3CA	Phosphatidylinositol-4,5-Bisphosphate 3-Kinase Catalytic Subunit Alpha	0.002106	7	7	7	0
PDYN	Prodynorphin	0.002624	4	4	3	1
ATM	ATM Serine/Threonine Kinase	0.003448	7	8	6	2
ARID2	AT-Rich Interaction Domain 2	0.005667	5	5	3	2
FBXW7	F-Box And WD Repeat Domain Containing 7	0.007291	5	5	4	1
ACVR1B	Activin A Receptor Type 1B	0.007554	5	5	5	0
BRCA2	BRCA2 DNA Repair Associated	0.007599	4	4	3	1
FAM214A	Family With Sequence Similarity 214 Member A	0.008006	4	4	3	1
TRO	Trophinin	0.010371	4	4	4	0
DNMBP	Dynamin Binding Protein	0.012701	4	4	3	1
ZSCAN1	Zinc Finger And SCAN Domain Containing 1	0.014597	4	4	4	0
ADAMTS18	ADAM Metallopeptidase With Thrombospondin Type 1 Motif 18	0.015212	5	5	4	1
MIB2	Mindbomb E3 Ubiquitin Protein Ligase 2	0.018198	6	7	7	0
FAM193A	Family With Sequence Similarity 193 Member A	0.01966	4	4	3	1
PSME4	Proteasome Activator Subunit 4	0.021745	4	4	4	0
OR5M3	Olfactory Receptor Family 5 Subfamily M Member 3	0.022966	4	4	4	0
FRAS1	Fraser Extracellular Matrix Complex Subunit 1	0.024063	4	5	4	1
SMARCA4	SWI/SNF Related, Matrix Associated, Actin Dependent Regulator Of Chromatin, Subfamily A, Member 4	0.025577	7	7	6	1

4.3. Atypical BRAF mutation pattern

Ten MSS and one MSI tumor (11/106, 10.4%) harbored a mutation in *BRAF*, see Figure 8. However, none of the mutations were V600E, the most common *BRAF* hotspot mutation in CRC. Instead, the mutation pattern included two other known but less studied hotspots: G469A found in two tumors and D594A/G/N in three tumors. Additionally, there were mutations located nearby (G466E, G596R, and K601N). All of the above-mentioned mutations were located in exons 11 or 15 and have been determined as somatic mutation hotspots in many cancers[221]. Read-level inspection uncovered one additional tumor that displayed a hotspot mutation in G469A which had not been called. This tumor also harbored another *BRAF* missense mutation. *BRAF* mutations included also two protein-truncating mutations, Q257X and A404fs. Generally, *BRAF* V600E and *KRAS* mutations are mutually exclusive. Here, four out of 11 *BRAF* mutant tumors harbored also a *KRAS* hotspot mutation. Comparison of the tumor and patient characteristics between *BRAF* mutant and *BRAF* wildtype cases did not display any significant differences.

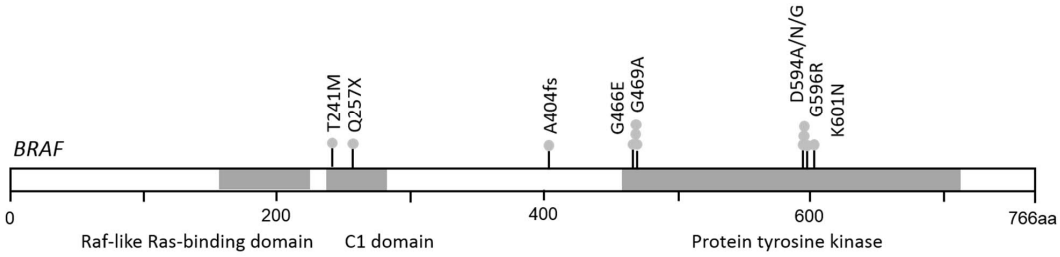

Figure 8. Mutation observed in *BRAF*. We identified altogether 12 *BRAF* (ENST00000288602) mutations in 11 tumors (10 in MSS and one in MSI). Reprinted under the Creative Commons Attribution (CC BY) license from Hänninen *et al*. (2018), PLoS Genet, PLoS.

4.4. ERBB receptor genes frequently affected

The most frequently mutated member of the ERBB receptor gene family was *ERBB2*, where mutations were detected in 15 tumors (15/106, 14%), see Figure 9. Of these, ten were MSS and five MSI. *ERBB2* did not reach significance level in the OncodriveFML analysis but it is frequently mutated in GI tumors, including SBA, and is a known therapy target[45,115,152,221,222]. Most of the identified mutations (14/18, 78%) clustered into four known hotspots, S310F/Y, R678Q, L755S and V842I. *ERBB2* mutations seemed to be more frequent in MSI than MSS tumors (OR, 3.98; 95% CI, 0.886–16.4; P = 0.0368; Fisher's exact test). Although with lower frequency, also the other three members of the ERBB family were mutated in the set: *EGFR* in one tumor, and *ERBB3* and *ERBB4* in nine tumors. *ERBB3* had two hotspot mutations - V104M/L found in three tumors and S864I in two tumors. *ERBB4* included one mutation hotspot, L897R/P, found in two tumors. Altogether 29 SBAs (27%) harbored a mutation in at least one of the ERBB genes. Hotspot mutations in different ERBB genes were mutually exclusive, Figure 9.

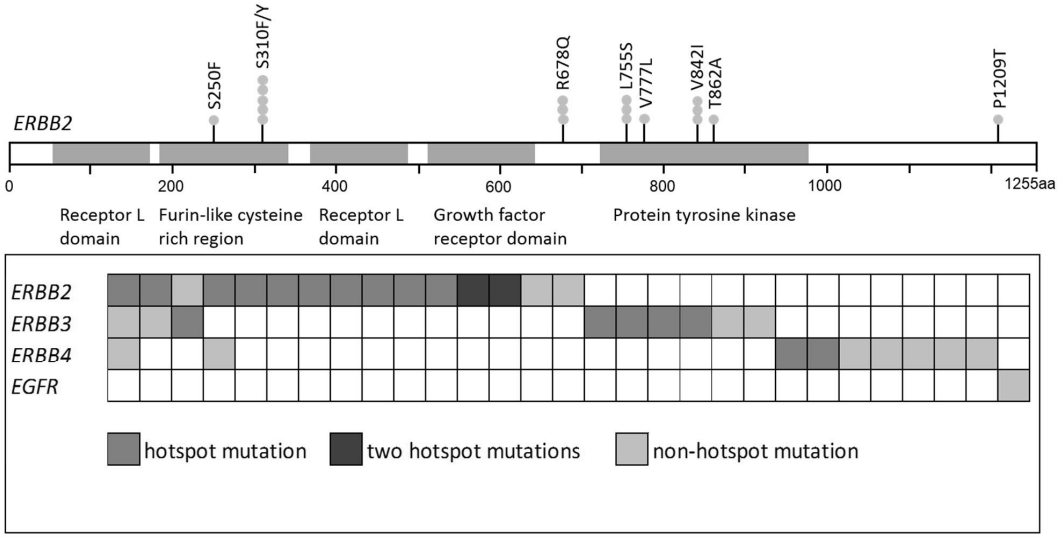

Figure 9. Mutation pattern in *ERBB2* and other ERBB receptor family genes. Mutations in *ERBB2* (ENST00000269571) grouped into four mutation hotspots (above). Below, SBAs (n=29) displaying a mutation in at least one member of ERBB receptor family genes are depicted in columns. All mutations are not shown as some of the samples with a hotspot mutation harbored also another mutation in the same gene. Reprinted under the Creative Commons Attribution (CC BY) license from Hänninen *et al*. (2018), PLoS Genet, PLoS.

4.5. Allelic imbalance events

AI analysis revealed a total of 840 gain and 1,541 loss events in the whole data set. MSI tumors harbored a significantly lower number of events than MSS tumors ($P = 1.95 \times 10^{-9}$; 95% CI, 0.16-0.39; negative binomial model). The most frequent AI event was partial or whole loss of chromosome 17p which was detected in 62/106 (58.5%) samples. This region includes *TP53* and non-synonymous *TP53* mutations co-occurred with the loss events in 41/50 (82.0%) of the mutated samples (OR, 7.43; 95% CI, 2.86-21.1, $P = 4.02 \times 10^{-6}$, Fisher's exact test). Also, *SOX9* (n = 44) and *SMAD4* (n = 46),

other significantly mutated known cancer genes, had a high frequency of chromosomal losses. Losses of chromosome or chromosomal arm were frequent events (n > 30) at the loci 3p, 8p, 9q, 12q, 15, 17, 18q, 19, and 22. Whereas, chromosomes 13 and 8q harbored a high frequency of gain events. Clear amplification was also seen in known cancer genes, such as *KRAS* (20/106, 18.9%), *BRAF* (19/106, 17.9%), and *PIK3CA* (16/106, 15.1%). Four samples showed a local amplification at the *ERBB2* locus, two of which also harbored a concurrent *ERBB2* hotspot mutation.

4.6. Mutational signatures

The mutational signature analysis was performed for all 106 SBAs. Signature 6, the known MSI signature, was identified in the 15 tumors designated as MSI[63]. These tumors were grouped separately, and the analysis was then repeated to the 91 MSS samples. This resulted in three mutational signatures that corresponded to known signatures 1A, 17, and U2. We detected an association between higher amount of mutations contributing to signature 1A and older age at diagnosis (increase per 10 years, 20%; 95% CI, 10–32%; $P = 4.32 \times 10^{-5}$; negative binomial model), similar phenomenon as seen in other cancers[63].

4.7. Variation between tumors from different small bowel segments

Finally, we compared tumors from the three small bowel segments. The three tumor groups contained rather similar mutation and AI counts, MSI frequencies, and mutated known cancer genes. However, they also displayed differences. In MSS SBAs, *APC* mutation frequency was lower in jejunal tumors (13.6%) compared to ileal (31.3%) and duodenal (37.5%) tumors. Whereas, *TP53* mutation frequency was lowest in duodenum (29.2%) followed by jejunal (56.8%) and ileal (56.3%) tumors.

Jejunal tumors showed the highest exposure to signature 1A. Compared to duodenal tumors, there was a 66% increase of expected mutation count (95% CI, 29%-110%; $P = 7.17 \times 10^{-5}$; negative binomial model). We observed no clear difference regarding signature 1A between ileal and duodenal tumors ($P = 0.348$; negative binomial model).

Of the known cancer signaling pathways, ERBB signaling was the most frequently mutated pathway in duodenal tumors (20/24, 83.3%). In jejunal and ileal tumors PI3K/AKT was the most frequently affected pathway (42/44, 95.5% and 12/16, 75.0%, respectively). Overall, ERBB and MAPK/ERK signaling displayed most notable differences between segments: ERBB signaling seemed to be less frequently affected in ileal tumors (9/16, 56.3%) than those of duodenum (20/24, 83.3%) and jejunum (38/44, 86.4%), ($P = 0.0463$, Fisher's exact test). Whereas, MAPK/ERK signaling was more frequently mutated in jejunal tumors (40/44, 90.9%) than duodenal (18/24, 75.0%) or ileal tumors (9/16, 56.3%) ($P = 9.06 \times 10^{-3}$, Fisher's exact test).

DISCUSSION

1. *ARID* gene mutations in CRCs with microsatellite instability (I)

Molecular characterization of different tumor types has led to the identification of many driver genes. Among them is *ARID1A* that has been recognized as a tumor suppressor gene in a variety of cancers. *ARID1A* is part of the ARID domain containing gene family that consists of 15 genes. They all contain a highly conserved ARID domain, bind to DNA either in a non-specific manner or specifically at AT-rich sites, and are involved in cell proliferation control, tissue-specific expression, and development[38,223]. Prior to our study, there had not been a systematic effort to study the mutation profile of all ARID-encoding genes in any cancer type. Our aim was to investigate the mutation spectrum and the role of ARID genes in MSI CRC.

Mining the exome data of 25 MSI CRCs, we found mutations in 12 out of the 15 ARID genes. Since the high background mutation rate in MSI tumors makes the data analysis challenging we applied a selection criterion of at least four mutated samples (including one nonsense or splice site mutation) for genes to qualify for further analyses. Four genes, *ARID1A*, *ARID1B*, *ARID2*, and *ARID4A*, fulfilled this criterion and emerged as the most potential candidates for additional studies.

ARID1A, *ARID1B*, and *ARID2* encode for components of the large SWI/SNF chromatin remodeling complex. Crucial function of the complex is to remodel nucleosomes and modulate transcription[224]. Since ARID domain allows the complex to bind DNA, mutations in these genes might disrupt the capability of the complex to bind DNA and regulate gene expression[223]. The SWI/SNF complex is affected commonly in cancer as mutations, deletions, and translocations in different subunits have been reported in around 20% of human tumors[49].

ARID1A encodes for BAF250a, a key component of the SWI/SNF chromatin remodeling complex[225]. *ARID1A* has emerged as a tumor suppressor with frequent inactivating mutations identified in e.g. ovarian, gastric, and pancreatic cancer[226–229]. *ARID1A* was the most frequently mutated of the ARID genes in our study displaying somatic mutations in 39% of tumors. The majority (63%) of the nonsynonymous mutations were frameshift changes, typical of MSI tumors. Mutations occurred along the length of the gene; a mutation pattern consistent with the reported spectra in other cancers. One exception was a deletion of one base pair in a mononucleotide repeat of seven G's in seven samples. This represents a typical MSI target gene mutation and indeed this hotspot mutation has been observed e.g. in MSI gastric cancer[228]. Some studies have reported a correlation between *ARID1A* and *TP53* mutations which varies, however, from an inverse relationship to a mutual inclusivity[49,230]. We detected no correlation between these genes. However, due to small sample size, no comprehensive conclusions could be made.

The driver role of ARID1A has been further affirmed by more recent studies, such as WGS efforts on pancreatic and gastric cancers[231,232]. Although *ARID1A* harbors mutations in diverse tumor types, these mutations occur often in a particular molecular or histological subtype, such as mutations enriched in gastric cancer and CRC displaying MSI[45,228] and ovarian clear-cell and endometrioid carcinomas[227]. A study on mice showed that loss of ARID1A results in invasive colon tumors exhibiting features associated specifically with MSI CRC[233]. They also observed that in the context

of *Apc*-mutant mouse, ARID1A inactivation seemed to block tumor formation suggesting also a possible oncogenic role for *ARID1A* in certain contexts. *ARID1A* mutations have also been suggested to have a prognostic role as loss of ARID1A seemed to shorten time to cancer recurrence and cancer-specific mortality[234].

ARID1B encodes for BAF250b, a member of BAF variant SWI/SNF complexes and mutually exclusive with BAF250a[225,235,236]. In our set, 13% of tumors harbored a mutation in *ARID1B*. Contrary to *ARID1A*, the majority of the mutations were one base pair substitutions that resulted in one nonsense and four missense changes. *ARID1B* mutations have been detected in different tumor types, such as breast, liver, and clear cell ovarian cancers[237–239]. In addition, *ARID1B* deletions have been observed e.g. in pancreatic cancer and childhood neuroblastoma[229,240]. *ARID1B* has been suggested to repress Wnt/β signaling and thus inactivation of *ARID1B* could lead to elevated Wnt-dependent transcription of target genes[241].

Three out of six *ARID1B* mutated tumors in our study harbored also a mutation in *ARID1A*. *ARID1A* and *ARID1B* mutations have been reported to co-occur and loss of *ARID1A* and *ARID1B* alleles have been suggested to cooperatively promote tumorigenesis[242]. However, this study showed *ARID1A*-deficient cancers to retain at least one functional *ARID1B* allele, suggesting a specific vulnerability compared to the non-mutated cells. This synthetic lethal relationship has been observed also in other studies and, in theory, could be used against *ARID1A*-deficient tumors by targeting ARID1B[243]. The approach has, however, many challenges as ARID1B does not contain small molecule binding sites and targeting ARID1B could increase cancer risk[48]. Also PARP inhibitors and anti-PD-L1 antibodies have been suggested as potential therapeutic approaches for *ARID1A*-deficient tumors[36,244]. In general, since many of the *ARID1A* mutations are inactivating and leading to loss of protein expression, *ARID1A* presents a difficult therapy target.

ARID2 encodes a part of the pPAF complex, another member of the SWI/SNF chromatin remodeling complex. The role of pPAF is to facilitate ligand-dependent transcriptional activation[245]. *ARID2* was first discovered as a putative tumor suppressor in hepatocellular cancer[246]. *ARID2* mutations have been reported in a variety of cancers, such as non-small cell lung adenocarcinoma, melanoma, and pancreatic cancer[247–249]. In our sample set, *ARID2* harbored a mutation in 13% of cancers. One nonsense mutation located in the ARID domain and thus could potentially disrupt its DNA binding ability. Subsequent to our study, *ARID2* has been suggested as a potential driver also in duodenal adenocarcinomas, papillary renal cell cancer, as well as hepatocellular cancer metastasis[153,250,251]. Albeit at a low frequency, *ARID2* was also significantly mutated in a pan-cancer analysis[60].

ARID4A is a ubiquitously expressed nuclear protein. Together with other proteins it binds directly to the retinoblastoma protein (pRB), a regulator of cell proliferation, and induces growth arrest[252]. *ARID4A* was the second most frequently mutated gene (20%) of the four candidate genes. One of the mutations located in the ARID domain. Prior to our study, *ARID4A* had been identified as a potential tumor suppressor in leukemia and decreased expression of *ARID4A* had been detected in laryngeal and oral tongue carcinomas[253,254]. More recently, studies have suggested *ARID4A* to also play a role in gastric and prostate cancer[255,256].

In addition to *ARID1A*, also other members, *ARID1B*, *ARID2*, and *ARID4A*, were rather frequently mutated in our set. Various types of mutations were present and scattered along the four genes, a mutation pattern typical of tumor suppressors and what is previously reported in *ARID1A*. As expected, little over one third of the mutations (38%) were frameshift changes. The mutation spectrum we observed indicates that rather than affecting specific functions of the protein, mutations lead more likely to partial or complete inactivation of the gene. These four genes appeared to be relatively frequently mutated also in two large genome sequencing studies on CRC that included 23 and 15 MSI CRCs[45,114]. Mutations in all four candidate genes were also found in MSS tumors of these studies, although at lower frequencies. Overall, the presence of typical tumor suppressor mutation patterns in genes previously associated with diverse cancer types suggests these genes may play a role also in MSI CRC.

To our knowledge, this was the first systematic effort to study the mutation spectrum of the ARID gene family. However, due to rather small sample size, this should be considered as a pilot study. The possibility that some mutations are passenger events resulting from the MSI phenotype cannot be excluded either. As another limitation to the study, larger structural alterations as well as non-coding regions remain uncharacterized. Future studies are still needed to validate the function of these mutations as well as better define the role of ARID genes in tumorigenesis.

2. New candidate oncogenes with mutation hotspots (II)

Most frequently mutated drivers have been discovered, however, the catalogue of mutations occurring at intermediate or low frequencies is still incomplete[45]. To date, there are only a few established oncogenes in MSI CRC. These include *BRAF*, *PIK3CA*, *CTNNB1*, and *KRAS*[100]. Oncogenes usually present recurrent mutations in specific mutation hotspots and it has been stated that pattern, rather than mutation frequency, should be considered when looking for significantly mutated genes[4]. In MSI tumors, the challenge is to distinguish true drivers among the passenger mutation load. These tumors, however, may provide a sensitive model system to detect novel oncogenic mutations by generating frequent mutations for possible selection during tumor progression. Since mutation hotspots are thought to be markers of positive selection in tumorigenesis, we hypothesized that by searching for these very specific mutation patterns we could detect new candidate oncogenes.

In this study, we sought to identify novel oncogenic driver genes by using exome data on 25 MSI CRCs and the corresponding normal samples as the discovery set. We identified 33 novel candidate oncogenes recurrently mutated at the same positions. Fourteen genes (*ANTXR1*, *CEP135*, *CRYBB1*, *MORC2*, *SLC36A1*, *GALNT9*, *PI15*, *KRT82*, *CNTF*, *GLDC*, *MBTPS1*, *OR9Q2*, *R3HDM1*, and *TTPAL*) harbored hotspot mutations also in the validation set. Of these, *ANTXR1* stood out as the most frequent target of hotspot mutations (2.8%). Two validated hotspots (*MORC2* S25A and *OR9Q2* A249T) were also found in one of the 10 studied CRC cell lines. In addition to CRC, a number of the 33 novel hotspot mutations were encountered in various extracolonic cancers. Albeit with low frequency, these hotspot mutations could have a role in the genesis in numerous cancers.

ANTXR1 (also named *TEM8*) encodes for a single-pass cell-surface protein that was first identified to be overexpressed on the tumor vasculature in human CRC and has been shown to play a role in

tumor angiogenesis[257,258]. TEM8 antibodies were also shown to exhibit antitumor activity in several different tumors in mice[258]. Subsequently, ANTXR1 has been shown to be overexpressed also on cancer cells in a variety of tumors, such as breast, neuroblastoma, and melanoma[259–261]. ANTXR1 has also been suggested to function as a biomarker of pancreatic ductal adenocarcinoma and triple-negative breast cancer stem cells[262,263]. Recently, studies have indicated overexpression of ANTXR1 to be a potential target for immunotherapy in triple-negative breast cancer and gastric cancer[264,265]. The biological function of *ANTXR1* combined with the observed mutation pattern makes *ANTXR1* the most potential candidate oncogenic driver in MSI CRC. Even though we found no expression of the mutant allele in the mature cancers, which makes the hotspot in question not an attractive therapy target per se, the exact mechanisms of the involved tumorigenesis are intriguing for further cancer biology research.

Also other genes highlighted by our study have been studied further in regards to their involvement in cancer. *MORC2* has been reported to facilitate ATP-dependent chromatin remodeling in response to DNA damage and to be upregulated in many cancers[266,267]. Studies have also suggested *MORC2* to promote genesis of e.g. gastric, breast, and liver cancers and contribute to the metastasis development in triple negative breast cancer[268–271]. Another example, *Glycine Decarboxylase* (*GLDC*), has been reported to be upregulated in many cancer types and one preclinical study suggested *GLDC* as a potential therapeutic target in non-small cell lung carcinoma[272]. Suppression of *GLDC* has also been indicated to suppress tumor cell growth in various cancer cell lines[273]. Whereas, expression of *GLDC* has been shown to associate with melanoma prognosis[274]. Also *Alpha Tocopherol Transfer Protein Like* (*TTPAL*) has recently been suggested to promote colorectal carcinogenesis[275]. They also associated TTPAL overexpression with worse survival.

Considering the strict criteria used in this study to define a hotspot mutation, the seemingly low mutation frequencies are not that surprising. As a comparison, *CTNNB1* and *KRAS* harbored hotspot mutations with the frequency of only 8% and 4% respectively in the discovery set (see Table 6). Subsequently, we have further characterized the SNV mutations in the discovery exome data set together with additional 12 genome-sequenced MSI CRCs[276]. The ranking of the top 72 most significantly mutated genes included three of our novel candidate oncogenes, *CMTM2*, *CRYBB1*, and *SLC36A1*. The effort also entailed the identification of additional potential oncogenes with a hotspot mutation. This yielded another seven novel candidates.

Table 6. Hotspot mutation frequencies of known oncogenes in the study and examples of the reported mutation frequencies of these genes in MSI CRC. Of note, the frequencies are not fully comparable due to varying sequencing targets: Mirabelli-Primdahl *et al.*[277] and Rajagopalan *et al.*[278] included most frequently mutated exon(s), Corso *et al.*[279] a mutation hotspot or two of the most frequently mutated exons, Lin *et al.*[141] a mutation hotspot panel, and Jonchere *et al.*[280] all exons of the genes.

Gene	Discovery set n=25	Validation set n=254	(Mirabelli-Primdahl *et al.* 1999) n=53	(Rajagopalan *et al.* 2002) n=49	(Corso *et al.* 2013) n=38	(Lin *et al.* 2015) n=29	(Jonchere *et al.* 2018) n=47
BRAF	32%	42%	NA	31%	16%	34%	53%
PIK3CA	20%	17%	NA	NA	7.9%	34%	38%
KRAS	4%	15%	NA	43%	21%	31%	21%
CTNNB1	8%	6.9%	25%	NA	NA	3%	NA

As to limitations of this study, it was conducted with a relatively small sample set and thus should be considered as a pilot study. We identified a set of candidate driver oncogenes, however, with this targeted approach other potential oncogenes activated by alternative ways remain uncharacterized. Furthermore, tumor heterogeneity and microenvironment are known to create challenges in mutation discovery. Further work on the characterization of the identified hotspot mutations in different tumor types together with functional experiments are still required to elucidate their part in tumorigenesis.

For improving diagnostics, new therapeutic interventions, and personalized care, it is essential to gather comprehensive knowledge on the spectrum of mutations in human cancers. In general, mutation hotspots in driver oncogenes create attractive therapeutic targets for small-molecule inhibitor drug development. For example, *BRAF* V600E mutations have been successfully utilized as a treatment target for patients with metastasized melanoma[281]. Often, however, targeted monotherapies fail to cure the patient due to pre-existing subclones that are resistant to the treatment[282]. Most cancer driver genes mutated in most patients occur at intermediate-frequency level (2-20%)[60]. Thus, it is of utmost importance to characterize and understand the role of these less frequently affected genes in tumorigenesis as well in order to provide therapeutic options for a larger number of patients and identify targets for combination treatment protocols. Here, MSI CRCs were used as a platform to find novel candidate oncogenes. Some of the hotspot mutations were also present in a small percentage of other tumors, implying they could play a role in a variety of tumor types. The catalogue of these mutation hotspots should serve as a valuable resource for further cancer research.

3. Synchronous CRCs within a patient display substantial variety (III)

Although SCRCs account for only a small fraction of CRC cases, given the prevalent nature of CRC, SCRCs represent a fairly frequent clinical challenge. Since the paired tumors arise in a common background of genetic and environmental factors, SCRCs present a unique model to study CRC genesis. Studies on SCRC have previously focused mainly on clinicopathological features and single genetic factors, such as the MSI status and known cancer genes including *APC*, *KRAS*, and *BRAF*. The underlying molecular mechanisms in SCRC have not been thoroughly investigated. We were therefore compelled to study these tumors in a more comprehensive manner. We exome-sequenced 23 tumor pairs and their corresponding normal tissue to evaluate the extent of their genetic overlap. Additionally, we analyzed the interplay between the mutation profile and immune cell invasion in the paired tumors. Based on clinical information, besides one case with Lynch syndrome, patients did not harbor any clear high risk factors for CRC.

First, we aimed to understand whether these simultaneous tumors share a common origin, or have they arisen independently. Based on the exome data, only five out of 23 pairs displayed one or more exact same genetic changes within them. Of the five pairs, three were composed of MSI-MSI tumors where the shared mutations accounted for only a small fraction of the mutations and represented types of changes commonly found in MSI tumors. The other two pairs were MSS-MSS and both pairs shared one mutation. One of these was a known mutation hotspot in *APC*, shared probably by coincidence. The lack of genetic overlap strongly suggests these paired tumors have independent origins.

While our study was ongoing, three other studies reported features of SCRCs based on exome-sequencing data[283–285]. In general, the median mutation counts per tumor (88.5 for MSS and 933 for MSI) were in line with these studies and what has been reported for solitary CRC[45,283–285]. Our results were in agreement with these three other exome studies that, in general, supported the idea of separate genetic origins. Whereas, one SNP study reported a subset of SCRCs to be monoclonal[286]. Single studies on synchronous tumors of other anatomical sites, such as lung and kidney, have shown these tumors to also display genetic heterogeneity[287,288].

Second, we wanted to explore whether the paired tumors would otherwise harbor similarities. We observed differences in the mutation statuses of the known CRC genes within pairs which is in line with previous results[283–285,289]. Further differences were seen when searching for other genes with possibly damaging truncating mutations in both tumors of a pair. This resulted mainly in genes with frameshift mutations in repeated sequences of MSI-MSI tumor pairs. These mutations had probably arisen by chance due to the replication errors characteristic of MSI tumors. The data suggested that the paired tumors might also utilize different signaling pathways. In general, genetic diversity was seen in all tumor pairs, regardless of a known predisposition syndrome background, the presence and number of adenomas, location of the tumors, or their distance from each other. In addition, the search for possible predisposing genetic factors did not result in any clear candidate genes.

We performed mutation signature analysis to evaluate whether the paired tumors in the same environment have developed through similar mutational processes. When looking at signature ts1, the paired tumors did resemble each other somewhat more than other tumors in the set. Whereas, signature ts2 did not differentiate whether the tumors belonged to the same patient. In general, the MSS tumors harbored similar signature contents both within the pairs and between patients. However, within over one third of the pairs the composition of mutation signatures differed. Therefore, these tumors might have undergone different mutational processes.

To further characterize the similarities and differences within tumor pairs, we assessed the immune infiltrate in the tumors by IS in 19 pairs. The combination of immune score and TNM classification and/or defined MMR status in CRC has been proposed to predict survival and therapy response[88,89,216,290]. There was no previous knowledge on how immune responses and thus immune cell levels vary within paired lesions in a patient. We observed that the intratumor immune reaction varied between cases and, in the majority of them, also within the tumor pair. Clinicopathologic features, such as mutation count, MSI status, or tumor location, did not clearly explain this variation. The immune reaction is thought to be affected by several factors, such as the tumor microenvironment, both genetic background of the patient and the genetic content of the tumor, and gut microbiota[89].

Studies have reported synchronous tumors to differ from solitary tumors. When compared to solitary tumors, the fraction of MSI tumors among SCRCs was slightly higher (22% vs. ~15%) which is in agreement with previous studies[24,81,83]. Some studies have reported mutation rate of known CRC genes, such as *BRAF*, *SMAD4*, *PIK3CA*, and *NRAS*, to vary in SCRCs compared to solitary tumors. We did not detect any clear differences.

The observed variation within paired tumors might affect genetic testing and therapeutic strategies. The results strengthen the concept that, for optimal outcome, both tumors should be considered when screening for drug targets or predictors of therapy response, such as *KRAS*, *NRAS*, *BRAF*, and *PIK3CA*. Many tumor pairs harbored a mutant *KRAS* in only one of the tumors - a situation where only one of the tumors might respond to anti-EGFR treatment. Although it has been suggested that, within a patient, a large majority of driver gene mutations are common to all metastatic lesions and thus a biopsy from a single metastasis represents accurately most of the functionally important mutations, synchronous primary tumors might present a more convoluted scenario as the metastases might have arisen from either of the genetically distinct tumors[291]. Some patients in our study displayed a pair of MSI-MSS tumors, indicating that the neoantigen levels might vary between tumors. This might have relevance if immune checkpoint inhibitor therapy is considered as a potential treatment. Incorporating IS as a measure of immune response might provide additional information on tumors that could respond to immune modulating therapy.

The study had some limitations. The research material included DNA extracted from FFPE material. Although steps were taken to ensure the efficient removal of possible false calls, we recognize that the data might still contain some low-frequency artefacts. However, these should not affect the main conclusions of the study as the aim was to compare paired tumors exome-wide and the mutations of the main interest were of good quality. Despite exome sequencing being highly informative, we were not able to study the non-coding genomic regions or epigenetic changes, such as gene methylation. Thus, it remains unknown what is their contribution to tumor multiplicity. Field effect has been one factor theorized to have an impact on tumorigenesis of synchronous tumors. In many cases the tumors were found on the same side of the colorectum and several patients had at least one additional adenoma. These might imply possible regional field effect. Some studies consider synchronicity as tumors that have appeared within a 6-month period. Among the strengths of this study, we concentrated solely on synchronous tumors that have been diagnosed simultaneously. Additionally, by analyzing tumors exome-wide, instead of concentrating on targeted genes, we were able to obtain a more comprehensive view of their genomic diversity.

In summary, we observed different mutation patterns with only a few shared mutations within paired tumors. This proposed a parallel evolution of SCRC without a common origin and thus confirmed previous observations. The mutation frequencies of known CRC genes resembled those seen in solitary CRCs, while some differences exist between previous studies on SCRC. Our study elucidated further the genetic background of SCRC and how immune responses varied within paired lesions. Regardless of their shared environment, CRCs that are found simultaneously within the patient exhibit great genetic variety and thus it is essential to examine both tumors when designing treatment. Additional studies are still needed to further explore the causes for tumor multiplicity.

4. Somatic mutation landscape of SBA (IV)

Our understanding of cancer genetics on multiple tumor types has rapidly increased with the utilization of NGS tools. Yet, due to their rarity, some cancer types remain less studied. One of such is SBA, an aggressive disease with limited treatment options. Prior to our study, relatively few large-scale studies on SBA had been conducted. These have either screened a set of known mutation

hotspots or cancer genes or exome-sequenced small sets of duodenal adenocarcinomas[115,120,152–154]. Since its incidence is rising and no evidence-based treatment recommendations have been available, there is a dire need for more knowledge on SBA biology. Thus, we conducted the thus far largest exome sequencing effort on a nationwide population-based SBA sample collection in order to characterize the common genetic changes occurring in this tumor type. This unique cohort included altogether 106 primary SBAs including tumors from all three small bowel segments.

The median mutation burden in the whole set was 3.96 mut/Mb which is in agreement with previous studies on SBA and similar to that seen in gastric cancer and CRC[67,115]. The fraction of MSI tumors were also similar in these tumor types[45,222]. We confirmed in a large set of SBAs that, as in CRC, MSI correlates with longer disease-specific survival also in SBA[292]. This was detected also in a more recent study[293]. Patients with MSI tumor might benefit from immunotherapy, suggesting a potential role for immunotherapy also in a fraction of SBA patients[92].

Exome data enabled us to search for genes that showed statistical evidence for positive selection in SBA. The most significantly altered genes in MSS SBA included *TP53*, *KRAS*, and *APC*. These genes were also frequently mutated in MSI SBA. AI analysis provided further support for the pivotal roles of these genes in SBA genesis, as high frequency of losses in *TP53* and *APC* and gains in *KRAS* was observed. The mutation frequencies of *TP53* and *KRAS* corresponded to what have previously been reported on SBA[115,120,152]. Compared to CRC and GA, *TP53* displayed similar mutation frequencies, whereas *KRAS* mutation frequencies resembled that of CRC but were clearly higher than in GAs[45,222]. *KRAS* has clinical relevance as mutations in this gene predict lack of response to EGFR inhibitors [294]. Additionally, the role of anti-EGFR therapy in RAS wild-type SBA has been suggested to be unclear based on a small clinical trial where treatment showed no response[91,295].

The role of *APC* mutations in SBA genesis has been under debate and it has been proposed that it would not be as essential as in CRC[296–298]. Especially, the lack of nonsense mutations has been reported. *APC* was relatively frequently mutated in our set, as reported in recent studies[115,153]. Furthermore, the majority of the observed mutations were protein-truncating. *APC* mutation frequency increased along the GI-tract, confirming previous findings[115]. Although the overall mutation frequency of *APC* was lower than in CRC, our results support the importance of *APC* also in the SBA pathogenesis. *APC* mutations had also been suggested to occur exclusively in SBA patients without IBD[115]. Our results implied, however, that a subset of SBAs with IBD background can harbor inactivating *APC* mutations.

Among the highest-ranking genes was also *BRAF*. Interestingly, we detected no V600E mutations, the most common hotspot for mutations e.g. in CRC. Instead, we observed an atypical hotspot pattern with mutations clustering in two other hotspots nearby. These were present exclusively in MSS tumors, as previously indicated in CRC[299]. These hotspots have been reported commonly in some tumor types, such as in lung adenocarcinomas and melanoma[300]. Four SBAs harbored both *BRAF* and *KRAS* mutations. This could at least partly be explained by the fact that three of these *BRAF* mutations were likely kinase-silencing or truncating changes. Co-occurrence of kinase-impaired *BRAF* and mutant *KRAS* has been observed in various cancers[299–301]. Shrock *et al.* reported only a minority of *BRAF* mutations in SBA to be V600E, together with our results highlighting the importance of atypical *BRAF* mutations in SBA[115]. Metastatic CRC harboring such atypical mutations have been

shown to have distinct clinicopathological features and improved OS compared to *BRAF* V600E mutant CRCs[118]. How different non-V600 *BRAF* mutants respond to therapy is currently under investigation. Overall, these results suggest that screening also for atypical *BRAF* mutations might be clinically relevant for guidance of personalized treatment plan.

In addition to the known cancer genes in SBA, the most relevant genes on exome-wide analysis included other recently reported and novel candidate driver genes. One recently reported candidate driver gene in SBA was *ATM*, a suggested barrier to dysplastic growth in bowel tumors and a potentially clinically relevant biomarker to predict PARP inhibitor sensitivity[302]. Another recently suggested potential driver in SBA was *ARID2*, frequently mutated in various cancers and further discussed above (section 1 of Discussion)[153]. The novel candidate drivers in SBA identified by this study, *ACVR1B*, *ACVR2A*, *BRCA2*, and *SMARCA4*, have been implicated previously as driver genes in many other tumor types. The mutation status of these novel candidate drivers might be clinically relevant due to their potential role as drug targets and/or predictors of therapy response. As an example, somatic *BRCA2* mutations in melanoma have been found to correlate with anti-PD-1 responsiveness[303]. Also, drugs targeting *BRCA2* are already being developed. *SMARCA4* is a key SWI/SNF chromatin remodeling gene[224]. Due to its tumor suppressive role, *SMARCA4* is not an ideal drug target. However, a recent study proposed that, due to synthetic lethality, CDK4/6 inhibitors might be effective against SMARCA4-deficient tumors[304].

The members of the ERBB gene family, *EGFR*, *ERBB2*, *ERBB3*, and *ERBB4*, are commonly amplified, overexpressed, or mutated in various malignancies[305]. Of these, *ERBB2* was the most frequently altered (14%) in our set. Unlike in CRC, *ERBB2* mutations were mainly point mutations. We detected four known mutation hotspots, of which one had not previously been reported in SBA. Additionally, we identified localized and strong amplification of *ERBB2* in four samples. Two of these harbored also a hotspot mutation. The other gene family members were less frequently mutated (0.9-8.5%). Both *ERBB3* and *ERBB4* harbored one hotspot. These hotspots have been previously identified in e.g. CRC and gastric cancer but, to our knowledge, not in SBA[306]. ERBB family members have potential clinical relevance and targeted therapies are under development. In fact, multiple drugs targeting ERBB2 and EGFR receptors are already in clinical use. Based on our results together with other reports, of these genes especially *ERBB2* could be considered as a potential therapeutic target in SBA[115,152,307]. Furthermore, a recent preclinical study suggested that *ERBB2* tyrosine kinase inhibitors could play a role in the management of SBA[308].

In addition to searching for single genes and possible therapy targets, we examined the tumor set for most frequently affected, well-known cancer related pathways and performed the first comprehensive mutation signature analysis of SBA. Overall, the most commonly affected pathways in MSS SBA were PI3K/AKT and ERBB signaling. Some variability between segments were noted, e.g. ERBB signaling was the most frequently affected pathway in duodenal SBAs and PI3K/AKT signaling in jejunal and ileal SBAs. More studies are needed, however, to clarify these differences in more detail. Mutational signature analysis revealed four signatures: 1A, 17, and U2 in MSS and 6 in MSI tumors. We observed segment-wise variation regarding signature 1A. The amount of signature 1A exposure has been shown to associate with older age at diagnosis[63]. Signature 1A exposure was associated with jejunal location, regardless that the patients with jejunal tumors were, on average, younger. One reason for this could be possible regional differences in DNA methylation or in the cell division rate

between small bowel segments. Based on mutational signatures, SBAs seem to closely resemble CRC as the majority of their associated signatures overlapped[63]. This could reflect them possibly sharing similar exposures, although the small bowel and colorectum represent different environments.

As a strength of our study, by obtaining registry-based data together with comprehensive clinicopathological data on all SBA cases, we were able to collect a large, population-based dataset. This included tumors from all three segments of the small bowel with defined location, allowing us to examine the differences between these subsets. Exome data allowed us to elucidate the molecular basis of SBA but, as a limitation, we were not able to study the non-coding region of the genome nor the epigenetic changes. We also acknowledge that, regardless of the strict filtering, the data may contain some rare germline mutations or low-frequency artefacts due to FFPE material. Despite the rather large dataset, we were not able to detect any obvious genetic reason for the difference in incidence between SBA and CRC.

To our knowledge, this was the first large-scale exome study conducted on SBAs. Taken together, these results indicate that SBA is a distinct tumor type with its unique set of significantly mutated genes. Heterogeneity in the mutation landscape indicates that several driver genes can affect SBA genesis. There may also be some variation between tumors originating from different segments. One of the clearest differences from CRC was the *BRAF* mutation spectrum which might be relevant when designing genetic testing and treatment of SBA. The results highlighted many potential treatment targets, both currently targetable (*ERBB2*, *BRAF*, and *BRCA2*) as well as new candidates (including *ERBB3*, *ERBB4*, *ATM*, *ACVR2A*, *ACVR1B*, and *SMARCA4*). Especially alterations in *ERBB2* and *BRCA2*, atypical *BRAF* hotspots, and MSI status have been discussed as options for more personalized treatment in SBA also more recently[91,307,308]. Thus far SBA has been treated following the same regimen as CRC in the advanced setting. These recent advances in characterizing SBA have led to improved understanding of the tumor type and to the first set of National Comprehensive Cancer Network (NCCN) guidelines for the treatment of SBA, published autumn 2019[309]. Further research is, however, required to obtain a better understanding of SBA. Finally, studies such as this are good examples that exome-wide characterization of a rare tumor type can provide more knowledge for the basis of more personalized treatment and thus should be considered valuable in other rare tumor types as well.

5. Toward future cancer treatments

Exome sequencing has enabled us to examine simultaneously the whole coding region of the genome with a reasonable price. By utilizing this technology, we have provided many potential novel drivers as well as further knowledge on tumor biology of bowel tumors. Examples of clinically relevant findings in this thesis work include the observed differences in SCRC pairs that can affect the treatment planning for patients with SCRC and assessment of the mutation landscape of SBA that indicate genetic screening for eg. atypical *BRAF* mutations to be beneficial for optimal treatment. A comprehensive view on genetic changes underlying tumorigenesis is needed to improve treatments as well as preventive measures. Knowledge on mutation content directs targeted treatment planning as well as helps identifying new potential candidates for drug development targets.

Through advancements of sequencing technology, research has also uncovered frequent mutations in genes involved in processes previously not associated with cancer and thus provided novel options to pursue in treating tumors more efficiently. One example is chromatin-related proteins, identified to be affected in over 50% of cancers[310]. In this study, we also detected multiple altered genes that encode for chromatin remodelers, such as *ARID1A*, *ARID1B*, *ARID2*, *SMARCA4*, and *MORC2*. We observed great heterogeneity in the mutational landscape of the bowel tumors indicating that many driver genes play a role in their biology. Since most cancer driver genes are known to be mutated at intermediate or lower frequencies, as also observed here, the characterization of these genes continues as an important effort to be able to provide optimal therapeutic options for a larger number of patients and more effective combination treatments. In the future, larger datasets and more advanced technologies, such as WGS and third-generation sequencing, aid in taking this effort forward. By understanding the molecular pathogenesis of bowel tumors paves the way towards identification of druggable targets with the goal to ultimately benefit the patients.

CONCLUSIONS AND FUTURE PROSPECTS

CRC is one of the most common cancers and due to extensive research, it is one of the best characterized tumor types. The most frequently mutated driver genes have been discovered but more insight is still needed on genes mutated with moderate or low frequency that contribute to tumorigenesis. Despite the rapid accumulation of knowledge on the genetic changes in CRC, until recently the clinical utility of molecular data has been rather limited. Specific CRC subtypes also require more research. As an example, MSI tumors represent a molecular subclass of CRC, where the detection of driver genes has been challenging and only few oncogenes are known. A subset of CRCs can also occur as multiple primary tumors residing in one patient. Increasing knowledge on the mutation content of these synchronous lesions helps to illustrate how they have arisen and how these cases should be treated. In contrast to CRC, SBA is a rare tumor type and thus less is known of its genetic background. It is an aggressive disease and there is a dire need for better understanding of its molecular characteristics. NGS has enabled the sequencing of large tumor sets and thus facilitated a more comprehensive view of the driving forces behind tumorigenesis. Large-scale genetic information together with clinicopathological data provides opportunities to understand the biology underlying tumor development and guides the development of treatments.

The aim of this thesis was to provide new insights into the molecular genetic background of CRC and SBA. Focus of studies I-II was on somatic mutations in MSI CRC to identify novel candidate driver genes, study III concentrated on understanding whether SCRCs within a patient resemble each other genetically and study IV aimed to characterize the somatic mutation content of SBA.

I

In study I, in order to examine the role of ARID gene family in MSI CRC, we characterized the mutation spectrum of these genes in 25 MSI CRCs, followed by a further analysis of the four most prominent candidate genes in an additional set of 21 MSI CRCs. We confirmed that *ARID1A* is frequently mutated in MSI CRC. In addition, *ARID1B*, *ARID2*, and *ARID4A* were rather frequently mutated and could represent potential candidate driver genes in MSI CRC. Altogether 61% of the patients had a tumor with a mutation in at least one of these four genes. This was the first attempt to characterize the mutations of the whole ARID gene family. In the future, functional efforts are needed to determine the mechanistic details of how these mutations might contribute to tumorigenesis. Additionally, further research is required to better define the role of ARID genes in various cancer types. Should alterations in these genes be involved in the development of MSI CRC and other tumor types, they could be utilized as predictors of therapy efficacy as well as in the development of targeted drug therapies.

II

Only few oncogenes are known in MSI CRC. Thus, in study II, we aimed to identify novel candidate oncogenes by searching for genes with somatic mutation hotspots. This mutation pattern is typical of oncogenes, a sign of selection in tumor development, and an attractive target for drug development. Exome sequencing data from 25 MSI CRCs and their corresponding normal tissues revealed 33 novel candidate oncogenes with a recurrent mutation hotspot. Of these, 14 displayed additional hotspot

mutations in a validation set of 254 MSI CRCs. Among them was *ANTXR1* that emerged as the most prominent candidate of the study. It is unknown why the hotspot mutations were not selected for in the mature tumors and further efforts are needed to understand the biological background of the phenomenon. In general, upregulation of *ANTXR1* has been suggested to promote tumor growth and progression[311]. Additionally, many therapies targeting *ANTXR1* have shown anti-tumor efficacy in preclinical models.

These findings underline the notion that CRCs are heterogeneous tumors whose mutation landscape comprises few frequently affected genes and many less frequently mutated genes. It is of great importance to continue creating a catalogue of driver genes in order to further characterize dysregulated pathways and potential drug targets. In this study, altogether 14% of the patients had a tumor that harbored at least one of the novel hotspot mutations and could potentially benefit from a targeted treatment. In the future, additional studies are needed to validate these findings in independent tumor series. Furthermore, functional efforts are necessary to thoroughly understand their pathogenic role in carcinogenesis. Taken together, this study uncovered a variety of novel recurrent candidate oncogene mutation hotspots to be further scrutinized as potential therapeutic targets.

III

In study III, in order to estimate the genetic overlap within SCRCs, we exome-sequenced 23 paired tumors and their corresponding normal samples. This study revealed great heterogeneity between SCRCs within each patient implicating that, despite similar environment and genetic background, these tumors may be driven by distinct molecular events and develop independently. In addition to the lack of shared somatic mutations, the amount of genetic overlap varied in other aspects such as mutation statutes of known CRC genes and signaling pathways. By analyzing IS, we observed that there was also variation in the immune reaction between patients and, in most of the cases, also within paired tumors.

Our study confirmed previous observations and elucidated further the genetic background of SCRC and variation of immune responses within tumors, improving our understanding of the diversity within SCRCs. These results strengthen the notion that SCRCs harbor variation in common drug targets as well as other potential targets, mutation burden, and single genetic changes that might affect response to treatment. Such variation within paired tumors included eg. the discordant MMR status that affected around 9% of the patients and the discordant *KRAS* mutation status that was present in 26% of the cases. Thus, this study should provide valuable information for the clinical management of patients with SCRC. Therapeutic decisions, especially regarding targeted therapies, should consider SCRCs as individual cancers and thus all lesions should be analyzed for optimal treatment outcome. Further studies are, however, still needed to decipher the reasons for tumor multiplicity.

IV

In study IV, in order to gain better understanding on the molecular genetic background of SBA, we gathered and analyzed a unique, population-based set of 106 primary SBAs, thus far the largest exome set of this tumor type. Our study confirmed the role of known drivers in this cancer as well as recently

reported potential drivers, and implicated possible novel drivers. The data showed some differences between SBAs from the three small bowel segments; however, the relevance of these differences requires further studies. Based on the results, SBAs harbor both similarities and differences with CRCs. The proportion of MSI tumors, mutation frequency of certain driver genes, and mutation signature content resembled those of CRC. MSI screening could thus also be warranted in SBA in view of immunotherapy. Whereas, notable differences were seen in e.g. potentially clinically relevant aspects between SBA and CRC, such as the atypical mutation spectrum of *BRAF* and the *ERBB2* mutations consisting mainly of point mutations. This study provided information also on other potential therapeutic targets or predictors of therapy response. Overall, comprehensive characterization of SBA yielded new and more thorough insights into the genetic background of this tumor type, providing further evidence that SBA is a genetically distinct tumor type. In the future, further validation of the findings in other sample sets and functional validation is warranted to thoroughly understand their role in SBA genesis. Altogether, this study improved our understanding of SBA biology that should ultimately be useful for the management of this cancer.

Future prospects

Over the past two decades, the field of oncology has undergone conceptual revolutions that have shaped our understanding of cancer. One enormous step forward has resulted from international efforts to sequence a large amount of cancers with base-by-base accuracy to obtain a comprehensive view on carcinogenesis. The development of high-throughput technologies and decrease in their costs have enabled the increase in sequencing intensity. This has resulted in vast amounts of information. Integration of mutation data with other data types, such as epigenomic (e.g. DNA methylation) and transcriptomic (mRNA and protein expression) data, has enhanced our ability to interpret the potential biological role of the identified alterations, including those of unknown significance. Furthermore, emerging technologies have brought novel ways to study tumorigenesis. Recent advances include third generation sequencing, also called long-read sequencing, that allows e.g. the detection of complicated structural changes on a haplotype level, single-cell sequencing that provides more detailed insight into tumor heterogeneity, CRISPR-Cas9 genome editing that enables the functional characterization of mutations, and methods, such as Hi-C and ATAC sequencing, that are used to study the regulatory genome.

Although deciphering the molecular pathogenesis of cancers is advancing, there are many challenges in designing targeted therapies. Most prevalent oncogenic mutations are currently undruggable. Thus, in addition to finding means to target these changes, identifying new targets is of great importance. Genes with mutational hotspots, as an example, represent attractive therapy targets. Also tumor suppressors, in general, depict a challenging treatment target, since driver mutations in these genes lead to functional loss of the protein. However, the search for alternative routes is ongoing. One suggested approach is synthetic lethality that is successfully used to target *BRCA* deficient breast and ovarian tumors with PARP inhibitors[312]. Furthermore, recent technical advances, such as the above-mentioned CRISPR-Cas9-based gene editing, have enabled more systematic screening of potential synthetic lethal drug targets[313]. Other therapeutic strategies are also being developed, such as targeting chromatin remodelers and harnessing the patient's own immune system[310,314].

Further challenges in treatment are introduced by tumor heterogeneity which both generates primary drug resistance and is the basis for genetic tumor evolution that drives secondary resistance. Understanding the critical signaling pathways for tumorigenesis is one of the main goals to target a wide range of tumor cells[4]. Signaling pathways, however, function differently in different tissues and thus a drug that is efficient in a certain tumor might not bring benefit in another. Such an example is vemurafenib that is successfully used in metastatic melanoma with *BRAF* V600E mutation but has no therapeutic effect in metastatic CRC harboring the same mutation[315]. Drug efficiency can also be affected by other accompanying genetic alterations (both germline and somatic) and tumor microenvironment. Thus, better understanding of the mutation content in different tissue contexts will enable more efficient targeted therapies. On the other hand, cancer subclasses arising from different tissues can also harbor similar molecular features[316]. Therefore, integration of diverse sets of cancers and grouping them based on their molecular relationships instead of anatomical location can be beneficial when exploring novel clinically actionable targets.

Cancer continues being a major public health problem worldwide. Thus, future research efforts are direly needed - also for the rarer tumor types. Although obtaining samples of a rare tumor type is not an easy task, these cancers affect many people and such studies can also help understanding the more common diseases. In the future, the aim is to improve utilization of cancer genomes for individualized therapeutic decisions, such as the ability to predict prognosis, to choose correct combinatorial treatment regimen that has efficacy for a particular genetic subtype, to be able to monitor therapy responses, and to identify drug-resistant subclones. Expanding information also on genetic predisposition, environmental exposures, and their combined effects provides means for better cancer risk assessment. It is imperative to improve understanding of the complex biology underlying tumor evolution to direct future efforts of reducing cancer burden through prevention and more efficient care - and to ultimately improve patients' outcome.

ACKNOWLEDGEMENTS

This work was carried out at the Department of Medical and Clinical Genetics and Applied Tumor Genomics Research Program, Faculty of Medicine, University of Helsinki, between 2012 and 2020. I thank the present and former directors for the excellent research facilities. This thesis was financially supported by personal grants from The Biomedicum Helsinki Foundation, The Cancer Society of Finland, Duodecim, The Einar and Karin Stroem Foundation, The Emil Aaltonen Foundation, The Finnish Gastroenterology Society, The Finnish-Norwegian Medical Foundation, The Ida Montin Foundation, The Instrumentarium Science Foundation, The Juhani Aho Foundation for Medical Research, The K. Albin Johanssons Foundation, The Maud Kuistila Memorial Foundation, The Orion Research Foundation, The Oskar Öflunds Foundation, and the Päivikki and Sakari Sohlberg Foundation. All are most sincerely thanked.

My deepest appreciation goes to my supervisor Lauri Aaltonen for granting me the opportunity to work in such a fantastic research group. Lauri's scientific excellence combined with his caring and humorous personality have created a unique research athosphere. I admire his infinite research ideas, devotion and enthusiasm towards science, and care and integrity towards the patients. I highly value the encouraging feedback and continuous support throughout the years. I am also grateful for the amazing opportunities to attend and present our work in conferences abroad. I thank Lauri for trusting in me and giving me responsibility to also work independently, and for his policy of keeping his door always open. I appreciate his ability to gather an amazing group of people to work with - it has been a privilige to be part of it.

Sari Tuupanen and Netta Mäkinen are sincerely thanked for being my other supervisors. Sari is thanked for being there at the beginning of this journey, introducing me to the research world, and teaching me a variety of research skills. Netta is thanked for her guidance and help, encouraging words, and heartfelt support towards the end of this effort. I admire Netta's passion for science and her kind and determined character. I also thank Netta for the company during conference trips and showing me around New York.

Reviewers of this thesis, professor Arto Mannermaa and docent Tero Hiekkalinna, are thanked for their constructive comments. Their expertize helped improving this work considerably. I also extend my thanks to Pia Österlund and Pirkko Mattila for their efforts on the thesis committee and their encouraging words.

I thank all the co-authors and collaborators for their invaluable contribution. Surgeons Heikki Järvinen, Jukka-Pekka Mecklin, Laura-Renkonen Sinisalo, and Anna Lepistö are thanked for the fantastic sample material; Jukka-Pekka Mecklin is also thanked for various interesting research ideas and discussions; Ari Ristimäki and Jan Böhm are thanked for their pathological expertize; Jussi and Minna Taipale are thanked for the efforts on exome sequencing; Toni Seppälä is thanked for the fruitful discussions and collaboration; Erkki-Wille Virta is thanked for the productive collaboration; and Claus L. Andersen is thanked for providing samples. I extend my thanks also to the organizers of the Center of Excellence meetings, especially Tiia, and the organizers of the Lynch syndrome meetings for creating such inspiring and high-quality research events.

My warmest thanks goes to the whole Aaltonen group, both past and present members, with whom working has truly been a pleasure. I am grateful for all the lively lunch table discussions, adventurous recreational days, fun parties, and meetings by the coffee machine. I thank the past and present senior researchers and post-docs, Auli, Anna, Davide, Eevi, Janne, Javier, Kimmo, Kristiina, Linda F., Linda vdB, Mervi, Niko, Outi, Pia, Rainer, Saija, Simona, and Tuomas, for all the scientific and practical help. Auli, Esa, and Janne are thanked for their assistance in various matters throughout the years, their good sense of humor, and for all the delightful conversations, Davide is thanked for his willingness to explain difficult concepts and enthusiasm for playing board games, Kimmo and Niko are thanked for their expertize and patient assistance with bioinformatics matters beoynd my understanding, Outi and Pia for the guidance in the ARID project, Linda vdB for being part of the CRC team and her warm personality, Eevi for her skillful help and positive spirit, and Mervi for her enthusiastic attitude, kind support, and friendship.

The past and present PhD students are greatly acknowledged: Alexandra, Aurora, Hanna, Heli K., Iikki, Jaana, Johanna, Kati, Lauri S., Maritta, Miika, Päivi, Riku K., Silva, Samantha, Tomas, and Tatiana. Alex is thanked for teaching me many skills in the lab, friendship, and her cheerful personality. Lauri S. and Iikki are thanked for the enjoyable discussions, peer support, and their delightful visits to the CRC room. Riku K. is thanked for the great team work on our common projects, his expertize and easy-going attitude, and for creating Baseplayer. Tomas is thanked for all the interesting scientific conversations, friendship, and his talented and invaluable help. I highly appreciate his eagerness and patience to explain things and his general can-do attitude.

I extend my special thanks to Alison, Aurora, Jaana, Johanna, Samantha, and Tatiana for being wonderful collagues and trusted friends. All are thanked for contributing to the supporting, creative, and pleasant working environment. Alison is thanked for sharing both the small and big steps in life, for her uplifting and valued presence, English language checking, and for being the best company for traveling and countless movies and BodyCombat classes. Aurora is thanked for her witty and caring character, creativity, and imaginative experiments. Jaana is thanked for her kind and compassionate spirit, lovely company working side-by-side, and for our memorable conference trips such as the windy Chicago. Johanna is thanked for introducing the lab to me, her insightful personality, and for being the constant peer support in scientific, clinical, and other areas in life. Sam is thanked for her sweet character and participating attitude. Tati is thanked for her genuine and caring personality, for sharing the whole doctoral path from the ARID project to writing the thesis, including our adventurous conference trip to San Diego, and for introducing the lovely Menorca. I am grateful for each day I got to work and spend time with this gang.

I greatly acknowledge Marjo, Sini, and Sirpa for creating such a welcoming and caring atmosphere, making it always easy and nice to turn to them for help. I appreciate all the cheerful conversations throughout the years. I am deeply grateful to Marjo for our seemless collaboration with the block collection effort, numerous laughs, and her kindhearted support in all matters. Sini is thanked for her company both at work and our gaming challenges, for handling the samples, and caring for everyone's well-being. Sirpa is thanked for her excellent secreterial assistance, kind personality, and taking such good care for all practical matters. Alison (again), Iina, Inga-Lill, and Justyna are warmly thanked for their invaluable and skillful help in the lab and for their delightful company. Sini and Iina are also

acknowledged for organizing events, such as the recreational days. Heikki and Jiri are thanked for their help with all IT-related matters and the fun conversations.

I warmly acknowledge Aino, Hande, Heikki R, Klaus, Kornelia, Magdalena, Maija, Petra, Riku L., Roosa, Sanna, Sofie, Viivi, and all the other past and present students in the lab. I thank Riku L. for his contribution to the statistical analysis, Roosa for her talented help with bioinformatics analysis and creating figures, Sofie for her energetic and positive attitude, and Petra for her lovely and caring personality, friendship, and the unforgetable trip to Tanzania.

I express my deepest gratitude towards my family and friends outside work. Elina and Marika are thanked for their long-lasting friendship throughout the years. I thank all members of the Sulkava rowing and Tough Viking teams for the great memories and the wonderful team spirit that made practices fun and got us to the finish line. Special thanks to Tiia (again) and Karen for their friendship, both in and outside of sports. Arvydas is heartily thanked for sharing the last part of the doctoral journey, for the encouragement, partnership, silly laughs, and his kind heart. I thank my family for the loving care and continuous support. I am deeply grateful to my parents, Eila and Seppo, for their endless help, wisdom, encouragement, and the wonderful cottage trips. Johanna and Heikki are greatly thanked for being such caring and simply the best siblings, Joonas and Linda are thanked for their delightful company, and Venla, my niece and goddaughter, is thanked for bringing joy and never-ending suprises to our lives. I thank Panda and Chili for the decade of loyal companionship.

Finally, I extend my deep gratitute to all patients who donated samples and made this research possible.

Helsinki, June 2020

Ulrika Hänninen

REFERENCES

1. Mehlen, P. & Puisieux, A. Metastasis: a question of life or death. *Nat. Rev. Cancer* **6,** 449–458 (2006).
2. Nowell, P. The clonal evolution of tumor cell populations. *Science* **194,** 23–28 (1976).
3. Stratton, M. R., Campbell, P. J. & Futreal, P. A. The cancer genome. *Nature* **458,** 719–724 (2009).
4. Vogelstein, B., Papadopoulos, N., Velculescu, V. E., Zhou, S., Diaz, L. A., Jr & Kinzler, K. W. Cancer genome landscapes. *Science* **339,** 1546–1558 (2013).
5. Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: the next generation. *Cell* **144,** 646–674 (2011).
6. Hanahan, D. & Weinberg, R. A. The hallmarks of cancer. *Cell* **100,** 57–70 (2000).
7. Loeb, L. A., Loeb, K. R. & Anderson, J. P. Multiple mutations and cancer. *Proc. Natl. Acad. Sci. U. S. A.* **100,** 776–781 (2003).
8. zur Hausen, H. Papillomaviruses and cancer: from basic studies to clinical application. *Nat. Rev. Cancer* **2,** 342–350 (2002).
9. Parsonnet, J., Friedman, G. D., Vandersteen, D. P., Chang, Y., Vogelman, J. H., Orentreich, N. & Sibley, R. K. Helicobacter pyloriInfection and the Risk of Gastric Carcinoma. *N. Engl. J. Med.* **325,** 1127–1131 (1991).
10. Tomasetti, C. & Vogelstein, B. Cancer etiology. Variation in cancer risk among tissues can be explained by the number of stem cell divisions. *Science* **347,** 78–81 (2015).
11. Greenman, C., Stephens, P., Smith, R., Dalgliesh, G. L., Hunter, C., Bignell, G., Davies, H., Teague, J., Butler, A., Stevens, C., Edkins, S., O'Meara, S., Vastrik, I., Schmidt, E. E., Avis, T., Barthorpe, S., Bhamra, G., Buck, G., Choudhury, B., Clements, J., Cole, J., Dicks, E., Forbes, S., Gray, K., Halliday, K., Harrison, R., Hills, K., Hinton, J., Jenkinson, A., Jones, D., Menzies, A., Mironenko, T., Perry, J., Raine, K., Richardson, D., Shepherd, R., Small, A., Tofts, C., Varian, J., Webb, T., West, S., Widaa, S., Yates, A., Cahill, D. P., Louis, D. N., Goldstraw, P., Nicholson, A. G., Brasseur, F., Looijenga, L., Weber, B. L., Chiew, Y.-E., DeFazio, A., Greaves, M. F., Green, A. R., Campbell, P., Birney, E., Easton, D. F., Chenevix-Trench, G., Tan, M.-H., Khoo, S. K., Teh, B. T., Yuen, S. T., Leung, S. Y., Wooster, R., Futreal, P. A. & Stratton, M. R. Patterns of somatic mutation in human cancer genomes. *Nature* **446,** 153–158 (2007).
12. Milholland, B., Dong, X., Zhang, L., Hao, X., Suh, Y. & Vijg, J. Differences between germline and somatic mutation rates in humans and mice. *Nat. Commun.* **8,** 15183 (2017).
13. Tomasetti, C., Li, L. & Vogelstein, B. Stem cell divisions, somatic mutations, cancer etiology, and cancer prevention. *Science* **355,** 1330–1334 (2017).
14. Tomasetti, C., Marchionni, L., Nowak, M. A., Parmigiani, G. & Vogelstein, B. Only three driver gene mutations are required for the development of lung and colorectal cancers. *Proc. Natl. Acad. Sci. U. S. A.* **112,** 118–123 (2015).
15. Bailey, M. H., Tokheim, C., Porta-Pardo, E., Sengupta, S., Bertrand, D., Weerasinghe, A., Colaprico, A., Wendl, M. C., Kim, J., Reardon, B., Kwok-Shing Ng, P., Jeong, K. J., Cao, S., Wang, Z., Gao, J., Gao, Q., Wang, F., Liu, E. M., Mularoni, L., Rubio-Perez, C., Nagarajan, N., Cortés-Ciriano, I., Zhou, D. C., Liang, W.-W., Hess, J. M., Yellapantula, V. D., Tamborero, D., Gonzalez-Perez, A., Suphavilai, C., Ko, J. Y., Khurana, E., Park, P. J., Van Allen, E. M., Liang, H., MC3 Working Group, Cancer Genome Atlas Research Network, Lawrence, M. S., Godzik, A., Lopez-Bigas, N., Stuart, J., Wheeler, D., Getz, G., Chen, K., Lazar, A. J., Mills, G. B., Karchin, R. & Ding, L. Comprehensive Characterization of Cancer Driver Genes and Mutations. *Cell* **174,** 1034–1035 (2018).
16. Croce, C. M. Oncogenes and cancer. *N. Engl. J. Med.* **358,** 502–511 (2008).
17. Reddy, E. P., Reynolds, R. K., Santos, E. & Barbacid, M. A point mutation is responsible for the acquisition of transforming properties by the T24 human bladder carcinoma oncogene. *Nature* **300,** 149–152 (1982).
18. Zehir, A., Benayed, R., Shah, R. H., Syed, A., Middha, S., Kim, H. R., Srinivasan, P., Gao, J., Chakravarty, D., Devlin, S. M., Hellmann, M. D., Barron, D. A., Schram, A. M., Hameed, M., Dogan, S., Ross, D. S., Hechtman, J. F., DeLair, D. F., Yao, J., Mandelker, D. L., Cheng, D. T., Chandramohan, R., Mohanty, A. S., Ptashkin, R. N., Jayakumaran, G., Prasad, M., Syed, M. H., Rema, A. B., Liu, Z. Y., Nafa, K., Borsu, L., Sadowska, J., Casanova, J., Bacares, R., Kiecka, I. J., Razumova, A., Son, J. B., Stewart, L., Baldi, T., Mullaney, K. A., Al-Ahmadie, H., Vakiani, E., Abeshouse, A. A., Penson, A. V., Jonsson, P., Camacho, N., Chang, M. T., Won, H. H., Gross, B. E., Kundra, R., Heins, Z. J., Chen, H.-W., Phillips, S., Zhang, H., Wang, J., Ochoa, A., Wills, J., Eubank, M., Thomas, S. B., Gardos, S. M., Reales, D. N., Galle, J., Durany, R., Cambria, R., Abida, W., Cercek, A., Feldman, D.

R., Gounder, M. M., Hakimi, A. A., Harding, J. J., Iyer, G., Janjigian, Y. Y., Jordan, E. J., Kelly, C. M., Lowery, M. A., Morris, L. G. T., Omuro, A. M., Raj, N., Razavi, P., Shoushtari, A. N., Shukla, N., Soumerai, T. E., Varghese, A. M., Yaeger, R., Coleman, J., Bochner, B., Riely, G. J., Saltz, L. B., Scher, H. I., Sabbatini, P. J., Robson, M. E., Klimstra, D. S., Taylor, B. S., Baselga, J., Schultz, N., Hyman, D. M., Arcila, M. E., Solit, D. B., Ladanyi, M. & Berger, M. F. Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients. *Nat. Med.* **23,** 703–713 (2017).

19. Davies, H., Bignell, G. R., Cox, C., Stephens, P., Edkins, S., Clegg, S., Teague, J., Woffendin, H., Garnett, M. J., Bottomley, W., Davis, N., Dicks, E., Ewing, R., Floyd, Y., Gray, K., Hall, S., Hawes, R., Hughes, J., Kosmidou, V., Menzies, A., Mould, C., Parker, A., Stevens, C., Watt, S., Hooper, S., Wilson, R., Jayatilake, H., Gusterson, B. A., Cooper, C., Shipley, J., Hargrave, D., Pritchard-Jones, K., Maitland, N., Chenevix-Trench, G., Riggins, G. J., Bigner, D. D., Palmieri, G., Cossu, A., Flanagan, A., Nicholson, A., Ho, J. W. C., Leung, S. Y., Yuen, S. T., Weber, B. L., Seigler, H. F., Darrow, T. L., Paterson, H., Marais, R., Marshall, C. J., Wooster, R., Stratton, M. R. & Futreal, P. A. Mutations of the BRAF gene in human cancer. *Nature* **417,** 949–954 (2002).
20. Dalla-Favera, R., Bregni, M., Erikson, J., Patterson, D., Gallo, R. C. & Croce, C. M. Human c-myc onc gene is located on the region of chromosome 8 that is translocated in Burkitt lymphoma cells. *Proc. Natl. Acad. Sci. U. S. A.* **79,** 7824–7827 (1982).
21. Taub, R., Kirsch, I., Morton, C., Lenoir, G., Swan, D., Tronick, S., Aaronson, S. & Leder, P. Translocation of the c-myc gene into the immunoglobulin heavy chain locus in human Burkitt lymphoma and murine plasmacytoma cells. *Proc. Natl. Acad. Sci. U. S. A.* **79,** 7837–7841 (1982).
22. Erikson, J., Nishikura, K., ar-Rushdi, A., Finan, J., Emanuel, B., Lenoir, G., Nowell, P. C. & Croce, C. M. Translocation of an immunoglobulin kappa locus to a region 3' of an unrearranged c-myc oncogene enhances c-myc transcription. *Proc. Natl. Acad. Sci. U. S. A.* **80,** 7581–7585 (1983).
23. Perner, S., Mosquera, J.-M., Demichelis, F., Hofer, M. D., Paris, P. L., Simko, J., Collins, C., Bismar, T. A., Chinnaiyan, A. M., De Marzo, A. M. & Rubin, M. A. TMPRSS2-ERG fusion prostate cancer: an early molecular event associated with invasion. *Am. J. Surg. Pathol.* **31,** 882–888 (2007).
24. Boland, C. R. & Goel, A. Microsatellite instability in colorectal cancer. *Gastroenterology* **138,** 2073–2087.e3 (2010).
25. Kinzler, K. W. & Vogelstein, B. Cancer-susceptibility genes. Gatekeepers and caretakers. *Nature* **386,** 761, 763 (1997).
26. Carethers, J. M. & Jung, B. H. Genetics and Genetic Biomarkers in Sporadic Colorectal Cancer. *Gastroenterology* **149,** 1177–1190.e3 (2015).
27. Knudson, A. G., Jr. Mutation and cancer: statistical study of retinoblastoma. *Proc. Natl. Acad. Sci. U. S. A.* **68,** 820–823 (1971).
28. Friend, S. H., Bernards, R., Rogelj, S., Weinberg, R. A., Rapaport, J. M., Albert, D. M. & Dryja, T. P. A human DNA segment with properties of the gene that predisposes to retinoblastoma and osteosarcoma. *Nature* **323,** 643–646 (1986).
29. Santarosa, M. & Ashworth, A. Haploinsufficiency for tumour suppressor genes: when you don't need to go all the way. *Biochim Biophys Acta Rev Cancer* **1654,** 105–122 (2004).
30. Berger, A. H., Knudson, A. G. & Pandolfi, P. P. A continuum model for tumour suppression. *Nature* **476,** 163–169 (2011).
31. Chen, Z., Trotman, L. C., Shaffer, D., Lin, H.-K., Dotan, Z. A., Niki, M., Koutcher, J. A., Scher, H. I., Ludwig, T., Gerald, W., Cordon-Cardo, C. & Pandolfi, P. P. Crucial role of p53-dependent cellular senescence in suppression of Pten-deficient tumorigenesis. *Nature* **436,** 725–730 (2005).
32. Fearon, E. R. & Vogelstein, B. A genetic model for colorectal tumorigenesis. *Cell* **61,** 759–767 (1990).
33. Vogelstein, B. & Kinzler, K. W. Cancer genes and the pathways they control. *Nat. Med.* **10,** 789–799 (2004).
34. Veigl, M. L., Kasturi, L., Olechnowicz, J., Ma, A. H., Lutterbaugh, J. D., Periyasamy, S., Li, G. M., Drummond, J., Modrich, P. L., Sedwick, W. D. & Markowitz, S. D. Biallelic inactivation of hMLH1 by epigenetic gene silencing, a novel mechanism causing human MSI cancers. *Proc. Natl. Acad. Sci. U. S. A.* **95,** 8698–8702 (1998).
35. Fanale, D., Maragliano, R., Bazan, V. & Russo, A. Caretakers and Gatekeepers. *eLS* 1–10 (2017). doi:10.1002/9780470015902.a0006048.pub2
36. Shen, J., Peng, Y., Wei, L., Zhang, W., Yang, L., Lan, L., Kapoor, P., Ju, Z., Mo, Q., Shih, I.-M., Uray, I. P., Wu, X., Brown, P. H., Shen, X., Mills, G. B. & Peng, G. ARID1A Deficiency Impairs the DNA Damage Checkpoint and Sensitizes Cells to PARP Inhibitors. *Cancer Discov.* **5,** 752–767 (2015).

37. Wu, R.-C., Wang, T.-L. & Shih, I.-M. The emerging roles of ARID1A in tumor suppression. *Cancer Biol. Ther.* **15,** 655–664 (2014).
38. Wilsker, D., Patsialou, A., Dallas, P. B. & Moran, E. ARID proteins: a diverse family of DNA binding proteins implicated in the control of cell growth, differentiation, and development. *Cell Growth Differ.* **13,** 95–106 (2002).
39. Michor, F., Iwasa, Y. & Nowak, M. A. Dynamics of cancer progression. *Nat. Rev. Cancer* **4,** 197–205 (2004).
40. Morris, L. G. T. & Chan, T. A. Therapeutic targeting of tumor suppressor genes. *Cancer* **121,** 1357–1368 (2015).
41. Negrini, S., Gorgoulis, V. G. & Halazonetis, T. D. Genomic instability — an evolving hallmark of cancer. *Nat. Rev. Mol. Cell Biol.* **11,** 220–228 (2010).
42. Thompson, S. L., Bakhoum, S. F. & Compton, D. A. Mechanisms of chromosomal instability. *Curr. Biol.* **20,** R285–95 (2010).
43. Shih, I. M., Zhou, W., Goodman, S. N., Lengauer, C., Kinzler, K. W. & Vogelstein, B. Evidence that genetic instability occurs at an early stage of colorectal tumorigenesis. *Cancer Res.* **61,** 818–822 (2001).
44. Bach, D.-H., Zhang, W. & Sood, A. K. Chromosomal Instability in Tumor Initiation and Development. *Cancer Res.* **79,** 3995–4002 (2019).
45. Cancer Genome Atlas Network. Comprehensive molecular characterization of human colon and rectal cancer. *Nature* **487,** 330–337 (2012).
46. Palles, C., Cazier, J.-B., Howarth, K. M., Domingo, E., Jones, A. M., Broderick, P., Kemp, Z., Spain, S. L., Guarino, E., Salguero, I., Sherborne, A., Chubb, D., Carvajal-Carmona, L. G., Ma, Y., Kaur, K., Dobbins, S., Barclay, E., Gorman, M., Martin, L., Kovac, M. B., Humphray, S., CORGI Consortium, WGS500 Consortium, Lucassen, A., Holmes, C. C., Bentley, D., Donnelly, P., Taylor, J., Petridis, C., Roylance, R., Sawyer, E. J., Kerr, D. J., Clark, S., Grimes, J., Kearsey, S. E., Thomas, H. J. W., McVean, G., Houlston, R. S. & Tomlinson, I. Germline mutations affecting the proofreading domains of POLE and POLD1 predispose to colorectal adenomas and carcinomas. *Nat. Genet.* **45,** 136–144 (2013).
47. Sharma, S., Kelly, T. K. & Jones, P. A. Epigenetics in cancer. *Carcinogenesis* **31,** 27–36 (2010).
48. Mathur, R. ARID1A loss in cancer: Towards a mechanistic understanding. *Pharmacol. Ther.* **190,** 15–23 (2018).
49. Kadoch, C., Hargreaves, D. C., Hodges, C., Elias, L., Ho, L., Ranish, J. & Crabtree, G. R. Proteomic and bioinformatic analysis of mammalian SWI/SNF complexes identifies extensive roles in human malignancy. *Nat. Genet.* **45,** 592–601 (2013).
50. Fletcher, O. & Houlston, R. S. Architecture of inherited susceptibility to common cancer. *Nat. Rev. Cancer* **10,** 353–361 (2010).
51. Nagy, R., Sweet, K. & Eng, C. Highly penetrant hereditary cancer syndromes. *Oncogene* **23,** 6445–6470 (2004).
52. Rahman, N. Realizing the promise of cancer predisposition genes. *Nature* **505,** 302–308 (2014).
53. Sud, A., Kinnersley, B. & Houlston, R. S. Genome-wide association studies of cancer: current insights and future perspectives. *Nat. Rev. Cancer* **17,** 692–704 (2017).
54. Consortium, I. H. G. S. & International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. *Nature* **409,** 860–921 (2001).
55. Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., Smith, H. O., Yandell, M., Evans, C. A., Holt, R. A., Gocayne, J. D., Amanatides, P., Ballew, R. M., Huson, D. H., Wortman, J. R., Zhang, Q., Kodira, C. D., Zheng, X. H., Chen, L., Skupski, M., Subramanian, G., Thomas, P. D., Zhang, J., Gabor Miklos, G. L., Nelson, C., Broder, S., Clark, A. G., Nadeau, J., McKusick, V. A., Zinder, N., Levine, A. J., Roberts, R. J., Simon, M., Slayman, C., Hunkapiller, M., Bolanos, R., Delcher, A., Dew, I., Fasulo, D., Flanigan, M., Florea, L., Halpern, A., Hannenhalli, S., Kravitz, S., Levy, S., Mobarry, C., Reinert, K., Remington, K., Abu-Threideh, J., Beasley, E., Biddick, K., Bonazzi, V., Brandon, R., Cargill, M., Chandramouliswaran, I., Charlab, R., Chaturvedi, K., Deng, Z., Di Francesco, V., Dunn, P., Eilbeck, K., Evangelista, C., Gabrielian, A. E., Gan, W., Ge, W., Gong, F., Gu, Z., Guan, P., Heiman, T. J., Higgins, M. E., Ji, R. R., Ke, Z., Ketchum, K. A., Lai, Z., Lei, Y., Li, Z., Li, J., Liang, Y., Lin, X., Lu, F., Merkulov, G. V., Milshina, N., Moore, H. M., Naik, A. K., Narayan, V. A., Neelam, B., Nusskern, D., Rusch, D. B., Salzberg, S., Shao, W., Shue, B., Sun, J., Wang, Z., Wang, A., Wang, X., Wang, J., Wei, M., Wides, R., Xiao, C., Yan, C., Yao, A., Ye, J., Zhan, M., Zhang, W., Zhang, H., Zhao, Q., Zheng, L., Zhong, F., Zhong, W., Zhu, S., Zhao, S., Gilbert, D.,

Baumhueter, S., Spier, G., Carter, C., Cravchik, A., Woodage, T., Ali, F., An, H., Awe, A., Baldwin, D., Baden, H., Barnstead, M., Barrow, I., Beeson, K., Busam, D., Carver, A., Center, A., Cheng, M. L., Curry, L., Danaher, S., Davenport, L., Desilets, R., Dietz, S., Dodson, K., Doup, L., Ferriera, S., Garg, N., Gluecksmann, A., Hart, B., Haynes, J., Haynes, C., Heiner, C., Hladun, S., Hostin, D., Houck, J., Howland, T., Ibegwam, C., Johnson, J., Kalush, F., Kline, L., Koduru, S., Love, A., Mann, F., May, D., McCawley, S., McIntosh, T., McMullen, I., Moy, M., Moy, L., Murphy, B., Nelson, K., Pfannkoch, C., Pratts, E., Puri, V., Qureshi, H., Reardon, M., Rodriguez, R., Rogers, Y. H., Romblad, D., Ruhfel, B., Scott, R., Sitter, C., Smallwood, M., Stewart, E., Strong, R., Suh, E., Thomas, R., Tint, N. N., Tse, S., Vech, C., Wang, G., Wetter, J., Williams, S., Williams, M., Windsor, S., Winn-Deen, E., Wolfe, K., Zaveri, J., Zaveri, K., Abril, J. F., Guigó, R., Campbell, M. J., Sjolander, K. V., Karlak, B., Kejariwal, A., Mi, H., Lazareva, B., Hatton, T., Narechania, A., Diemer, K., Muruganujan, A., Guo, N., Sato, S., Bafna, V., Istrail, S., Lippert, R., Schwartz, R., Walenz, B., Yooseph, S., Allen, D., Basu, A., Baxendale, J., Blick, L., Caminha, M., Carnes-Stine, J., Caulk, P., Chiang, Y. H., Coyne, M., Dahlke, C., Mays, A., Dombroski, M., Donnelly, M., Ely, D., Esparham, S., Fosler, C., Gire, H., Glanowski, S., Glasser, K., Glodek, A., Gorokhov, M., Graham, K., Gropman, B., Harris, M., Heil, J., Henderson, S., Hoover, J., Jennings, D., Jordan, C., Jordan, J., Kasha, J., Kagan, L., Kraft, C., Levitsky, A., Lewis, M., Liu, X., Lopez, J., Ma, D., Majoros, W., McDaniel, J., Murphy, S., Newman, M., Nguyen, T., Nguyen, N., Nodell, M., Pan, S., Peck, J., Peterson, M., Rowe, W., Sanders, R., Scott, J., Simpson, M., Smith, T., Sprague, A., Stockwell, T., Turner, R., Venter, E., Wang, M., Wen, M., Wu, D., Wu, M., Xia, A., Zandieh, A. & Zhu, X. The sequence of the human genome. *Science* **291,** 1304–1351 (2001).

56. International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. *Nature* **431,** 931–945 (2004).
57. 1000 Genomes Project Consortium, Auton, A., Brooks, L. D., Durbin, R. M., Garrison, E. P., Kang, H. M., Korbel, J. O., Marchini, J. L., McCarthy, S., McVean, G. A. & Abecasis, G. R. A global reference for human genetic variation. *Nature* **526,** 68–74 (2015).
58. Lek, M., Karczewski, K. J., Minikel, E. V., Samocha, K. E., Banks, E., Fennell, T., O'Donnell-Luria, A. H., Ware, J. S., Hill, A. J., Cummings, B. B., Tukiainen, T., Birnbaum, D. P., Kosmicki, J. A., Duncan, L. E., Estrada, K., Zhao, F., Zou, J., Pierce-Hoffman, E., Berghout, J., Cooper, D. N., Deflaux, N., DePristo, M., Do, R., Flannick, J., Fromer, M., Gauthier, L., Goldstein, J., Gupta, N., Howrigan, D., Kiezun, A., Kurki, M. I., Moonshine, A. L., Natarajan, P., Orozco, L., Peloso, G. M., Poplin, R., Rivas, M. A., Ruano-Rubio, V., Rose, S. A., Ruderfer, D. M., Shakir, K., Stenson, P. D., Stevens, C., Thomas, B. P., Tiao, G., Tusie-Luna, M. T., Weisburd, B., Won, H.-H., Yu, D., Altshuler, D. M., Ardissino, D., Boehnke, M., Danesh, J., Donnelly, S., Elosua, R., Florez, J. C., Gabriel, S. B., Getz, G., Glatt, S. J., Hultman, C. M., Kathiresan, S., Laakso, M., McCarroll, S., McCarthy, M. I., McGovern, D., McPherson, R., Neale, B. M., Palotie, A., Purcell, S. M., Saleheen, D., Scharf, J. M., Sklar, P., Sullivan, P. F., Tuomilehto, J., Tsuang, M. T., Watkins, H. C., Wilson, J. G., Daly, M. J., MacArthur, D. G. & Exome Aggregation Consortium. Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536,** 285–291 (2016).
59. Karczewski, K. J., Francioli, L. C., Tiao, G., Cummings, B. B., Alföldi, J., Wang, Q., Collins, R. L., Laricchia, K. M., Ganna, A., Birnbaum, D. P., Gauthier, L. D., Brand, H., Solomonson, M., Watts, N. A., Rhodes, D., Singer-Berk, M., England, E. M., Seaby, E. G., Kosmicki, J. A., Walters, R. K., Tashman, K., Farjoun, Y., Banks, E., Poterba, T., Wang, A., Seed, C., Whiffin, N., Chong, J. X., Samocha, K. E., Pierce-Hoffman, E., Zappala, Z., O'Donnell-Luria, A. H., Minikel, E. V., Weisburd, B., Lek, M., Ware, J. S., Vittal, C., Armean, I. M., Bergelson, L., Cibulskis, K., Connolly, K. M., Covarrubias, M., Donnelly, S., Ferriera, S., Gabriel, S., Gentry, J., Gupta, N., Jeandet, T., Kaplan, D., Llanwarne, C., Munshi, R., Novod, S., Petrillo, N., Roazen, D., Ruano-Rubio, V., Saltzman, A., Schleicher, M., Soto, J., Tibbetts, K., Tolonen, C., Wade, G., Talkowski, M. E., Genome Aggregation Database Consortium, Neale, B. M., Daly, M. J. & MacArthur, D. G. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **581,** 434–443 (2020).
60. Lawrence, M. S., Stojanov, P., Mermel, C. H., Robinson, J. T., Garraway, L. A., Golub, T. R., Meyerson, M., Gabriel, S. B., Lander, E. S. & Getz, G. Discovery and saturation analysis of cancer genes across 21 tumour types. *Nature* **505,** 495–501 (2014).
61. Hutter, C. & Zenklusen, J. C. The Cancer Genome Atlas: Creating Lasting Value beyond Its Data. *Cell* **173,** 283–285 (2018).
62. ICGC/TCGA Pan-Cancer Analysis of Whole Genomes Consortium. Pan-cancer analysis of whole genomes. *Nature* **578,** 82–93 (2020).
63. Alexandrov, L. B., Nik-Zainal, S., Wedge, D. C., Aparicio, S. A. J. R., Behjati, S., Biankin, A. V.,

Bignell, G. R., Bolli, N., Borg, A., Børresen-Dale, A.-L., Boyault, S., Burkhardt, B., Butler, A. P., Caldas, C., Davies, H. R., Desmedt, C., Eils, R., Eyfjörd, J. E., Foekens, J. A., Greaves, M., Hosoda, F., Hutter, B., Ilicic, T., Imbeaud, S., Imielinski, M., Jäger, N., Jones, D. T. W., Jones, D., Knappskog, S., Kool, M., Lakhani, S. R., López-Otín, C., Martin, S., Munshi, N. C., Nakamura, H., Northcott, P. A., Pajic, M., Papaemmanuil, E., Paradiso, A., Pearson, J. V., Puente, X. S., Raine, K., Ramakrishna, M., Richardson, A. L., Richter, J., Rosenstiel, P., Schlesner, M., Schumacher, T. N., Span, P. N., Teague, J. W., Totoki, Y., Tutt, A. N. J., Valdés-Mas, R., van Buuren, M. M., van 't Veer, L., Vincent-Salomon, A., Waddell, N., Yates, L. R., Australian Pancreatic Cancer Genome Initiative, ICGC Breast Cancer Consortium, ICGC MMML-Seq Consortium, ICGC PedBrain, Zucman-Rossi, J., Futreal, P. A., McDermott, U., Lichter, P., Meyerson, M., Grimmond, S. M., Siebert, R., Campo, E., Shibata, T., Pfister, S. M., Campbell, P. J. & Stratton, M. R. Signatures of mutational processes in human cancer. *Nature* **500,** 415–421 (2013).

64. Tomasetti, C., Vogelstein, B. & Parmigiani, G. Half or more of the somatic mutations in cancers of self-renewing tissues originate prior to tumor initiation. *Proc. Natl. Acad. Sci. U. S. A.* **110,** 1999–2004 (2013).
65. Govindan, R., Ding, L., Griffith, M., Subramanian, J., Dees, N. D., Kanchi, K. L., Maher, C. A., Fulton, R., Fulton, L., Wallis, J., Chen, K., Walker, J., McDonald, S., Bose, R., Ornitz, D., Xiong, D., You, M., Dooling, D. J., Watson, M., Mardis, E. R. & Wilson, R. K. Genomic landscape of non-small cell lung cancer in smokers and never-smokers. *Cell* **150,** 1121–1134 (2012).
66. Yang, J. & Adli, M. Mapping and Making Sense of Noncoding Mutations in the Genome. *Cancer Res.* **79,** 4309-4314 (2019).
67. Lawrence, M. S., Stojanov, P., Polak, P., Kryukov, G. V., Cibulskis, K., Sivachenko, A., Carter, S. L., Stewart, C., Mermel, C. H., Roberts, S. A., Kiezun, A., Hammerman, P. S., McKenna, A., Drier, Y., Zou, L., Ramos, A. H., Pugh, T. J., Stransky, N., Helman, E., Kim, J., Sougnez, C., Ambrogio, L., Nickerson, E., Shefler, E., Cortés, M. L., Auclair, D., Saksena, G., Voet, D., Noble, M., DiCara, D., Lin, P., Lichtenstein, L., Heiman, D. I., Fennell, T., Imielinski, M., Hernandez, B., Hodis, E., Baca, S., Dulak, A. M., Lohr, J., Landau, D.-A., Wu, C. J., Melendez-Zajgla, J., Hidalgo-Miranda, A., Koren, A., McCarroll, S. A., Mora, J., Crompton, B., Onofrio, R., Parkin, M., Winckler, W., Ardlie, K., Gabriel, S. B., Roberts, C. W. M., Biegel, J. A., Stegmaier, K., Bass, A. J., Garraway, L. A., Meyerson, M., Golub, T. R., Gordenin, D. A., Sunyaev, S., Lander, E. S. & Getz, G. Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* **499,** 214–218 (2013).
68. Wood, L. D., Parsons, D. W., Jones, S., Lin, J., Sjöblom, T., Leary, R. J., Shen, D., Boca, S. M., Barber, T., Ptak, J., Silliman, N., Szabo, S., Dezso, Z., Ustyanksky, V., Nikolskaya, T., Nikolsky, Y., Karchin, R., Wilson, P. A., Kaminker, J. S., Zhang, Z., Croshaw, R., Willis, J., Dawson, D., Shipitsin, M., Willson, J. K. V., Sukumar, S., Polyak, K., Park, B. H., Pethiyagoda, C. L., Pant, P. V. K., Ballinger, D. G., Sparks, A. B., Hartigan, J., Smith, D. R., Suh, E., Papadopoulos, N., Buckhaults, P., Markowitz, S. D., Parmigiani, G., Kinzler, K. W., Velculescu, V. E. & Vogelstein, B. The genomic landscapes of human breast and colorectal cancers. *Science* **318,** 1108–1113 (2007).
69. Raphael, B. J., Dobson, J. R., Oesper, L. & Vandin, F. Identifying driver mutations in sequenced cancer genomes: computational approaches to enable precision medicine. *Genome Med.* **6,** 5 (2014).
70. Siegel, R. L., Miller, K. D. & Jemal, A. Cancer statistics, 2019. *CA Cancer J. Clin.* **69,** 7–34 (2019).
71. Bilimoria, K. Y., Bentrem, D. J., Wayne, J. D., Ko, C. Y., Bennett, C. L. & Talamonti, M. S. Small bowel cancer in the United States: changes in epidemiology, treatment, and survival over the last 20 years. *Ann. Surg.* **249,** 63–71 (2009).
72. Baran, B., Mert Ozupek, N., Yerli Tetik, N., Acar, E., Bekcioglu, O. & Baskin, Y. Difference Between Left-Sided and Right-Sided Colorectal Cancer: A Focused Review of Literature. *Gastroenterol. Res. Pract.* **11,** 264–273 (2018).
73. Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A. & Jemal, A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: Cancer J. Clin.* **68,** 394–424 (2018).
74. Araghi, M., Soerjomataram, I., Jenkins, M., Brierley, J., Morris, E., Bray, F. & Arnold, M. Global trends in colorectal cancer mortality: projections to the year 2035. *Int. J. Cancer* **144,** 2992–3000 (2019).
75. Song, M., Garrett, W. S. & Chan, A. T. Nutrients, Foods, and Colorectal Cancer Prevention. *Gastroenterology* **148,** 1244–1260.e16 (2015).
76. Chia, W. K., Ali, R. & Toh, H. C. Aspirin as adjuvant therapy for colorectal cancer—reinterpreting paradigms. *Nat Rev Clin Oncol.* **9,** 561–570 (2012).

77. Burn, J., Sheth, H., Elliott, F., Reed, L., Macrae, F., Mecklin, J.-P., Möslein, G., McRonald, F. E., Bertario, L., Evans, D. G., Gerdes, A.-M., Ho, J. W. C., Lindblom, A., Morrison, P. J., Rashbass, J., Ramesar, R., Seppälä, T., Thomas, H. J. W., Pylvänäinen, K., Borthwick, G. M., Mathers, J. C., Bishop, D. T., Boussioutas, A., Brewer, C., Cook, J., Eccles, D., Ellis, A., Hodgson, S. V., Lubinski, J., Maher, E. R., Porteous, M. E. M., Sampson, J., Scott, R. J. & Side, L. Cancer prevention with aspirin in hereditary colorectal cancer (Lynch syndrome), 10-year follow-up and registry-based 20-year data in the CAPP2 study: a double-blind, randomised, placebo-controlled trial. *Lancet* **395,** 1855–1863 (2020).
78. Mulder, S. A., Kranse, R., Damhuis, R. A., de Wilt, J. H. W., Th. Ouwendijk, R. J., Kuipers, E. J. & van Leerdam, M. E. Prevalence and prognosis of synchronous colorectal cancer: A Dutch population-based study. *Cancer Epidemiol.* **35,** 442–447 (2011).
79. Lam, A. K.-Y., Carmichael, R., Buettner, P. G., Gopalan, V., Ho, Y.-H. & Siu, S. Clinicopathological significance of synchronous carcinoma in colorectal cancer. *Am. J. Surg.* **202,** 39–44 (2011).
80. Latournerie, M., Jooste, V., Cottet, V., Lepage, C., Faivre, J. & Bouvier, A.-M. Epidemiology and prognosis of synchronous colorectal cancers. *Br. J. Surg.* **95,** 1528–1533 (2008).
81. Nosho, K., Kure, S., Irahara, N., Shima, K., Baba, Y., Spiegelman, D., Meyerhardt, J. A., Giovannucci, E. L., Fuchs, C. S. & Ogino, S. A prospective cohort study shows unique epigenetic, genetic, and prognostic features of synchronous colorectal cancers. *Gastroenterology* **137,** 1609–20.e1–3 (2009).
82. Giovannucci, E. & Ogino, S. DNA methylation, field effects, and colorectal cancer. *J. Natl. Cancer Inst.* **97,** 1317–1319 (2005).
83. Hu, H., Chang, D. T., Nikiforova, M. N., Kuan, S.-F. & Pai, R. K. Clinicopathologic features of synchronous colorectal carcinoma: A distinct subset arising from multiple sessile serrated adenomas and associated with high levels of microsatellite instability and favorable prognosis. *Am. J. Surg. Pathol.* **37,** 1660–1670 (2013).
84. Bennett, C. M., Coleman, H. G., Veal, P. G., Cantwell, M. M., Lau, C. C. L. & Murray, L. J. Lifestyle factors and small intestine adenocarcinoma risk: A systematic review and meta-analysis. *Cancer Epidemiol.* **39,** 265–273 (2015).
85. Raghav, K. & Overman, M. J. Small bowel adenocarcinomas--existing evidence and evolving paradigms. *Nat. Rev. Clin. Oncol.* **10,** 534–544 (2013).
86. Amin, M. B., Edge, S. B., Greene, F. L., Byrd, D. R., Brookland, R. K., Washington, M. K., Gershenwald, J. E., Compton, C. C., Hess, K. R., Sullivan, D. C., Milburn Jessup, J., Brierley, J. D., Gaspar, L. E., Schilsky, R. L., Balch, C. M., Winchester, D. P., Asare, E. A., Madera, M., Gress, D. M. & Meyer, L. R. *AJCC Cancer Staging Manual*. (Springer, 2018).
87. Galon, J., Costes, A., Sanchez-Cabo, F., Kirilovsky, A., Mlecnik, B., Lagorce-Pagès, C., Tosolini, M., Camus, M., Berger, A., Wind, P., Zinzindohoué, F., Bruneval, P., Cugnenc, P.-H., Trajanoski, Z., Fridman, W.-H. & Pagès, F. Type, density, and location of immune cells within human colorectal tumors predict clinical outcome. *Science* **313,** 1960–1964 (2006).
88. Galon, J., Mlecnik, B., Bindea, G., Angell, H. K., Berger, A., Lagorce, C., Lugli, A., Zlobec, I., Hartmann, A., Bifulco, C., Nagtegaal, I. D., Palmqvist, R., Masucci, G. V., Botti, G., Tatangelo, F., Delrio, P., Maio, M., Laghi, L., Grizzi, F., Asslaber, M., D'Arrigo, C., Vidal-Vanaclocha, F., Zavadova, E., Chouchane, L., Ohashi, P. S., Hafezi-Bakhtiari, S., Wouters, B. G., Roehrl, M., Nguyen, L., Kawakami, Y., Hazama, S., Okuno, K., Ogino, S., Gibbs, P., Waring, P., Sato, N., Torigoe, T., Itoh, K., Patel, P. S., Shukla, S. N., Wang, Y., Kopetz, S., Sinicrope, F. A., Scripcariu, V., Ascierto, P. A., Marincola, F. M., Fox, B. A. & Pagès, F. Towards the introduction of the 'Immunoscore' in the classification of malignant tumours. *J.Pathol.* **232,** 199–209 (2014).
89. Pagès, F., Mlecnik, B., Marliot, F., Bindea, G., Ou, F.-S., Bifulco, C., Lugli, A., Zlobec, I., Rau, T. T., Berger, M. D., Nagtegaal, I. D., Vink-Börger, E., Hartmann, A., Geppert, C., Kolwelter, J., Merkel, S., Grützmann, R., Van den Eynde, M., Jouret-Mourin, A., Kartheuser, A., Léonard, D., Remue, C., Wang, J. Y., Bavi, P., Roehrl, M. H. A., Ohashi, P. S., Nguyen, L. T., Han, S., MacGregor, H. L., Hafezi-Bakhtiari, S., Wouters, B. G., Masucci, G. V., Andersson, E. K., Zavadova, E., Vocka, M., Spacek, J., Petruzelka, L., Konopasek, B., Dundr, P., Skalova, H., Nemejcova, K., Botti, G., Tatangelo, F., Delrio, P., Ciliberto, G., Maio, M., Laghi, L., Grizzi, F., Fredriksen, T., Buttard, B., Angelova, M., Vasaturo, A., Maby, P., Church, S. E., Angell, H. K., Lafontaine, L., Bruni, D., El Sissy, C., Haicheur, N., Kirilovsky, A., Berger, A., Lagorce, C., Meyers, J. P., Paustian, C., Feng, Z., Ballesteros-Merino, C., Dijkstra, J., van de Water, C., van Lent-van Vliet, S., Knijn, N., Mușină, A.-M., Scripcariu, D.-V., Popivanova, B., Xu, M., Fujita, T., Hazama, S., Suzuki, N., Nagano, H., Okuno, K., Torigoe, T., Sato, N., Furuhata, T., Takemasa, I., Itoh, K., Patel, P. S., Vora, H. H., Shah, B., Patel, J. B., Rajvik, K. N., Pandya, S. J., Shukla, S. N., Wang, Y., Zhang, G., Kawakami, Y., Marincola, F. M., Ascierto, P. A.,

Sargent, D. J., Fox, B. A. & Galon, J. International validation of the consensus Immunoscore for the classification of colon cancer: a prognostic and accuracy study. *Lancet* **391,** 2128–2139 (2018).
90. Van Cutsem, E., Cervantes, A., Adam, R., Sobrero, A., Van Krieken, J. H., Aderka, D., Aranda Aguilar, E., Bardelli, A., Benson, A., Bodoky, G., Ciardiello, F., D'Hoore, A., Diaz-Rubio, E., Douillard, J.-Y., Ducreux, M., Falcone, A., Grothey, A., Gruenberger, T., Haustermans, K., Heinemann, V., Hoff, P., Köhne, C.-H., Labianca, R., Laurent-Puig, P., Ma, B., Maughan, T., Muro, K., Normanno, N., Österlund, P., Oyen, W. J. G., Papamichael, D., Pentheroudakis, G., Pfeiffer, P., Price, T. J., Punt, C., Ricke, J., Roth, A., Salazar, R., Scheithauer, W., Schmoll, H. J., Tabernero, J., Taïeb, J., Tejpar, S., Wasan, H., Yoshino, T., Zaanan, A. & Arnold, D. ESMO consensus guidelines for the management of patients with metastatic colorectal cancer. *Ann. Oncol.* **27,** 1386–1422 (2016).
91. Pedersen, K. S., Raghav, K. & Overman, M. J. Small Bowel Adenocarcinoma: Etiology, Presentation, and Molecular Alterations. *J. Natl. Compr. Canc. Netw.* **17,** 1135–1141 (2019).
92. Le, D. T., Durham, J. N., Smith, K. N., Wang, H., Bartlett, B. R., Aulakh, L. K., Lu, S., Kemberling, H., Wilt, C., Luber, B. S., Wong, F., Azad, N. S., Rucki, A. A., Laheru, D., Donehower, R., Zaheer, A., Fisher, G. A., Crocenzi, T. S., Lee, J. J., Greten, T. F., Duffy, A. G., Ciombor, K. K., Eyring, A. D., Lam, B. H., Joe, A., Kang, S. P., Holdhoff, M., Danilova, L., Cope, L., Meyer, C., Zhou, S., Goldberg, R. M., Armstrong, D. K., Bever, K. M., Fader, A. N., Taube, J., Housseau, F., Spetzler, D., Xiao, N., Pardoll, D. M., Papadopoulos, N., Kinzler, K. W., Eshleman, J. R., Vogelstein, B., Anders, R. A. & Diaz, L. A., Jr. Mismatch repair deficiency predicts response of solid tumors to PD-1 blockade. *Science* **357,** 409–413 (2017).
93. Sartore-Bianchi, A., Trusolino, L., Martino, C., Bencardino, K., Lonardi, S., Bergamo, F., Zagonel, V., Leone, F., Depetris, I., Martinelli, E., Troiani, T., Ciardiello, F., Racca, P., Bertotti, A., Siravegna, G., Torri, V., Amatu, A., Ghezzi, S., Marrapese, G., Palmeri, L., Valtorta, E., Cassingena, A., Lauricella, C., Vanzulli, A., Regge, D., Veronese, S., Comoglio, P. M., Bardelli, A., Marsoni, S. & Siena, S. Dual-targeted therapy with trastuzumab and lapatinib in treatment-refractory, KRAS codon 12/13 wild-type, HER2-positive metastatic colorectal cancer (HERACLES): a proof-of-concept, multicentre, open-label, phase 2 trial. *Lancet Oncol.* **17,** 738–746 (2016).
94. Humphries, A. & Wright, N. A. Colonic crypt organization and tumorigenesis. *Nat. Rev. Cancer* **8,** 415–424 (2008).
95. Clevers, H. The intestinal crypt, a prototype stem cell compartment. *Cell* **154,** 274–284 (2013).
96. Kim, J. Y., Siegmund, K. D., Tavaré, S. & Shibata, D. Age-related human small intestine methylation: evidence for stem cell niches. *BMC Med.* **3,** 10 (2005).
97. Fevr, T., Robine, S., Louvard, D. & Huelsken, J. Wnt/b-Catenin Is Essential for Intestinal Homeostasis and Maintenance of Intestinal Stem Cells. *Mol. Cell. Biol.* **27,** 7551–7559 (2007).
98. Barker, N., Ridgway, R. A., van Es, J. H., van de Wetering, M., Begthel, H., van den Born, M., Danenberg, E., Clarke, A. R., Sansom, O. J. & Clevers, H. Crypt stem cells as the cells-of-origin of intestinal cancer. *Nature* **457,** 608–611 (2009).
99. Sellner, F. Investigations on the significance of the adenoma-carcinoma sequence in the small bowel. *Cancer* **66,** 702–715 (1990).
100. Fearon, E. R. Molecular genetics of colorectal cancer. *Annu. Rev. Pathol.* **6,** 479–507 (2011).
101. Aparicio, T., Zaanan, A., Mary, F., Afchain, P., Manfredi, S. & Evans, T. R. J. Small Bowel Adenocarcinoma. *Gastroenterol. Clin. North Am.* **45,** 447–457 (2016).
102. Rajagopalan, H., Nowak, M. A., Vogelstein, B. & Lengauer, C. The significance of unstable chromosomes in colorectal cancer. *Nat. Rev. Cancer* **3,** 695–701 (2003).
103. Jones, S., Chen, W., Parmigiani, G., Diehl, F., Beerenwinkel, N., Antal, T., Traulsen, A., Nowak, M. A., Siegel, C., Velculescu, V. E., Kinzler, K. W., Vogelstein, B., Willis, J. & Markowitz, S. D. Comparative lesion sequencing provides insights into tumor evolution. *Proc. Natl. Acad. Sci. U. S. A.* **105,** 4283–4288 (2008).
104. Riihimäki, M., Hemminki, A., Sundquist, J. & Hemminki, K. Patterns of metastasis in colon and rectal cancer. *Sci. Rep.* **6,** 29765 (2016).
105. Legué, L. M., Bernards, N., Gerritse, S. L., van Oudheusden, T. R., de Hingh, I. H. J. T., Creemers, G.-J. M., Ten Tije, A. J. & Lemmens, V. E. P. P. Trends in incidence, treatment and survival of small bowel adenocarcinomas between 1999 and 2013: a population-based study in The Netherlands. *Acta Oncol.* **55,** 1183–1189 (2016).
106. Goelz, S. E., Vogelstein, B., Hamilton, S. R. & Feinberg, A. P. Hypomethylation of DNA from benign and malignant human colon neoplasms. *Science* **228,** 187–190 (1985).
107. Kim, M. S., Lee, J. & Sidransky, D. DNA methylation markers in colorectal cancer. *Cancer Metastasis*

Rev. **29,** 181–206 (2010).
108. Issa, J.-P. CpG island methylator phenotype in cancer. *Nat. Rev. Cancer* **4,** 988–993 (2004).
109. Aparicio, T., Zaanan, A., Svrcek, M., Laurent-Puig, P., Carrere, N., Manfredi, S., Locher, C. & Afchain, P. Small bowel adenocarcinoma: epidemiology, risk factors, diagnosis and treatment. *Dig. Liver Dis.* **46,** 97–104 (2014).
110. Vogelstein, B., Fearon, E. R., Hamilton, S. R., Kern, S. E., Preisinger, A. C., Leppert, M., Nakamura, Y., White, R., Smits, A. M. & Bos, J. L. Genetic alterations during colorectal-tumor development. *N. Engl. J. Med.* **319,** 525–532 (1988).
111. Grady, W. M. Genomic instability and colon cancer. *Cancer Metastasis Rev.* **23,** 11–27 (2004).
112. Kinzler, K. W. & Vogelstein, B. Lessons from hereditary colorectal cancer. *Cell* **87,** 159–170 (1996).
113. He, T. Identification of c-MYC as a Target of the APC Pathway. *Science* **281,** 1509–1512 (1998).
114. Seshagiri, S., Stawiski, E. W., Durinck, S., Modrusan, Z., Storm, E. E., Conboy, C. B., Chaudhuri, S., Guan, Y., Janakiraman, V., Jaiswal, B. S., Guillory, J., Ha, C., Dijkgraaf, G. J. P., Stinson, J., Gnad, F., Huntley, M. A., Degenhardt, J. D., Haverty, P. M., Bourgon, R., Wang, W., Koeppen, H., Gentleman, R., Starr, T. K., Zhang, Z., Largaespada, D. A., Wu, T. D. & de Sauvage, F. J. Recurrent R-spondin fusions in colon cancer. *Nature* **488,** 660–664 (2012).
115. Schrock, A. B., Devoe, C. E., McWilliams, R., Sun, J., Aparicio, T., Stephens, P. J., Ross, J. S., Wilson, R., Miller, V. A., Ali, S. M. & Overman, M. J. Genomic Profiling of Small-Bowel Adenocarcinoma. *JAMA Oncol* **3,** 1546–1553 (2017).
116. Deng, G. BRAF Mutation Is Frequently Present in Sporadic Colorectal Cancer with Methylated hMLH1, But Not in Hereditary Nonpolyposis Colorectal Cancer. *Clin. Cancer Res.* **10,** 191–195 (2004).
117. Fang, J. Y. & Richardson, B. C. The MAPK signalling pathways and colorectal cancer. *Lancet Oncol.* **6,** 322–327 (2005).
118. Jones, J. C., Renfro, L. A., Al-Shamsi, H. O., Schrock, A. B., Rankin, A., Zhang, B. Y., Kasi, P. M., Voss, J. S., Leal, A. D., Sun, J., Ross, J., Ali, S. M., Hubbard, J. M., Kipp, B. R., McWilliams, R. R., Kopetz, S., Wolff, R. A. & Grothey, A. BRAF Mutations Define a Clinically Distinct Molecular Subtype of Metastatic Colorectal Cancer. *J. Clin. Oncol.* **35,** 2624–2630 (2017).
119. Shimada, Y., Tajima, Y., Nagahashi, M., Ichikawa, H., Oyanagi, H., Okuda, S., Takabe, K. & Wakai, T. Clinical Significance of BRAF Non-V600E Mutations in Colorectal Cancer: A Retrospective Study of Two Institutions. *J. Surg. Res.* **232,** 72–81 (2018).
120. Alvi, M. A., McArt, D. G., Kelly, P., Fuchs, M.-A., Alderdice, M., McCabe, C. M., Bingham, V., McGready, C., Tripathi, S., Emmert-Streib, F., Loughrey, M. B., McQuaid, S., Maxwell, P., Hamilton, P. W., Turkington, R., James, J. A., Wilson, R. H. & Salto-Tellez, M. Comprehensive molecular pathology analysis of small bowel adenocarcinoma reveals novel targets with potential for clinical utility. *Oncotarget* **6,** 20863–20874 (2015).
121. Misale, S., Yaeger, R., Hobor, S., Scala, E., Janakiraman, M., Liska, D., Valtorta, E., Schiavo, R., Buscarino, M., Siravegna, G., Bencardino, K., Cercek, A., Chen, C.-T., Veronese, S., Zanon, C., Sartore-Bianchi, A., Gambacorta, M., Gallicchio, M., Vakiani, E., Boscaro, V., Medico, E., Weiser, M., Siena, S., Di Nicolantonio, F., Solit, D. & Bardelli, A. Emergence of KRAS mutations and acquired resistance to anti-EGFR therapy in colorectal cancer. *Nature* **486,** 532–536 (2012).
122. De Roock, W., Claes, B., Bernasconi, D., De Schutter, J., Biesmans, B., Fountzilas, G., Kalogeras, K. T., Kotoula, V., Papamichael, D., Laurent-Puig, P., Penault-Llorca, F., Rougier, P., Vincenzi, B., Santini, D., Tonini, G., Cappuzzo, F., Frattini, M., Molinari, F., Saletti, P., De Dosso, S., Martini, M., Bardelli, A., Siena, S., Sartore-Bianchi, A., Tabernero, J., Macarulla, T., Di Fiore, F., Gangloff, A. O., Ciardiello, F., Pfeiffer, P., Qvortrup, C., Hansen, T. P., Van Cutsem, E., Piessevaux, H., Lambrechts, D., Delorenzi, M. & Tejpar, S. 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. *Lancet Oncol.* **11,** 753–762 (2010).
123. Laurent-Puig, P., Cayre, A., Manceau, G., Buc, E., Bachet, J.-B., Lecomte, T., Rougier, P., Lievre, A., Landi, B., Boige, V., Ducreux, M., Ychou, M., Bibeau, F., Bouché, O., Reid, J., Stone, S. & Penault-Llorca, F. Analysis of PTEN, BRAF, and EGFR status in determining benefit from cetuximab therapy in wild-type KRAS metastatic colon cancer. *J. Clin. Oncol.* **27,** 5924–5930 (2009).
124. Baker, S. J., Preisinger, A. C., Jessup, J. M., Paraskeva, C., Markowitz, S., Willson, J. K., Hamilton, S. & Vogelstein, B. p53 gene mutations occur in combination with 17p allelic deletions as late events in colorectal tumorigenesis. *Cancer Res.* **50,** 7717–7722 (1990).
125. Baker, S., Fearon, E., Nigro, J., Hamilton, Preisinger, A., Jessup, J., vanTuinen, P., Ledbetter, D.,

Barker, D., Nakamura, Y., White, R. & Vogelstein, B. Chromosome 17 deletions and p53 gene mutations in colorectal carcinomas. *Science* **244,** 217–221 (1989).
126. Cahill, D. P., Lengauer, C., Yu, J., Riggins, G. J., Willson, J. K. V., Markowitz, S. D., Kinzler, K. W. & Vogelstein, B. Mutations of mitotic checkpoint genes in human cancers. *Nature* **392,** 300–303 (1998).
127. Michel, L. S., Liberal, V., Chatterjee, A., Kirchwegger, R., Pasche, B., Gerald, W., Dobles, M., Sorger, P. K., Murty, V. V. & Benezra, R. MAD2 haplo-insufficiency causes premature anaphase and chromosome instability in mammalian cells. *Nature* **409,** 355–359 (2001).
128. Fodde, R., Kuipers, J., Rosenberg, C., Smits, R., Kielman, M., Gaspar, C., van Es, J. H., Breukel, C., Wiegant, J., Giles, R. H. & Clevers, H. Mutations in the APC tumour suppressor gene cause chromosomal instability. *Nat. Cell Biol.* **3,** 433–438 (2001).
129. Aaltonen, L. A., Peltomäki, P., Leach, F. S., Sistonen, P., Pylkkänen, L., Mecklin, J. P., Järvinen, H., Powell, S. M., Jen, J. & Hamilton, S. R. Clues to the pathogenesis of familial colorectal cancer. *Science* **260,** 812–816 (1993).
130. Ionov, Y., Peinado, M. A., Malkhosyan, S., Shibata, D. & Perucho, M. Ubiquitous somatic mutations in simple repeated sequences reveal a new mechanism for colonic carcinogenesis. *Nature* **363,** 558–561 (1993).
131. Thibodeau, S. N., Bren, G. & Schaid, D. Microsatellite instability in cancer of the proximal colon. *Science* **260,** 816–819 (1993).
132. Kane, M. F., Loda, M., Gaida, G. M., Lipman, J., Mishra, R., Goldman, H., Jessup, J. M. & Kolodner, R. Methylation of the hMLH1 promoter correlates with lack of expression of hMLH1 in sporadic colon tumors and mismatch repair-defective human tumor cell lines. *Cancer Res.* **57,** 808–811 (1997).
133. Carethers, J. M., Smith, E. J., Behling, C. A., Nguyen, L., Tajima, A., Doctolero, R. T., Cabrera, B. L., Goel, A., Arnold, C. A., Miyai, K. & Boland, C. R. Use of 5-fluorouracil and survival in patients with microsatellite-unstable colorectal cancer. *Gastroenterology* **126,** 394–401 (2004).
134. Malesci, A., Laghi, L., Bianchi, P., Delconte, G., Randolph, A., Torri, V., Carnaghi, C., Doci, R., Rosati, R., Montorsi, M., Roncalli, M., Gennari, L. & Santoro, A. Reduced likelihood of metastases in patients with microsatellite-unstable colorectal cancer. *Clin. Cancer Res.* **13,** 3831–3839 (2007).
135. Kakar, S., Burgart, L. J., Thibodeau, S. N., Rabe, K. G., Petersen, G. M., Goldberg, R. M. & Lindor, N. M. Frequency of loss of hMLH1 expression in colorectal carcinoma increases with advancing age. *Cancer* **97,** 1421–1427 (2003).
136. Jiricny, J. The multifaceted mismatch-repair system. *Nat. Rev. Mol. Cell Biol.* **7,** 335–346 (2006).
137. Parsons, R., Li, G. M., Longley, M. J., Fang, W. H., Papadopoulos, N., Jen, J., de la Chapelle, A., Kinzler, K. W., Vogelstein, B. & Modrich, P. Hypermutability and mismatch repair deficiency in RER+ tumor cells. *Cell* **75,** 1227–1236 (1993).
138. Boland, C. R., Thibodeau, S. N., Hamilton, S. R., Sidransky, D., Eshleman, J. R., Burt, R. W., Meltzer, S. J., Rodriguez-Bigas, M. A., Fodde, R., Ranzani, G. N. & Srivastava, S. A National Cancer Institute Workshop on Microsatellite Instability for cancer detection and familial predisposition: development of international criteria for the determination of microsatellite instability in colorectal cancer. *Cancer Res.* **58,** 5248–5257 (1998).
139. Markowitz, S., Wang, J., Myeroff, L., Parsons, R., Sun, L., Lutterbaugh, J., Fan, R. S., Zborowska, E., Kinzler, K. W. & Vogelstein, B. Inactivation of the type II TGF-beta receptor in colon cancer cells with microsatellite instability. *Science* **268,** 1336–1338 (1995).
140. Jung, B., Doctolero, R. T., Tajima, A., Nguyen, A. K., Keku, T., Sandler, R. S. & Carethers, J. M. Loss of activin receptor type 2 protein expression in microsatellite unstable colon cancers. *Gastroenterology* **126,** 654–659 (2004).
141. Lin, E. I., Tseng, L.-H., Gocke, C. D., Reil, S., Le, D. T., Azad, N. S. & Eshleman, J. R. Mutational profiling of colorectal cancers with microsatellite instability. *Oncotarget* **6,** 42334–42344 (2015).
142. Umar, A., Boland, C. R., Terdiman, J. P., Syngal, S., de la Chapelle, A., Rüschoff, J., Fishel, R., Lindor, N. M., Burgart, L. J., Hamelin, R., Hamilton, S. R., Hiatt, R. A., Jass, J., Lindblom, A., Lynch, H. T., Peltomaki, P., Ramsey, S. D., Rodriguez-Bigas, M. A., Vasen, H. F. A., Hawk, E. T., Barrett, J. C., Freedman, A. N. & Srivastava, S. Revised Bethesda Guidelines for hereditary nonpolyposis colorectal cancer (Lynch syndrome) and microsatellite instability. *J. Natl. Cancer Inst.* **96,** 261–268 (2004).
143. Kim, T.-M., Laird, P. W. & Park, P. J. The landscape of microsatellite instability in colorectal and endometrial cancer genomes. *Cell* **155,** 858–868 (2013).
144. Zhang, L., Peng, Y. & Peng, G. Mismatch repair-based stratification for immune checkpoint blockade therapy. *Am. J. Cancer Res.* **8,** 1977–1988 (2018).

145. Salipante, S. J., Scroggins, S. M., Hampel, H. L., Turner, E. H. & Pritchard, C. C. Microsatellite instability detection by next generation sequencing. *Clin. Chem.* **60,** 1192–1199 (2014).
146. Hause, R. J., Pritchard, C. C., Shendure, J. & Salipante, S. J. Classification and characterization of microsatellite instability across 18 cancer types. *Nat. Med.* **22,** 1342–1350 (2016).
147. Stadler, Z. K., Battaglin, F., Middha, S., Hechtman, J. F., Tran, C., Cercek, A., Yaeger, R., Segal, N. H., Varghese, A. M., Reidy-Lagunes, D. L., Kemeny, N. E., Salo-Mullen, E. E., Ashraf, A., Weiser, M. R., Garcia-Aguilar, J., Robson, M. E., Offit, K., Arcila, M. E., Berger, M. F., Shia, J., Solit, D. B. & Saltz, L. B. Reliable Detection of Mismatch Repair Deficiency in Colorectal Cancers Using Mutational Load in Next-Generation Sequencing Panels. *J. Clin. Oncol.* **34,** 2141–2147 (2016).
148. Kandoth, C., McLellan, M. D., Vandin, F., Ye, K., Niu, B., Lu, C., Xie, M., Zhang, Q., McMichael, J. F., Wyczalkowski, M. A., Leiserson, M. D. M., Miller, C. A., Welch, J. S., Walter, M. J., Wendl, M. C., Ley, T. J., Wilson, R. K., Raphael, B. J. & Ding, L. Mutational landscape and significance across 12 major cancer types. *Nature* **502,** 333–339 (2013).
149. Katainen, R., Dave, K., Pitkänen, E., Palin, K., Kivioja, T., Välimäki, N., Gylfe, A. E., Ristolainen, H., Hänninen, U. A., Cajuso, T., Kondelin, J., Tanskanen, T., Mecklin, J.-P., Järvinen, H., Renkonen-Sinisalo, L., Lepistö, A., Kaasinen, E., Kilpivaara, O., Tuupanen, S., Enge, M., Taipale, J. & Aaltonen, L. A. CTCF/cohesin-binding sites are frequently mutated in cancer. *Nat. Genet.* **47,** 818–821 (2015).
150. Orlando, G., Law, P. J., Cornish, A. J., Dobbins, S. E., Chubb, D., Broderick, P., Litchfield, K., Hariri, F., Pastinen, T., Osborne, C. S., Taipale, J. & Houlston, R. S. Promoter capture Hi-C-based identification of recurrent noncoding mutations in colorectal cancer. *Nat. Genet.* **50,** 1375–1380 (2018).
151. Guinney, J., Dienstmann, R., Wang, X., de Reyniès, A., Schlicker, A., Soneson, C., Marisa, L., Roepman, P., Nyamundanda, G., Angelino, P., Bot, B. M., Morris, J. S., Simon, I. M., Gerster, S., Fessler, E., De Sousa E Melo, F., Missiaglia, E., Ramay, H., Barras, D., Homicsko, K., Maru, D., Manyam, G. C., Broom, B., Boige, V., Perez-Villamil, B., Laderas, T., Salazar, R., Gray, J. W., Hanahan, D., Tabernero, J., Bernards, R., Friend, S. H., Laurent-Puig, P., Medema, J. P., Sadanandam, A., Wessels, L., Delorenzi, M., Kopetz, S., Vermeulen, L. & Tejpar, S. The consensus molecular subtypes of colorectal cancer. *Nat. Med.* **21,** 1350–1356 (2015).
152. Laforest, A., Aparicio, T., Zaanan, A., Silva, F. P., Didelot, A., Desbeaux, A., Le Corre, D., Benhaim, L., Pallier, K., Aust, D., Pistorius, S., Blons, H., Svrcek, M. & Laurent-Puig, P. ERBB2 gene as a potential therapeutic target in small bowel adenocarcinoma. *Eur. J. Cancer* **50,** 1740–1746 (2014).
153. Yuan, W., Zhang, Z., Dai, B., Wei, Q., Liu, J., Liu, Y., Liu, Y., He, L. & Zhou, D. Whole-exome sequencing of duodenal adenocarcinoma identifies recurrent Wnt/β-catenin signaling pathway mutations. *Cancer* **122,** 1689–1696 (2016).
154. Gingras, M.-C., Covington, K. R., Chang, D. K., Donehower, L. A., Gill, A. J., Ittmann, M. M., Creighton, C. J., Johns, A. L., Shinbrot, E., Dewal, N., Fisher, W. E., Australian Pancreatic Cancer Genome Initiative, Pilarsky, C., Grützmann, R., Overman, M. J., Jamieson, N. B., Van Buren, G., 2nd, Drummond, J., Walker, K., Hampton, O. A., Xi, L., Muzny, D. M., Doddapaneni, H., Lee, S. L., Bellair, M., Hu, J., Han, Y., Dinh, H. H., Dahdouli, M., Samra, J. S., Bailey, P., Waddell, N., Pearson, J. V., Harliwong, I., Wang, H., Aust, D., Oien, K. A., Hruban, R. H., Hodges, S. E., McElhany, A., Saengboonmee, C., Duthie, F. R., Grimmond, S. M., Biankin, A. V., Wheeler, D. A. & Gibbs, R. A. Ampullary Cancers Harbor ELF3 Tumor Suppressor Gene Mutations and Exhibit Frequent WNT Dysregulation. *Cell Rep.* **14,** 907–919 (2016).
155. Haan, J. C., Buffart, T. E., Eijk, P. P., van de Wiel, M. A., van Wieringen, W. N., Howdle, P. D., Mulder, C. J. J., van de Velde, C. J., Quirke, P., Nagtegaal, I. D., van Grieken, N. C. T., Grabsch, H., Meijer, G. A. & Ylstra, B. Small bowel adenocarcinoma copy number profiles are more closely related to colorectal than to gastric cancers. *Ann. Oncol.* **23,** 367–374 (2012).
156. Lichtenstein, P., Holm, N. V., Verkasalo, P. K., Iliadou, A., Kaprio, J., Koskenvuo, M., Pukkala, E., Skytthe, A. & Hemminki, K. Environmental and Heritable Factors in the Causation of Cancer — Analyses of Cohorts of Twins from Sweden, Denmark, and Finland. *N. Engl. J. Med.* **343,** 78–85 (2000).
157. Jasperson, K. W., Tuohy, T. M., Neklason, D. W. & Burt, R. W. Hereditary and Familial Colon Cancer. *Gastroenterology* **138,** 2044–2058 (2010).
158. Valle, L. Genetic predisposition to colorectal cancer: where we stand and future perspectives. *World J. Gastroenterol.* **20,** 9828–9849 (2014).
159. Heitzer, E. & Tomlinson, I. Replicative DNA polymerase mutations in cancer. *Curr. Opin. Genet. Dev.* **24,** 107–113 (2014).
160. de la Chapelle, A. Genetic predisposition to colorectal cancer. *Nat. Rev. Cancer* **4,** 769–780 (2004).

161. Laken, S. J., Petersen, G. M., Gruber, S. B., Oddoux, C., Ostrer, H., Giardiello, F. M., Hamilton, S. R., Hampel, H., Markowitz, A., Klimstra, D., Jhanwar, S., Winawer, S., Offit, K., Luce, M. C., Kinzler, K. W. & Vogelstein, B. Familial colorectal cancer in Ashkenazim due to a hypermutable tract in APC. *Nat. Genet.* **17,** 79–83 (1997).
162. Law, P. J., Timofeeva, M., Fernandez-Rozadilla, C., Broderick, P., Studd, J., Fernandez-Tajes, J., Farrington, S., Svinti, V., Palles, C., Orlando, G., Sud, A., Holroyd, A., Penegar, S., Theodoratou, E., Vaughan-Shaw, P., Campbell, H., Zgaga, L., Hayward, C., Campbell, A., Harris, S., Deary, I. J., Starr, J., Gatcombe, L., Pinna, M., Briggs, S., Martin, L., Jaeger, E., Sharma-Oates, A., East, J., Leedham, S., Arnold, R., Johnstone, E., Wang, H., Kerr, D., Kerr, R., Maughan, T., Kaplan, R., Al-Tassan, N., Palin, K., Hänninen, U. A., Cajuso, T., Tanskanen, T., Kondelin, J., Kaasinen, E., Sarin, A.-P., Eriksson, J. G., Rissanen, H., Knekt, P., Pukkala, E., Jousilahti, P., Salomaa, V., Ripatti, S., Palotie, A., Renkonen-Sinisalo, L., Lepistö, A., Böhm, J., Mecklin, J.-P., Buchanan, D. D., Win, A.-K., Hopper, J., Jenkins, M. E., Lindor, N. M., Newcomb, P. A., Gallinger, S., Duggan, D., Casey, G., Hoffmann, P., Nöthen, M. M., Jöckel, K.-H., Easton, D. F., Pharoah, P. D. P., Peto, J., Canzian, F., Swerdlow, A., Eeles, R. A., Kote-Jarai, Z., Muir, K., Pashayan, N., PRACTICAL consortium, Harkin, A., Allan, K., McQueen, J., Paul, J., Iveson, T., Saunders, M., Butterbach, K., Chang-Claude, J., Hoffmeister, M., Brenner, H., Kirac, I., Matošević, P., Hofer, P., Brezina, S., Gsur, A., Cheadle, J. P., Aaltonen, L. A., Tomlinson, I., Houlston, R. S. & Dunlop, M. G. Association analyses identify 31 new risk loci for colorectal cancer susceptibility. *Nat. Commun.* **10,** 2154 (2019).
163. Schmit, S. L., Edlund, C. K., Schumacher, F. R., Gong, J., Harrison, T. A., Huyghe, J. R., Qu, C., Melas, M., Van Den Berg, D. J., Wang, H., Tring, S., Plummer, S. J., Albanes, D., Alonso, M. H., Amos, C. I., Anton, K., Aragaki, A. K., Arndt, V., Barry, E. L., Berndt, S. I., Bezieau, S., Bien, S., Bloomer, A., Boehm, J., Boutron-Ruault, M.-C., Brenner, H., Brezina, S., Buchanan, D. D., Butterbach, K., Caan, B. J., Campbell, P. T., Carlson, C. S., Castelao, J. E., Chan, A. T., Chang-Claude, J., Chanock, S. J., Cheng, I., Cheng, Y.-W., Chin, L. S., Church, J. M., Church, T., Coetzee, G. A., Cotterchio, M., Cruz Correa, M., Curtis, K. R., Duggan, D., Easton, D. F., English, D., Feskens, E. J. M., Fischer, R., FitzGerald, L. M., Fortini, B. K., Fritsche, L. G., Fuchs, C. S., Gago-Dominguez, M., Gala, M., Gallinger, S. J., Gauderman, W. J., Giles, G. G., Giovannucci, E. L., Gogarten, S. M., Gonzalez-Villalpando, C., Gonzalez-Villalpando, E. M., Grady, W. M., Greenson, J. K., Gsur, A., Gunter, M., Haiman, C. A., Hampe, J., Harlid, S., Harju, J. F., Hayes, R. B., Hofer, P., Hoffmeister, M., Hopper, J. L., Huang, S.-C., Huerta, J. M., Hudson, T. J., Hunter, D. J., Idos, G. E., Iwasaki, M., Jackson, R. D., Jacobs, E. J., Jee, S. H., Jenkins, M. A., Jia, W.-H., Jiao, S., Joshi, A. D., Kolonel, L. N., Kono, S., Kooperberg, C., Krogh, V., Kuehn, T., Küry, S., LaCroix, A., Laurie, C. A., Lejbkowicz, F., Lemire, M., Lenz, H.-J., Levine, D., Li, C. I., Li, L., Lieb, W., Lin, Y., Lindor, N. M., Liu, Y.-R., Loupakis, F., Lu, Y., Luh, F., Ma, J., Mancao, C., Manion, F. J., Markowitz, S. D., Martin, V., Matsuda, K., Matsuo, K., McDonnell, K. J., McNeil, C. E., Milne, R., Molina, A. J., Mukherjee, B., Murphy, N., Newcomb, P. A., Offit, K., Omichessan, H., Palli, D., Cotoré, J. P. P., Pérez-Mayoral, J., Pharoah, P. D., Potter, J. D., Qu, C., Raskin, L., Rennert, G., Rennert, H. S., Riggs, B. M., Schafmayer, C., Schoen, R. E., Sellers, T. A., Seminara, D., Severi, G., Shi, W., Shibata, D., Shu, X.-O., Siegel, E. M., Slattery, M. L., Southey, M., Stadler, Z. K., Stern, M. C., Stintzing, S., Taverna, D., Thibodeau, S. N., Thomas, D. C., Trichopoulou, A., Tsugane, S., Ulrich, C. M., van Duijnhoven, F. J. B., van Guelpan, B., Vijai, J., Virtamo, J., Weinstein, S. J., White, E., Win, A. K., Wolk, A., Woods, M., Wu, A. H., Wu, K., Xiang, Y.-B., Yen, Y., Zanke, B. W., Zeng, Y.-X., Zhang, B., Zubair, N., Kweon, S.-S., Figueiredo, J. C., Zheng, W., Marchand, L. L., Lindblom, A., Moreno, V., Peters, U., Casey, G., Hsu, L., Conti, D. V. & Gruber, S. B. Novel Common Genetic Susceptibility Loci for Colorectal Cancer. *J. Natl. Cancer Inst.* **111,** 146–157 (2019).
164. Orlando, G., Law, P. J., Palin, K., Tuupanen, S., Gylfe, A., Hänninen, U. A., Cajuso, T., Tanskanen, T., Kondelin, J., Kaasinen, E., Sarin, A.-P., Kaprio, J., Eriksson, J. G., Rissanen, H., Knekt, P., Pukkala, E., Jousilahti, P., Salomaa, V., Ripatti, S., Palotie, A., Järvinen, H., Renkonen-Sinisalo, L., Lepistö, A., Böhm, J., Mecklin, J.-P., Al-Tassan, N. A., Palles, C., Martin, L., Barclay, E., Tenesa, A., Farrington, S., Timofeeva, M. N., Meyer, B. F., Wakil, S. M., Campbell, H., Smith, C. G., Idziaszczyk, S., Maughan, T. S., Kaplan, R., Kerr, R., Kerr, D., Buchanan, D. D., Win, A. K., Hopper, J., Jenkins, M., Lindor, N. M., Newcomb, P. A., Gallinger, S., Conti, D., Schumacher, F., Casey, G., Taipale, J., Cheadle, J. P., Dunlop, M. G., Tomlinson, I. P., Aaltonen, L. A. & Houlston, R. S. Variation at 2q35 (PNKD and TMBIM1) influences colorectal cancer risk and identifies a pleiotropic effect with inflammatory bowel disease. *Hum. Mol. Genet.* **25,** 2349–2359 (2016).
165. Lu, Y., Kweon, S.-S., Tanikawa, C., Jia, W.-H., Xiang, Y.-B., Cai, Q., Zeng, C., Schmit, S. L., Shin,

A., Matsuo, K., Jee, S. H., Kim, D.-H., Kim, J., Wen, W., Shi, J., Guo, X., Li, B., Wang, N., Zhang, B., Li, X., Shin, M.-H., Li, H.-L., Ren, Z., Oh, J. H., Oze, I., Ahn, Y.-O., Jung, K. J., Conti, D. V., Schumacher, F. R., Rennert, G., Jenkins, M. A., Campbell, P. T., Hoffmeister, M., Casey, G., Gruber, S. B., Gao, J., Gao, Y.-T., Pan, Z.-Z., Kamatani, Y., Zeng, Y.-X., Shu, X.-O., Long, J., Matsuda, K. & Zheng, W. Large-Scale Genome-Wide Association Study of East Asians Identifies Loci Associated With Risk for Colorectal Cancer. *Gastroenterology* **156,** 1455–1466 (2019).

166. Tomlinson, I., the CORGI Consortium, Webb, E., Carvajal-Carmona, L., Broderick, P., Kemp, Z., Spain, S., Penegar, S., Chandler, I., Gorman, M., Wood, W., Barclay, E., Lubbe, S., Martin, L., Sellick, G., Jaeger, E., Hubner, R., Wild, R., Rowan, A., Fielding, S., Howarth, K., Silver, A., Atkin, W., Muir, K., Logan, R., Kerr, D., Johnstone, E., Sieber, O., Gray, R., Thomas, H., Peto, J., Cazier, J.-B. & Houlston, R. A genome-wide association scan of tag SNPs identifies a susceptibility variant for colorectal cancer at 8q24.21. *Nat. Genet.* **39,** 984–988 (2007).
167. Tuupanen, S., Turunen, M., Lehtonen, R., Hallikas, O., Vanharanta, S., Kivioja, T., Björklund, M., Wei, G., Yan, J., Niittymäki, I., Mecklin, J.-P., Järvinen, H., Ristimäki, A., Di-Bernardo, M., East, P., Carvajal-Carmona, L., Houlston, R. S., Tomlinson, I., Palin, K., Ukkonen, E., Karhu, A., Taipale, J. & Aaltonen, L. A. The common colorectal cancer predisposition SNP rs6983267 at chromosome 8q24 confers potential to enhanced Wnt signaling. *Nat. Genet.* **41,** 885–890 (2009).
168. Sur, I. K., Hallikas, O., Vähärautio, A., Yan, J., Turunen, M., Enge, M., Taipale, M., Karhu, A., Aaltonen, L. A. & Taipale, J. Mice lacking a Myc enhancer that includes human SNP rs6983267 are resistant to intestinal tumors. *Science* **338,** 1360–1363 (2012).
169. Watson, P., Vasen, H. F. A., Mecklin, J.-P., Bernstein, I., Aarnio, M., Järvinen, H. J., Myrhøj, T., Sunde, L., Wijnen, J. T. & Lynch, H. T. The risk of extra-colonic, extra-endometrial cancer in the Lynch syndrome. *Int. J. Cancer* **123,** 444–449 (2008).
170. Sankila, R., Aaltonen, L. A., Järvinen, H. J. & Mecklin, J. P. Better survival rates in patients with MLH1-associated hereditary colorectal cancer. *Gastroenterology* **110,** 682–687 (1996).
171. Domingo, E., Laiho, P., Ollikainen, M., Pinto, M., Wang, L., French, A. J., Westra, J., Frebourg, T., Espín, E., Armengol, M., Hamelin, R., Yamamoto, H., Hofstra, R. M. W., Seruca, R., Lindblom, A., Peltomäki, P., Thibodeau, S. N., Aaltonen, L. A. & Schwartz, S., Jr. BRAF screening as a low-cost effective strategy for simplifying HNPCC genetic testing. *J. Med. Genet.* **41,** 664–668 (2004).
172. Peltomäki, P., Aaltonen, L. A., Sistonen, P., Pylkkänen, L., Mecklin, J. P., Järvinen, H., Green, J. S., Jass, J. R., Weber, J. L. & Leach, F. S. Genetic mapping of a locus predisposing to human colorectal cancer. *Science* **260,** 810–812 (1993).
173. Fishel, R., Lescoe, M. K., Rao, M. R., Copeland, N. G., Jenkins, N. A., Garber, J., Kane, M. & Kolodner, R. The human mutator gene homolog MSH2 and its association with hereditary nonpolyposis colon cancer. *Cell* **77,** 1 p following 166 (1994).
174. Leach, F. S., Nicolaides, N. C., Papadopoulos, N., Liu, B., Jen, J., Parsons, R., Peltomäki, P., Sistonen, P., Aaltonen, L. A., Nyström-Lahti, M., -Y. Guan, X., Zhang, J., Meltzer, P. S., Yu, J.-W., Kao, F.-T., Chen, D. J., Cerosaletti, K. M., Keith Fournier, R. E., Todd, S., Lewis, T., Leach, R. J., Naylor, S. L., Weissenbach, J., Mecklin, J.-P., Järvinen, H., Petersen, G. M., Hamilton, S. R., Green, J., Jass, J., Watson, P., Lynch, H. T., Trent, J. M., de la Chapelle, A., Kinzler, K. W. & Vogelstein, B. Mutations of a mutS homolog in hereditary nonpolyposis colorectal cancer. *Cell* **75,** 1215–1225 (1993).
175. Lindblom, A., Tannergård, P., Werelius, B. & Nordenskjöld, M. Genetic mapping of a second locus predisposing to hereditary non-polyposis colon cancer. *Nat. Genet.* **5,** 279–282 (1993).
176. Bronner, C. E., Eric Bronner, C., Baker, S. M., Morrison, P. T., Warren, G., Smith, L. G., Lescoe, M. K., Kane, M., Earabino, C., Lipford, J., Lindblom, A., Tannergård, P., Bollag, R. J., Godwin, A. R., Ward, D. C., Nordenskjold, M., Fishel, R., Kolodner, R. & Michael Liskay, R. Mutation in the DNA mismatch repair gene homologue hMLH 1 is associated with hereditary non-polyposis colon cancer. *Nature* **368,** 258–261 (1994).
177. Papadopoulos, N., Nicolaides, N., Wei, Y., Ruben, S., Carter, K., Rosen, C., Haseltine, W., Fleischmann, R., Fraser, C., Adams, M. & Et, A. Mutation of a mutL homolog in hereditary colon cancer. *Science* **263,** 1625–1629 (1994).
178. Nicolaides, N. C., Papadopoulos, N., Liu, B., Wei, Y. F., Carter, K. C., Ruben, S. M., Rosen, C. A., Haseltine, W. A., Fleischmann, R. D. & Fraser, C. M. Mutations of two PMS homologues in hereditary nonpolyposis colon cancer. *Nature* **371,** 75–80 (1994).
179. Miyaki, M., Konishi, M., Tanaka, K., Kikuchi-Yanoshita, R., Muraoka, M., Yasuno, M., Igari, T., Koike, M., Chiba, M. & Mori, T. Germline mutation of MSH6 as the cause of hereditary nonpolyposis colorectal cancer. *Nat. Genet.* **17,** 271–272 (1997).

180. Kovacs, M. E., Papp, J., Szentirmay, Z., Otto, S. & Olah, E. Deletions removing the last exon of TACSTD1 constitute a distinct class of mutations predisposing to Lynch syndrome. *Hum. Mutat.* **30,** 197–203 (2009).
181. Ligtenberg, M. J. L., Kuiper, R. P., Chan, T. L., Goossens, M., Hebeda, K. M., Voorendt, M., Lee, T. Y. H., Bodmer, D., Hoenselaar, E., Hendriks-Cornelissen, S. J. B., Tsui, W. Y., Kong, C. K., Brunner, H. G., van Kessel, A. G., Yuen, S. T., van Krieken, J. H. J. M., Leung, S. Y. & Hoogerbrugge, N. Heritable somatic methylation and inactivation of MSH2 in families with Lynch syndrome due to deletion of the 3' exons of TACSTD1. *Nat. Genet.* **41,** 112–117 (2009).
182. Peltomäki, P. Lynch syndrome genes. *Fam. Cancer* **4,** 227–232 (2005).
183. Nyström-Lahti, M., Kristo, P., Nicolaides, N. C., Chang, S.-Y., Aaltonen, L. A., Moisio, A.-L., Järvinen, H. J., Mecklin, J.-P., Kinzler, K. W., Vogelstein, B., De La Chapelle, A. & Peltomäki, P. Founding mutations and Alu-mediated recombination in hereditary colon cancer. *Nat. Med.* **1,** 1203–1206 (1995).
184. Vasen, H. F. A., -P. Mecklin, J., Meera Khan, P. & Lynch, H. T. The International Collaborative Group on Hereditary Non-Polyposis Colorectal Cancer (ICG-HNPCC). *Dis. Colon Rectum* **34,** 424–425 (1991).
185. Vasen, H., Watson, P., Mecklin, J. & Lynch, H. New clinical criteria for hereditary nonpolyposis colorectal cancer (HNPCC, Lynch syndrome) proposed by the International Collaborative Group on HNPCC. *Gastroenterology* **116,** 1453–1456 (1999).
186. Rodriguez-Bigas, M. A., Boland, C. R., Hamilton, S. R., Henson, D. E., Jass, J. R., Khan, P. M., Lynch, H., Perucho, M., Smyrk, T., Sobin, L. & Srivastava, S. A National Cancer Institute Workshop on Hereditary Nonpolyposis Colorectal Cancer Syndrome: meeting highlights and Bethesda guidelines. *J. Natl. Cancer Inst.* **89,** 1758–1762 (1997).
187. Heald, B., Plesec, T., Liu, X., Pai, R., Patil, D., Moline, J., Sharp, R. R., Burke, C. A., Kalady, M. F., Church, J. & Eng, C. Implementation of universal microsatellite instability and immunohistochemistry screening for diagnosing lynch syndrome in a large academic medical center. *J. Clin. Oncol.* **31,** 1336–1340 (2013).
188. de Vos tot Nederveen Cappel, W. H., Järvinen, H. J., Lynch, P. M., Engel, C., Mecklin, J.-P. & Vasen, H. F. A. Colorectal surveillance in Lynch syndrome families. *Fam. Cancer* **12,** 261–265 (2013).
189. Seppälä, T., in collaboration with The Mallorca Group, Pylvänäinen, K., Evans, D. G., Järvinen, H., Renkonen-Sinisalo, L., Bernstein, I., Holinski-Feder, E., Sala, P., Lindblom, A., Macrae, F., Blanco, I., Sijmons, R., Jeffries, J., Vasen, H., Burn, J., Nakken, S., Hovig, E., Rødland, E. A., Tharmaratnam, K., de Vos tot Nederveen Cappel, W. H., Hill, J., Wijnen, J., Jenkins, M., Genuardi, M., Green, K., Lalloo, F., Sunde, L., Mints, M., Bertario, L., Pineda, M., Navarro, M., Morak, M., Frayling, I. M., Plazzer, J.-P., Sampson, J. R., Capella, G., Möslein, G., Mecklin, J.-P. & Møller, P. Colorectal cancer incidence in path_MLH1 carriers subjected to different follow-up protocols: a Prospective Lynch Syndrome Database report. *Hered. Cancer Clin. Pract.* **15,** 18 (2017).
190. Syngal, S., Brand, R. E., Church, J. M., Giardiello, F. M., Hampel, H. L. & Burt, R. W. ACG Clinical Guideline: Genetic Testing and Management of Hereditary Gastrointestinal Cancer Syndromes. *Am. J Gastroenterol.* **110,** 223–262 (2015).
191. Jarvinen, H. J. Epidemiology of familial adenomatous polyposis in Finland: impact of family screening on the colorectal cancer rate and survival. *Gut* **33,** 357–360 (1992).
192. Half, E., Bercovich, D. & Rozen, P. Familial adenomatous polyposis. *Orphanet J. Rare Dis.* **4,** 22 (2009).
193. Bisgaard, M. L., Fenger, K., Bülow, S., Niebuhr, E. & Mohr, J. Familial adenomatous polyposis (FAP): frequency, penetrance, and mutation rate. *Hum. Mutat.* **3,** 121–125 (1994).
194. Burt, R. W., Leppert, M. F., Slattery, M. L., Samowitz, W. S., Spirio, L. N., Kerber, R. A., Kuwada, S. K., Neklason, D. W., DiSario, J. A., Lyon, E., Preston Hughes, J., Chey, W. Y. & White, R. L. Genetic testing and phenotype in a large kindred with attenuated familial adenomatous polyposis. *Gastroenterology* **127,** 444–451 (2004).
195. Al-Tassan, N., Chmiel, N. H., Maynard, J., Fleming, N., Livingston, A. L., Williams, G. T., Hodges, A. K., Rhodri Davies, D., David, S. S., Sampson, J. R. & Cheadle, J. P. Inherited variants of MYH associated with somatic G:C→T:A mutations in colorectal tumors. *Nat. Genet.* **30,** 227–232 (2002).
196. Farrington, S. M., Tenesa, A., Barnetson, R., Wiltshire, A., Prendergast, J., Porteous, M., Campbell, H. & Dunlop, M. G. Germline Susceptibility to Colorectal Cancer Due to Base-Excision Repair Gene Defects. *Am. J. Hum. Genet.* **77,** 112–119 (2005).
197. Vogt, S., Jones, N., Christian, D., Engel, C., Nielsen, M., Kaufmann, A., Steinke, V., Vasen, H. F.,

Propping, P., Sampson, J. R., Hes, F. J. & Aretz, S. Expanded extracolonic tumor spectrum in MUTYH-associated polyposis. *Gastroenterology* **137,** 1976–85.e1–10 (2009).
198. Hemminki, A., Markie, D., Tomlinson, I., Avizienyte, E., Roth, S., Loukola, A., Bignell, G., Warren, W., Aminoff, M., Höglund, P., Järvinen, H., Kristo, P., Pelin, K., Ridanpää, M., Salovaara, R., Toro, T., Bodmer, W., Olschwang, S., Olsen, A. S., Stratton, M. R., de la Chapelle, A. & Aaltonen, L. A. A serine/threonine kinase gene defective in Peutz–Jeghers syndrome. *Nature* **391,** 184–187 (1998).
199. Gammon, A., Jasperson, K., Kohlmann, W. & Burt, R. W. Hamartomatous polyposis syndromes. *Best Pract. Res. Clin. Gastroenterol.* **23,** 219–231 (2009).
200. Howe, J. R., Roth, S., Ringold, J. C., Summers, R. W., Järvinen, H. J., Sistonen, P., Tomlinson, I. P., Houlston, R. S., Bevan, S., Mitros, F. A., Stone, E. M. & Aaltonen, L. A. Mutations in the SMAD4/DPC4 gene in juvenile polyposis. *Science* **280,** 1086–1088 (1998).
201. Howe, J. R., Bair, J. L., Sayed, M. G., Anderson, M. E., Mitros, F. A., Petersen, G. M., Velculescu, V. E., Traverso, G. & Vogelstein, B. Germline mutations of the gene encoding bone morphogenetic protein receptor 1A in juvenile polyposis. *Nat. Genet.* **28,** 184–187 (2001).
202. Aaltonen, L. A., Salovaara, R., Kristo, P., Canzian, F., Hemminki, A., Peltomäki, P., Chadwick, R. B., Kääriäinen, H., Eskelinen, M., Järvinen, H., Mecklin, J. P. & de la Chapelle, A. Incidence of hereditary nonpolyposis colorectal cancer and the feasibility of molecular screening for the disease. *N. Engl. J. Med.* **338,** 1481–1487 (1998).
203. Salovaara, R., Loukola, A., Kristo, P., Kääriäinen, H., Ahtola, H., Eskelinen, M., Härkönen, N., Julkunen, R., Kangas, E., Ojala, S., Tulikoura, J., Valkamo, E., Järvinen, H., Mecklin, J. P., Aaltonen, L. A. & de la Chapelle, A. Population-based molecular detection of hereditary nonpolyposis colorectal cancer. *J. Clin. Oncol.* **18,** 2193–2200 (2000).
204. Tanskanen, T., Gylfe, A. E., Katainen, R., Taipale, M., Renkonen-Sinisalo, L., Mecklin, J.-P., Järvinen, H., Tuupanen, S., Kilpivaara, O., Vahteristo, P. & Aaltonen, L. A. Exome sequencing in diagnostic evaluation of colorectal cancer predisposition in young patients. *Scand. J. Gastroenterol.* **48,** 672–678 (2013).
205. Okonechnikov, K., Conesa, A. & García-Alcalde, F. Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. *Bioinformatics* **32,** 292–294 (2016).
206. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25,** 1754–1760 (2009).
207. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R. & 1000 Genome Project Data Processing Subgroup. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25,** 2078–2079 (2009).
208. DePristo, M. A., Banks, E., Poplin, R., Garimella, K. V., Maguire, J. R., Hartl, C., Philippakis, A. A., del Angel, G., Rivas, M. A., Hanna, M., McKenna, A., Fennell, T. J., Kernytsky, A. M., Sivachenko, A. Y., Cibulskis, K., Gabriel, S. B., Altshuler, D. & Daly, M. J. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **43,** 491–498 (2011).
209. Hiltemann, S., Jenster, G., Trapman, J., van der Spek, P. & Stubbs, A. Discriminating somatic and germline mutations in tumor DNA samples without matching normals. *Genome Res.* **25,** 1382–1390 (2015).
210. Katainen, R., Donner, I., Cajuso, T., Kaasinen, E., Palin, K., Mäkinen, V., Aaltonen, L. A. & Pitkänen, E. Discovery of potential causative mutations in human coding and noncoding genome with the interactive software BasePlayer. *Nat. Protoc.* **13,** 2580–2600 (2018).
211. Untergasser, A., Nijveen, H., Rao, X., Bisseling, T., Geurts, R. & Leunissen, J. A. M. Primer3Plus, an enhanced web interface to Primer3. *Nucleic Acids Res.* **35,** W71–W74 (2007).
212. Touloumis, A., Agresti, A. & Kateri, M. GEE for multinomial responses using a local odds ratios parameterization. *Biometrics* **69,** 633–640 (2013).
213. Alexandrov, L. B., Nik-Zainal, S., Wedge, D. C., Campbell, P. J. & Stratton, M. R. Deciphering signatures of mutational processes operative in human cancer. *Cell Rep.* **3,** 246–259 (2013).
214. Staaf, J., Lindgren, D., Vallon-Christersson, J., Isaksson, A., Göransson, H., Juliusson, G., Rosenquist, R., Höglund, M., Borg, A. & Ringnér, M. Segmentation-based detection of allelic imbalance and loss-of-heterozygosity in cancer cells using whole genome SNP arrays. *Genome Biol.* **9,** R136 (2008).
215. Bankhead, P., Loughrey, M. B., Fernández, J. A., Dombrowski, Y., McArt, D. G., Dunne, P. D., McQuaid, S., Gray, R. T., Murray, L. J., Coleman, H. G., James, J. A., Salto-Tellez, M. & Hamilton, P. W. QuPath: Open source software for digital pathology image analysis. *Sci. Rep.* **7,** 16878 (2017).
216. Wirta, E.-V., Seppälä, T., Friman, M., Väyrynen, J., Ahtiainen, M., Kautiainen, H., Kuopio, T., Kellokumpu, I., Mecklin, J.-P. & Böhm, J. Immunoscore in mismatch repair-proficient and -deficient

colon cancer. *J Pathol Clin Res.* **3,** 203–213 (2017).
217. Gylfe, A. E., Kondelin, J., Turunen, M., Ristolainen, H., Katainen, R., Pitkänen, E., Kaasinen, E., Rantanen, V., Tanskanen, T., Varjosalo, M., Lehtonen, H., Palin, K., Taipale, M., Taipale, J., Renkonen-Sinisalo, L., Järvinen, H., Böhm, J., Mecklin, J.-P., Ristimäki, A., Kilpivaara, O., Tuupanen, S., Karhu, A., Vahteristo, P. & Aaltonen, L. A. Identification of candidate oncogenes in human colorectal cancers with microsatellite instability. *Gastroenterology* **145,** 540–3.e22 (2013).
218. Cerami, E., Gao, J., Dogrusoz, U., Gross, B. E., Sumer, S. O., Aksoy, B. A., Jacobsen, A., Byrne, C. J., Heuer, M. L., Larsson, E., Antipin, Y., Reva, B., Goldberg, A. P., Sander, C. & Schultz, N. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov.* **2,** 401–404 (2012).
219. Forbes, S. A., Bindal, N., Bamford, S., Cole, C., Kok, C. Y., Beare, D., Jia, M., Shepherd, R., Leung, K., Menzies, A., Teague, J. W., Campbell, P. J., Stratton, M. R. & Futreal, P. A. COSMIC: mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer. *Nucleic Acids Res.* **39,** D945–50 (2011).
220. Consortium, T. I. C. G. & The International Cancer Genome Consortium. International network of cancer genome projects. *Nature* **464,** 993–998 (2010).
221. Chang, M. T., Asthana, S., Gao, S. P., Lee, B. H., Chapman, J. S., Kandoth, C., Gao, J., Socci, N. D., Solit, D. B., Olshen, A. B., Schultz, N. & Taylor, B. S. Identifying recurrent mutations in cancer reveals widespread lineage diversity and mutational specificity. *Nat. Biotechnol.* **34,** 155–163 (2016).
222. Cancer Genome Atlas Research Network. Comprehensive molecular characterization of gastric adenocarcinoma. *Nature* **513,** 202–209 (2014).
223. Patsialou, A. DNA-binding properties of ARID family proteins. *Nucleic Acids Res.* **33,** 66–80 (2005).
224. Wilson, B. G. & Roberts, C. W. M. SWI/SNF nucleosome remodellers and cancer. *Nat. Rev. Cancer* **11,** 481–492 (2011).
225. Wang, X., Nagl, N. G., Wilsker, D., Van Scoy, M., Pacchione, S., Yaciuk, P., Dallas, P. B. & Moran, E. Two related ARID family proteins are alternative subunits of human SWI/SNF complexes. *Biochem. J* **383,** 319–325 (2004).
226. Jones, S., Wang, T.-L., Shih, I.-M., Mao, T.-L., Nakayama, K., Roden, R., Glas, R., Slamon, D., Diaz, L. A., Vogelstein, B., Kinzler, K. W., Velculescu, V. E. & Papadopoulos, N. Frequent Mutations of Chromatin Remodeling Gene ARID1A in Ovarian Clear Cell Carcinoma. *Science* **330,** 228–231 (2010).
227. Wiegand, K. C., Shah, S. P., Al-Agha, O. M., Zhao, Y., Tse, K., Zeng, T., Senz, J., McConechy, M. K., Anglesio, M. S., Kalloger, S. E., Yang, W., Heravi-Moussavi, A., Giuliany, R., Chow, C., Fee, J., Zayed, A., Prentice, L., Melnyk, N., Turashvili, G., Delaney, A. D., Madore, J., Yip, S., McPherson, A. W., Ha, G., Bell, L., Fereday, S., Tam, A., Galletta, L., Tonin, P. N., Provencher, D., Miller, D., Jones, S. J. M., Moore, R. A., Morin, G. B., Oloumi, A., Boyd, N., Aparicio, S. A., Shih, I.-M., Mes-Masson, A.-M., Bowtell, D. D., Hirst, M., Gilks, B., Marra, M. A. & Huntsman, D. G. ARID1A mutations in endometriosis-associated ovarian carcinomas. *N. Engl. J. Med.* **363,** 1532–1543 (2010).
228. Wang, K., Kan, J., Yuen, S. T., Shi, S. T., Chu, K. M., Law, S., Chan, T. L., Kan, Z., Chan, A. S. Y., Tsui, W. Y., Lee, S. P., Ho, S. L., Chan, A. K. W., Cheng, G. H. W., Roberts, P. C., Rejto, P. A., Gibson, N. W., Pocalyko, D. J., Mao, M., Xu, J. & Leung, S. Y. Exome sequencing identifies frequent mutation of ARID1A in molecular subtypes of gastric cancer. *Nat Genet.* **43,** 1219–1223 (2011).
229. Shain, A. H., Giacomini, C. P., Matsukuma, K., Karikari, C. A., Bashyam, M. D., Hidalgo, M., Maitra, A. & Pollack, J. R. Convergent structural alterations define SWItch/Sucrose NonFermentable (SWI/SNF) chromatin remodeler as a central tumor suppressive complex in pancreatic cancer. *Proc. Natl. Acad. Sci. U. S. A.* **109,** E252–E259 (2012).
230. Shain, A. H. & Pollack, J. R. The spectrum of SWI/SNF mutations, ubiquitous in human cancers. *PLoS One* **8,** e55119 (2013).
231. Waddell, N., Pajic, M., Patch, A.-M., Chang, D. K., Kassahn, K. S., Bailey, P., Johns, A. L., Miller, D., Nones, K., Quek, K., Quinn, M. C. J., Robertson, A. J., Fadlullah, M. Z. H., Bruxner, T. J. C., Christ, A. N., Harliwong, I., Idrisoglu, S., Manning, S., Nourse, C., Nourbakhsh, E., Wani, S., Wilson, P. J., Markham, E., Cloonan, N., Anderson, M. J., Fink, J. L., Holmes, O., Kazakoff, S. H., Leonard, C., Newell, F., Poudel, B., Song, S., Taylor, D., Waddell, N., Wood, S., Xu, Q., Wu, J., Pinese, M., Cowley, M. J., Lee, H. C., Jones, M. D., Nagrial, A. M., Humphris, J., Chantrill, L. A., Chin, V., Steinmann, A. M., Mawson, A., Humphrey, E. S., Colvin, E. K., Chou, A., Scarlett, C. J., Pinho, A. V., Giry-Laterriere, M., Rooman, I., Samra, J. S., Kench, J. G., Pettitt, J. A., Merrett, N. D., Toon, C., Epari, K., Nguyen, N. Q., Barbour, A., Zeps, N., Jamieson, N. B., Graham, J. S., Niclou, S. P., Bjerkvig, R., Grützmann, R., Aust, D., Hruban, R. H., Maitra, A., Iacobuzio-Donahue, C. A.,

Wolfgang, C. L., Morgan, R. A., Lawlor, R. T., Corbo, V., Bassi, C., Falconi, M., Zamboni, G., Tortora, G., Tempero, M. A., Australian Pancreatic Cancer Genome Initiative, Gill, A. J., Eshleman, J. R., Pilarsky, C., Scarpa, A., Musgrove, E. A., Pearson, J. V., Biankin, A. V. & Grimmond, S. M. Whole genomes redefine the mutational landscape of pancreatic cancer. *Nature* **518,** 495–501 (2015).

232. Wang, K., Yuen, S. T., Xu, J., Lee, S. P., Yan, H. H. N., Shi, S. T., Siu, H. C., Deng, S., Chu, K. M., Law, S., Chan, K. H., Chan, A. S. Y., Tsui, W. Y., Ho, S. L., Chan, A. K. W., Man, J. L. K., Foglizzo, V., Ng, M. K., Chan, A. S., Ching, Y. P., Cheng, G. H. W., Xie, T., Fernandez, J., Li, V. S. W., Clevers, H., Rejto, P. A., Mao, M. & Leung, S. Y. Whole-genome sequencing and comprehensive molecular profiling identify new driver mutations in gastric cancer. *Nat. Genet.* **46,** 573–582 (2014).
233. Mathur, R., Alver, B. H., San Roman, A. K., Wilson, B. G., Wang, X., Agoston, A. T., Park, P. J., Shivdasani, R. A. & Roberts, C. W. M. ARID1A loss impairs enhancer-mediated gene regulation and drives colon cancer in mice. *Nat Genet.* **49,** 296–302 (2017).
234. Luchini, C., Veronese, N., Solmi, M., Cho, H., Kim, J.-H., Chou, A., Gill, A. J., Faraj, S. F., Chaux, A., Netto, G. J., Nakayama, K., Kyo, S., Lee, S. Y., Kim, D.-W., Yousef, G. M., Scorilas, A., Nelson, G. S., Köbel, M., Kalloger, S. E., Schaeffer, D. F., Yan, H.-B., Liu, F., Yokoyama, Y., Zhang, X., Pang, D., Lichner, Z., Sergi, G., Manzato, E., Capelli, P., Wood, L. D., Scarpa, A. & Correll, C. U. Prognostic role and implications of mutation status of tumor suppressor gene ARID1A in cancer: a systematic review and meta-analysis. *Oncotarget* **6,** (2015).
235. Wu, J. N. & Roberts, C. W. M. ARID1A Mutations in Cancer: Another Epigenetic Tumor Suppressor? *Cancer Discov.* **3,** 35–43 (2013).
236. Nagl, N. G., Wang, X., Patsialou, A., Van Scoy, M. & Moran, E. Distinct mammalian SWI/SNF chromatin remodeling complexes with opposing roles in cell-cycle control. *EMBO J.* **26,** 752–763 (2007).
237. Stephens, P. J., Tarpey, P. S., Davies, H., Van Loo, P., Greenman, C., Wedge, D. C., Nik-Zainal, S., Martin, S., Varela, I., Bignell, G. R., Yates, L. R., Papaemmanuil, E., Beare, D., Butler, A., Cheverton, A., Gamble, J., Hinton, J., Jia, M., Jayakumar, A., Jones, D., Latimer, C., Lau, K. W., McLaren, S., McBride, D. J., Menzies, A., Mudie, L., Raine, K., Rad, R., Chapman, M. S., Teague, J., Easton, D., Langerød, A., Oslo Breast Cancer Consortium (OSBREAC), Lee, M. T. M., Shen, C.-Y., Tee, B. T. K., Huimin, B. W., Broeks, A., Vargas, A. C., Turashvili, G., Martens, J., Fatima, A., Miron, P., Chin, S.-F., Thomas, G., Boyault, S., Mariani, O., Lakhani, S. R., van de Vijver, M., van 't Veer, L., Foekens, J., Desmedt, C., Sotiriou, C., Tutt, A., Caldas, C., Reis-Filho, J. S., Aparicio, S. A. J. R., Salomon, A. V., Børresen-Dale, A.-L., Richardson, A. L., Campbell, P. J., Futreal, P. A. & Stratton, M. R. The landscape of cancer genes and mutational processes in breast cancer. *Nature* **486,** 400–404 (2012).
238. Fujimoto, A., Totoki, Y., Abe, T., Boroevich, K. A., Hosoda, F., Nguyen, H. H., Aoki, M., Hosono, N., Kubo, M., Miya, F., Arai, Y., Takahashi, H., Shirakihara, T., Nagasaki, M., Shibuya, T., Nakano, K., Watanabe-Makino, K., Tanaka, H., Nakamura, H., Kusuda, J., Ojima, H., Shimada, K., Okusaka, T., Ueno, M., Shigekawa, Y., Kawakami, Y., Arihiro, K., Ohdan, H., Gotoh, K., Ishikawa, O., Ariizumi, S.-I., Yamamoto, M., Yamada, T., Chayama, K., Kosuge, T., Yamaue, H., Kamatani, N., Miyano, S., Nakagama, H., Nakamura, Y., Tsunoda, T., Shibata, T. & Nakagawa, H. Whole-genome sequencing of liver cancers identifies etiological influences on mutation patterns and recurrent mutations in chromatin regulators. *Nat Genet.* **44,** 760–764 (2012).
239. Murakami, R., Matsumura, N., Brown, J. B., Higasa, K., Tsutsumi, T., Kamada, M., Abou-Taleb, H., Hosoe, Y., Kitamura, S., Yamaguchi, K., Abiko, K., Hamanishi, J., Baba, T., Koshiyama, M., Okuno, Y., Yamada, R., Matsuda, F., Konishi, I. & Mandai, M. Exome Sequencing Landscape Analysis in Ovarian Clear Cell Carcinoma Shed Light on Key Chromosomal Regions and Mutation Gene Networks. *Am. J. Pathol.* **187,** 2246–2258 (2017).
240. Sausen, M., Leary, R. J., Jones, S., Wu, J., Reynolds, C. P., Liu, X., Blackford, A., Parmigiani, G., Diaz, L. A., Jr, Papadopoulos, N., Vogelstein, B., Kinzler, K. W., Velculescu, V. E. & Hogarty, M. D. Integrated genomic analyses identify ARID1A and ARID1B alterations in the childhood cancer neuroblastoma. *Nat. Genet.* **45,** 12–17 (2013).
241. Vasileiou, G., Ekici, A. B., Uebe, S., Zweier, C., Hoyer, J., Engels, H., Behrens, J., Reis, A. & Hadjihannas, M. V. Chromatin-Remodeling-Factor ARID1B Represses Wnt/β-Catenin Signaling. *Am J Hum Genet.* **97,** 445–456 (2015).
242. Helming, K. C., Wang, X., Wilson, B. G., Vazquez, F., Haswell, J. R., Manchester, H. E., Kim, Y., Kryukov, G. V., Ghandi, M., Aguirre, A. J., Jagani, Z., Wang, Z., Garraway, L. A., Hahn, W. C. & Roberts, C. W. M. ARID1B is a specific vulnerability in ARID1A-mutant cancers. *Nat. Med.* **20,** 251–254 (2014).

243. Kelso, T. W. R., Porter, D. K., Amaral, M. L., Shokhirev, M. N., Benner, C. & Hargreaves, D. C. Chromatin accessibility underlies synthetic lethality of SWI/SNF subunits in ARID1A-mutant cancers. *eLife* **6,** (2017).
244. Shen, J., Ju, Z., Zhao, W., Wang, L., Peng, Y., Ge, Z., Nagel, Z. D., Zou, J., Wang, C., Kapoor, P., Ma, X., Ma, D., Liang, J., Song, S., Liu, J., Samson, L. D., Ajani, J. A., Li, G.-M., Liang, H., Shen, X., Mills, G. B. & Peng, G. ARID1A deficiency promotes mutability and potentiates therapeutic antitumor immunity unleashed by immune checkpoint blockade. *Nat. Med.* **24,** 556–562 (2018).
245. Yan, Z. PBAF chromatin-remodeling complex requires a novel specificity subunit, BAF200, to regulate expression of selective interferon-responsive genes. *Genes Dev.* **19,** 1662–1667 (2005).
246. Li, M., Zhao, H., Zhang, X., Wood, L. D., Anders, R. A., Choti, M. A., Pawlik, T. M., Daniel, H. D., Kannangai, R., Offerhaus, G. J. A., Velculescu, V. E., Wang, L., Zhou, S., Vogelstein, B., Hruban, R. H., Papadopoulos, N., Cai, J., Torbenson, M. S. & Kinzler, K. W. Inactivating mutations of the chromatin remodeling gene ARID2 in hepatocellular carcinoma. *Nat. Genet.* **43,** 828–829 (2011).
247. Manceau, G., Letouzé, E., Guichard, C., Didelot, A., Cazes, A., Corté, H., Fabre, E., Pallier, K., Imbeaud, S., Le Pimpec-Barthes, F., Zucman-Rossi, J., Laurent-Puig, P. & Blons, H. Recurrent inactivating mutations of ARID2 in non-small cell lung carcinoma. *Int J Cancer* **132,** 2217–2221 (2013).
248. Hodis, E., Watson, I. R., Kryukov, G. V., Arold, S. T., Imielinski, M., Theurillat, J.-P., Nickerson, E., Auclair, D., Li, L., Place, C., DiCara, D., Ramos, A. H., Lawrence, M. S., Cibulskis, K., Sivachenko, A., Voet, D., Saksena, G., Stransky, N., Onofrio, R. C., Winckler, W., Ardlie, K., Wagle, N., Wargo, J., Chong, K., Morton, D. L., Stemke-Hale, K., Chen, G., Noble, M., Meyerson, M., Ladbury, J. E., Davies, M. A., Gershenwald, J. E., Wagner, S. N., Hoon, D. S. B., Schadendorf, D., Lander, E. S., Gabriel, S. B., Getz, G., Garraway, L. A. & Chin, L. A Landscape of Driver Mutations in Melanoma. *Cell* **150,** 251–263 (2012).
249. Biankin, A. V., Waddell, N., Kassahn, K. S., Gingras, M.-C., Muthuswamy, L. B., Johns, A. L., Miller, D. K., Wilson, P. J., Patch, A.-M., Wu, J., Chang, D. K., Cowley, M. J., Gardiner, B. B., Song, S., Harliwong, I., Idrisoglu, S., Nourse, C., Nourbakhsh, E., Manning, S., Wani, S., Gongora, M., Pajic, M., Scarlett, C. J., Gill, A. J., Pinho, A. V., Rooman, I., Anderson, M., Holmes, O., Leonard, C., Taylor, D., Wood, S., Xu, Q., Nones, K., Fink, J. L., Christ, A., Bruxner, T., Cloonan, N., Kolle, G., Newell, F., Pinese, M., Mead, R. S., Humphris, J. L., Kaplan, W., Jones, M. D., Colvin, E. K., Nagrial, A. M., Humphrey, E. S., Chou, A., Chin, V. T., Chantrill, L. A., Mawson, A., Samra, J. S., Kench, J. G., Lovell, J. A., Daly, R. J., Merrett, N. D., Toon, C., Epari, K., Nguyen, N. Q., Barbour, A., Zeps, N., Australian Pancreatic Cancer Genome Initiative, Kakkar, N., Zhao, F., Wu, Y. Q., Wang, M., Muzny, D. M., Fisher, W. E., Brunicardi, F. C., Hodges, S. E., Reid, J. G., Drummond, J., Chang, K., Han, Y., Lewis, L. R., Dinh, H., Buhay, C. J., Beck, T., Timms, L., Sam, M., Begley, K., Brown, A., Pai, D., Panchal, A., Buchner, N., De Borja, R., Denroche, R. E., Yung, C. K., Serra, S., Onetto, N., Mukhopadhyay, D., Tsao, M.-S., Shaw, P. A., Petersen, G. M., Gallinger, S., Hruban, R. H., Maitra, A., Iacobuzio-Donahue, C. A., Schulick, R. D., Wolfgang, C. L., Morgan, R. A., Lawlor, R. T., Capelli, P., Corbo, V., Scardoni, M., Tortora, G., Tempero, M. A., Mann, K. M., Jenkins, N. A., Perez-Mancera, P. A., Adams, D. J., Largaespada, D. A., Wessels, L. F. A., Rust, A. G., Stein, L. D., Tuveson, D. A., Copeland, N. G., Musgrove, E. A., Scarpa, A., Eshleman, J. R., Hudson, T. J., Sutherland, R. L., Wheeler, D. A., Pearson, J. V., McPherson, J. D., Gibbs, R. A. & Grimmond, S. M. Pancreatic cancer genomes reveal aberrations in axon guidance pathway genes. *Nature* **491,** 399–405 (2012).
250. Kovac, M., Navas, C., Horswell, S., Salm, M., Bardella, C., Rowan, A., Stares, M., Castro-Giner, F., Fisher, R., de Bruin, E. C., Kovacova, M., Gorman, M., Makino, S., Williams, J., Jaeger, E., Jones, A., Howarth, K., Larkin, J., Pickering, L., Gore, M., Nicol, D. L., Hazell, S., Stamp, G., O'Brien, T., Challacombe, B., Matthews, N., Phillimore, B., Begum, S., Rabinowitz, A., Varela, I., Chandra, A., Horsfield, C., Polson, A., Tran, M., Bhatt, R., Terracciano, L., Eppenberger-Castori, S., Protheroe, A., Maher, E., El Bahrawy, M., Fleming, S., Ratcliffe, P., Heinimann, K., Swanton, C. & Tomlinson, I. Recurrent chromosomal gains and heterogeneous driver mutations characterise papillary renal cancer evolution. *Nat. Commun.* **6,** 6336 (2015).
251. Jiang, H., Cao, H.-J., Ma, N., Bao, W.-D., Wang, J.-J., Chen, T.-W., Zhang, E.-B., Yuan, Y.-M., Ni, Q.-Z., Zhang, F.-K., Ding, X.-F., Zheng, Q.-W., Wang, Y.-K., Zhu, M., Wang, X., Feng, J., Zhang, X.-L., Cheng, S.-Q., Ma, D.-J., Qiu, L., Li, J.-J. & Xie, D. Chromatin remodeling factor ARID2 suppresses hepatocellular carcinoma metastasis via DNMT1-Snail axis. *Proc. Natl. Acad. Sci. U. S. A.***117,** 4770-4780 (2020).
252. Lai, A., Marcellus, R. C., Corbeil, H. B. & Branton, P. E. RBP1 induces growth arrest by repression of

E2F-dependent transcription. *Oncogene* **18,** 2091–2100 (1999).
253. Wu, M.-Y., Eldin, K. W. & Beaudet, A. L. Identification of chromatin remodeling genes Arid4a and Arid4b as leukemia suppressor genes. *J. Natl. Cancer Inst.* **100,** 1247–1259 (2008).
254. Rodrigues-Lisoni, F. C., Head and Neck Genome Project GENCAPO, Peitl, P., Vidotto, A., Polachini, G. M., Maniglia, J. V., Carmona-Raphe, J., Cunha, B. R., Henrique, T., Souza, C. F., Teixeira, R. A. P., Fukuyama, E. E., Michaluart, P., de Carvalho, M. B., Oliani, S. M. & Tajara, E. H. Genomics and proteomics approaches to the study of cancer-stroma interactions. *BMC Med Genomics* **3,** 14 (2010).
255. Hung, P.-S., Chen, C.-Y., Chen, W.-T., Kuo, C.-Y., Fang, W.-L., Huang, K.-H., Chiu, P.-C. & Lo, S.-S. miR-376c promotes carcinogenesis and serves as a plasma marker for gastric carcinoma. *PLoS One* **12,** e0177346 (2017).
256. Liang, Y.-K., Han, Z.-D., Lu, J.-M., Liu, Z.-Z., Zhuo, Y.-J., Zhu, X.-J., Chen, J.-X., Ye, J.-H., Liang, Y.-X., He, H.-C. & Zhong, W.-D. Downregulation of ARID4A and ARID4B promote tumor progression and directly regulated by microRNA-30d in patient with prostate cancer. *J. Cell. Biochem.* **119,** 7245–7255 (2018).
257. St Croix, B., Rago, C., Velculescu, V., Traverso, G., Romans, K. E., Montgomery, E., Lal, A., Riggins, G. J., Lengauer, C., Vogelstein, B. & Kinzler, K. W. Genes expressed in human tumor endothelium. *Science* **289,** 1197–1202 (2000).
258. Chaudhary, A., Hilton, M. B., Seaman, S., Haines, D. C., Stevenson, S., Lemotte, P. K., Tschantz, W. R., Zhang, X. M., Saha, S., Fleming, T. & St Croix, B. TEM8/ANTXR1 blockade inhibits pathological angiogenesis and potentiates tumoricidal responses against multiple cancer types. *Cancer Cell* **21,** 212–226 (2012).
259. Bachran, C. & Leppla, S. H. Tumor Targeting and Drug Delivery by Anthrax Toxin. *Toxins* **8,** (2016).
260. Gutwein, L. G., Al-Quran, S. Z., Fernando, S., Fletcher, B. S., Copeland, E. M. & Grobmyer, S. R. Tumor endothelial marker 8 expression in triple-negative breast cancer. *Anticancer Res.* **31,** 3417–3422 (2011).
261. Teicher, B., Biemann, H.-P., Woude, G. V., Duesbery, N., Frankel, A., Shankara, S., Perricone, M., Yoshida, H., Kataoka, S., Guyre, C., Boutin, P., Menon, K. & Rouleau, C. The systemic administration of lethal toxin achieves a growth delay of human melanoma and neuroblastoma xenografts: Assessment of receptor contribution. *Int J Oncol.* **32,** 739-48 (2008).
262. Chen, D., Bhat-Nakshatri, P., Goswami, C., Badve, S. & Nakshatri, H. ANTXR1, a stem cell-enriched functional biomarker, connects collagen signaling to cancer stem-like cells and metastasis in breast cancer. *Cancer Res.* **73,** 5821–5833 (2013).
263. Alcalá, S., Martinelli, P., Hermann, P. C., Heeschen, C. & Sainz, B., Jr. The Anthrax Toxin Receptor 1 (ANTXR1) Is Enriched in Pancreatic Cancer Stem Cells Derived from Primary Tumor Cultures. *Stem Cells Int.* **2019,** 1378639 (2019).
264. Byrd, T. T., Fousek, K., Pignata, A., Szot, C., Samaha, H., Seaman, S., Dobrolecki, L., Salsman, V. S., Oo, H. Z., Bielamowicz, K., Landi, D., Rainusso, N., Hicks, J., Powell, S., Baker, M. L., Wels, W. S., Koch, J., Sorensen, P. H., Deneen, B., Ellis, M. J., Lewis, M. T., Hegde, M., Fletcher, B. S., St. Croix, B. & Ahmed, N. TEM8/ANTXR1-Specific CAR T Cells as a Targeted Therapy for Triple-Negative Breast Cancer. *Cancer Res.* **78,** 489–500 (2018).
265. Sotoudeh, M., Shakeri, R., Dawsey, S. M., Sharififard, B., Ahmadbeigi, N. & Naderi, M. ANTXR1 (TEM8) overexpression in gastric adenocarcinoma makes the protein a potential target of immunotherapy. *Cancer Immunol. Immunother.* **68,** 1597–1603 (2019).
266. Li, D.-Q., Nair, S. S., Ohshiro, K., Kumar, A., Nair, V. S., Pakala, S. B., Reddy, S. D. N., Gajula, R. P., Eswaran, J., Aravind, L. & Kumar, R. MORC2 signaling integrates phosphorylation-dependent, ATPase-coupled chromatin remodeling during the DNA damage response. *Cell Rep.* **2,** 1657–1669 (2012).
267. Ding, Q.-S., Zhang, L., Wang, B.-C., Zeng, Z., Zou, X.-Q., Cao, P.-B., Zhou, G.-M., Tang, M., Wu, L., Wu, L.-L., Yu, H.-G., Guo, Y. & Zhou, F.-X. Aberrant high expression level of MORC2 is a common character in multiple cancers. *Hum. Pathol.* **76,** 58–67 (2018).
268. Wang, G., Song, Y., Liu, T., Wang, C., Zhang, Q., Liu, F., Cai, X., Miao, Z., Xu, H., Xu, H., Cao, L. & Li, F. PAK1-mediated MORC2 phosphorylation promotes gastric tumorigenesis. *Oncotarget* **6,** (2015).
269. Liao, X.-H., Zhang, Y., Dong, W.-J., Shao, Z.-M. & Li, D.-Q. Chromatin remodeling protein MORC2 promotes breast cancer invasion and metastasis through a PRD domain-mediated interaction with CTNND1. *Oncotarget* **8,** 97941–97954 (2017).
270. Wang, T., Qin, Z.-Y., Wen, L.-Z., Guo, Y., Liu, Q., Lei, Z.-J., Pan, W., Liu, K.-J., Wang, X.-W., Lai, S.-J., Sun, W.-J., Wei, Y.-L., Liu, L., Guo, L., Chen, Y.-Q., Wang, J., Xiao, H.-L., Bian, X.-W., Chen,

D.-F. & Wang, B. Epigenetic restriction of Hippo signaling by MORC2 underlies stemness of hepatocellular carcinoma cells. *Cell Death Differ.* **25,** 2086–2100 (2018).
271. Zhang, F.-L., Cao, J.-L., Xie, H.-Y., Sun, R., Yang, L.-F., Shao, Z.-M. & Li, D.-Q. Cancer-Associated MORC2-Mutant M276I Regulates an hnRNPM-Mediated CD44 Splicing Switch to Promote Invasion and Metastasis in Triple-Negative Breast Cancer. *Cancer Res.* **78,** 5780–5792 (2018).
272. Lin, J., Lee, J. H. J., Paramasivam, K., Pathak, E., Wang, Z., Pramono, Z. A. D., Lim, B., Wee, K. B. & Surana, U. Induced-Decay of Glycine Decarboxylase Transcripts as an Anticancer Therapeutic Strategy for Non-Small-Cell Lung Carcinoma. *Mol. Ther. Nucleic Acids* **9,** 263–273 (2017).
273. Woo, C. C., Kaur, K., Chan, W. X., Teo, X. Q. & Lee, T. H. P. Inhibiting Glycine Decarboxylase Suppresses Pyruvate-to-Lactate Metabolism in Lung Cancer Cells. *Front. Oncol.* **8,** 196 (2018).
274. Jäger, K., Larribère, L., Wu, H., Weiss, C., Gebhardt, C. & Utikal, J. Expression of Neural Crest Markers GLDC and ERRFI1 is Correlated with Melanoma Prognosis. *Cancers* **11,** (2019).
275. Gou, H., Liang, J. Q., Zhang, L., Chen, H., Zhang, Y., Li, R., Wang, X., Ji, J., Tong, J. H., To, K.-F., Sung, J. J. Y., Chan, F. K. L., Fang, J.-Y. & Yu, J. TTPAL Promotes Colorectal Tumorigenesis by Stabilizing TRIP6 to Activate Wnt/β-Catenin Signaling. *Cancer Res.* **79,** 3332–3346 (2019).
276. Kondelin, J., Salokas, K., Saarinen, L., Ovaska, K., Rauanheimo, H., Plaketti, R.-M., Hamberg, J., Liu, X., Yadav, L., Gylfe, A. E., Cajuso, T., Hänninen, U. A., Palin, K., Ristolainen, H., Katainen, R., Kaasinen, E., Tanskanen, T., Aavikko, M., Taipale, M., Taipale, J., Renkonen-Sinisalo, L., Lepistö, A., Koskensalo, S., Böhm, J., Mecklin, J.-P., Ongen, H., Dermitzakis, E. T., Kilpivaara, O., Vahteristo, P., Turunen, M., Hautaniemi, S., Tuupanen, S., Karhu, A., Välimäki, N., Varjosalo, M., Pitkänen, E. & Aaltonen, L. A. Comprehensive evaluation of coding region point mutations in microsatellite-unstable colorectal cancer. *EMBO Mol. Med.* **10,** (2018).
277. Mirabelli-Primdahl, L., Gryfe, R., Kim, H., Millar, A., Luceri, C., Dale, D., Holowaty, E., Bapat, B., Gallinger, S. & Redston, M. Beta-catenin mutations are specific for colorectal carcinomas with microsatellite instability but occur in endometrial carcinomas irrespective of mutator pathway. *Cancer Res.* **59,** 3346–3351 (1999).
278. Rajagopalan, H., Bardelli, A., Lengauer, C., Kinzler, K. W., Vogelstein, B. & Velculescu, V. E. Tumorigenesis: RAF/RAS oncogenes and mismatch-repair status. *Nature* **418,** 934 (2002).
279. Corso, G., Pascale, V., Flauti, G., Ferrara, F., Marrelli, D. & Roviello, F. Oncogenic mutations and microsatellite instability phenotype predict specific anatomical subsite in colorectal cancer patients. *Eur. J. Hum. Genet.* **21,** 1383–1388 (2013).
280. Jonchere, V., Marisa, L., Greene, M., Virouleau, A., Buhard, O., Bertrand, R., Svrcek, M., Cervera, P., Goloudina, A., Guillerm, E., Coulet, F., Landman, S., Ratovomanana, T., Job, S., Ayadi, M., Elarouci, N., Armenoult, L., Merabtene, F., Dumont, S., Parc, Y., Lefèvre, J. H., André, T., Fléjou, J.-F., Guilloux, A., Collura, A., de Reyniès, A. & Duval, A. Identification of Positively and Negatively Selected Driver Gene Mutations Associated With Colorectal Cancer With Microsatellite Instability. *Cell Mol Gastroenterol Hepatol* **6,** 277–300 (2018).
281. Chapman, P. B., Hauschild, A., Robert, C., Haanen, J. B., Ascierto, P., Larkin, J., Dummer, R., Garbe, C., Testori, A., Maio, M., Hogg, D., Lorigan, P., Lebbe, C., Jouary, T., Schadendorf, D., Ribas, A., O'Day, S. J., Sosman, J. A., Kirkwood, J. M., Eggermont, A. M. M., Dreno, B., Nolop, K., Li, J., Nelson, B., Hou, J., Lee, R. J., Flaherty, K. T., McArthur, G. A. & BRIM-3 Study Group. Improved survival with vemurafenib in melanoma with BRAF V600E mutation. *N. Engl. J. Med.* **364,** 2507–2516 (2011).
282. Diaz, L. A., Jr, Williams, R. T., Wu, J., Kinde, I., Hecht, J. R., Berlin, J., Allen, B., Bozic, I., Reiter, J. G., Nowak, M. A., Kinzler, K. W., Oliner, K. S. & Vogelstein, B. The molecular evolution of acquired resistance to targeted EGFR blockade in colorectal cancers. *Nature* **486,** 537–540 (2012).
283. Cereda, M., Gambardella, G., Benedetti, L., Iannelli, F., Patel, D., Basso, G., Guerra, R. F., Mourikis, T. P., Puccio, I., Sinha, S., Laghi, L., Spencer, J., Rodriguez-Justo, M. & Ciccarelli, F. D. Patients with genetically heterogeneous synchronous colorectal cancer carry rare damaging germline mutations in immune-related genes. *Nat. Commun.* **7,** 12072 (2016).
284. Di, J., Yang, H., Jiang, B., Wang, Z., Ji, J. & Su, X. Whole exome sequencing reveals intertumor heterogeneity and distinct genetic origins of sporadic synchronous colorectal cancer. *Int J Cancer* **142,** 927–939 (2018).
285. Wang, X., Fang, H., Cheng, Y., Li, L., Sun, X., Fu, T., Huang, P., Zhang, A., Feng, Z., Li, C., Huang, X., Li, G., Du, P., Yang, H., Fang, X., Li, F., Gao, Q. & Liu, B. The molecular landscape of synchronous colorectal cancer reveals genetic heterogeneity. *Carcinogenesis* **39,** 708–718 (2018).
286. Perea, J., García, J. L., Corchete, L., Lumbreras, E., Arriba, M., Rueda, D., Tapial, S., Pérez, J., Vieiro,

V., Rodríguez, Y., Brandáriz, L., García-Arranz, M., García-Olmo, D., Goel, A., Urioste, M. & Sarmiento, R. G. Redefining synchronous colorectal cancers based on tumor clonality. *Int. J. Cancer* **144,** 1596–1608 (2019).
287. Liu, Y., Zhang, J., Li, L., Yin, G., Zhang, J., Zheng, S., Cheung, H., Wu, N., Lu, N., Mao, X., Yang, L., Zhang, J., Zhang, L., Seth, S., Chen, H., Song, X., Liu, K., Xie, Y., Zhou, L., Zhao, C., Han, N., Chen, W., Zhang, S., Chen, L., Cai, W., Li, L., Shen, M., Xu, N., Cheng, S., Yang, H., Lee, J. J., Correa, A., Fujimoto, J., Behrens, C., Chow, C.-W., William, W. N., Heymach, J. V., Hong, W. K., Swisher, S., Wistuba, I. I., Wang, J., Lin, D., Liu, X., Futreal, P. A. & Gao, Y. Genomic heterogeneity of multiple synchronous lung cancer. *Nat. Commun.* **7,** 13200 (2016).
288. Fisher, R., Horswell, S., Rowan, A., Salm, M. P., de Bruin, E. C., Gulati, S., McGranahan, N., Stares, M., Gerlinger, M., Varela, I., Crockford, A., Favero, F., Quidville, V., André, F., Navas, C., Grönroos, E., Nicol, D., Hazell, S., Hrouda, D., O'Brien, T., Matthews, N., Phillimore, B., Begum, S., Rabinowitz, A., Biggs, J., Bates, P. A., McDonald, N. Q., Stamp, G., Spencer-Dene, B., Hsieh, J. J., Xu, J., Pickering, L., Gore, M., Larkin, J. & Swanton, C. Development of synchronous VHL syndrome tumors reveals contingencies and constraints to tumor evolution. *Genome Biol.* **15,** 433 (2014).
289. Jesinghaus, M., Pfarr, N., Kloor, M., Endris, V., Tavernar, L., Muckenhuber, A., von Knebel Doeberitz, M., Penzel, R., Weichert, W. & Stenzinger, A. Genetic heterogeneity in synchronous colorectal cancers impacts genotyping approaches and therapeutic strategies. *Genes Chromosom Cancer* **55,** 268–277 (2016).
290. Mlecnik, B., Bindea, G., Angell, H. K., Maby, P., Angelova, M., Tougeron, D., Church, S. E., Lafontaine, L., Fischer, M., Fredriksen, T., Sasso, M., Bilocq, A. M., Kirilovsky, A., Obenauf, A. C., Hamieh, M., Berger, A., Bruneval, P., Tuech, J.-J., Sabourin, J.-C., Le Pessot, F., Mauillon, J., Rafii, A., Laurent-Puig, P., Speicher, M. R., Trajanoski, Z., Michel, P., Sesboüe, R., Frebourg, T., Pagès, F., Valge-Archer, V., Latouche, J.-B. & Galon, J. Integrative Analyses of Colorectal Cancer Show Immunoscore Is a Stronger Predictor of Patient Survival Than Microsatellite Instability. *Immunity* **44,** 698–711 (2016).
291. Reiter, J. G., Makohon-Moore, A. P., Gerold, J. M., Heyde, A., Attiyeh, M. A., Kohutek, Z. A., Tokheim, C. J., Brown, A., DeBlasio, R. M., Niyazov, J., Zucker, A., Karchin, R., Kinzler, K. W., Iacobuzio-Donahue, C. A., Vogelstein, B. & Nowak, M. A. Minimal functional driver gene heterogeneity among untreated metastases. *Science* **361,** 1033–1037 (2018).
292. Gryfe, R., Kim, H., Hsieh, E. T., Aronson, M. D., Holowaty, E. J., Bull, S. B., Redston, M. & Gallinger, S. Tumor microsatellite instability and clinical outcome in young patients with colorectal cancer. *N. Engl. J. Med.* **342,** 69–77 (2000).
293. González, I., Goyal, B., Xia, M. D., Pai, R. K. & Ma, C. DNA mismatch repair deficiency but not ARID1A loss is associated with prognosis in small intestinal adenocarcinoma. *Hum Pathol.* **85,** 18–26 (2019).
294. Karapetis, C. S., Khambata-Ford, S., Jonker, D. J., O'Callaghan, C. J., Tu, D., Tebbutt, N. C., Simes, R. J., Chalchal, H., Shapiro, J. D., Robitaille, S., Price, T. J., Shepherd, L., Au, H.-J., Langer, C., Moore, M. J. & Zalcberg, J. R. K-ras mutations and benefit from cetuximab in advanced colorectal cancer. *N. Engl. J. Med.* **359,** 1757–1765 (2008).
295. Gulhati, P., Raghav, K., Shroff, R., Varadhachary, G., Javle, M., Qiao, W., Wang, H., Morris, J., Wolff, R. & Overman, M. J. Phase II Study of Panitumumab in RAS Wild-Type Metastatic Adenocarcinoma of Small Bowel or Ampulla of Vater. *The Oncologist* **23,** 277–e26 (2018).
296. Diosdado, B., Buffart, T. E., Watkins, R., Carvalho, B., Ylstra, B., Tijssen, M., Bolijn, A. S., Lewis, F., Maude, K., Verbeke, C., Nagtegaal, I. D., Grabsch, H., Mulder, C. J. J., Quirke, P., Howdle, P. & Meijer, G. A. High-resolution array comparative genomic hybridization in sporadic and celiac disease-related small bowel adenocarcinomas. *Clin. Cancer Res.* **16,** 1391–1401 (2010).
297. Bläker, H., Helmchen, B., Bönisch, A., Aulmann, S., Penzel, R., Otto, H. F. & Rieker, R. J. Mutational activation of the RAS-RAF-MAPK and the Wnt pathway in small intestinal adenocarcinomas. *Scand. J. Gastroenterol.* **39,** 748–753 (2004).
298. Wheeler, J. M. D., Warren, B. F., Mortensen, N. J. M., Kim, H. C., Biddolph, S. C., Elia, G., Beck, N. E., Williams, G. T., Shepherd, N. A., Bateman, A. C. & Bodmer, W. F. An insight into the genetic pathway of adenocarcinoma of the small intestine. *Gut* **50,** 218–223 (2002).
299. Cremolini, C., Di Bartolomeo, M., Amatu, A., Antoniotti, C., Moretto, R., Berenato, R., Perrone, F., Tamborini, E., Aprile, G., Lonardi, S., Sartore-Bianchi, A., Fontanini, G., Milione, M., Lauricella, C., Siena, S., Falcone, A., de Braud, F., Loupakis, F. & Pietrantonio, F. BRAF codons 594 and 596 mutations identify a new molecular subtype of metastatic colorectal cancer at favorable prognosis. *Ann.*

Oncol. **26,** 2092–2097 (2015).
300. Zheng, G., Tseng, L.-H., Chen, G., Haley, L., Illei, P., Gocke, C. D., Eshleman, J. R. & Lin, M.-T. Clinical detection and categorization of uncommon and concomitant mutations involving BRAF. *BMC Cancer* **15,** 779 (2015).
301. Heidorn, S. J., Milagre, C., Whittaker, S., Nourry, A., Niculescu-Duvas, I., Dhomen, N., Hussain, J., Reis-Filho, J. S., Springer, C. J., Pritchard, C. & Marais, R. Kinase-dead BRAF and oncogenic RAS cooperate to drive tumor progression through CRAF. *Cell* **140,** 209–221 (2010).
302. Cremona, C. A. & Behrens, A. ATM signalling and cancer. *Oncogene* **33,** 3351–3360 (2014).
303. Hugo, W., Zaretsky, J. M., Sun, L., Song, C., Moreno, B. H., Hu-Lieskovan, S., Berent-Maoz, B., Pang, J., Chmielowski, B., Cherry, G., Seja, E., Lomeli, S., Kong, X., Kelley, M. C., Sosman, J. A., Johnson, D. B., Ribas, A. & Lo, R. S. Genomic and Transcriptomic Features of Response to Anti-PD-1 Therapy in Metastatic Melanoma. *Cell* **165,** 35–44 (2016).
304. Xue, Y., Meehan, B., Fu, Z., Wang, X. Q. D., Fiset, P. O., Rieker, R., Levins, C., Kong, T., Zhu, X., Morin, G., Skerritt, L., Herpel, E., Venneti, S., Martinez, D., Judkins, A. R., Jung, S., Camilleri-Broet, S., Gonzalez, A. V., Guiot, M.-C., Lockwood, W. W., Spicer, J. D., Agaimy, A., Pastor, W. A., Dostie, J., Rak, J., Foulkes, W. D. & Huang, S. SMARCA4 loss is synthetic lethal with CDK4/6 inhibition in non-small cell lung cancer. *Nat. Commun.* **10,** 557 (2019).
305. Arteaga, C. L. & Engelman, J. A. ERBB receptors: from oncogene discovery to basic science to mechanism-based cancer therapeutics. *Cancer Cell* **25,** 282–303 (2014).
306. Jaiswal, B. S., Kljavin, N. M., Stawiski, E. W., Chan, E., Parikh, C., Durinck, S., Chaudhuri, S., Pujara, K., Guillory, J., Edgar, K. A., Janakiraman, V., Scholz, R.-P., Bowman, K. K., Lorenzo, M., Li, H., Wu, J., Yuan, W., Peters, B. A., Kan, Z., Stinson, J., Mak, M., Modrusan, Z., Eigenbrot, C., Firestein, R., Stern, H. M., Rajalingam, K., Schaefer, G., Merchant, M. A., Sliwkowski, M. X., de Sauvage, F. J. & Seshagiri, S. Oncogenic ERBB3 mutations in human cancers. *Cancer Cell* **23,** 603–617 (2013).
307. Quaas, A., Heydt, C., Waldschmidt, D., Alakus, H., Zander, T., Goeser, T., Kasper, P., Bruns, C., Brunn, A., Roth, W., Hartmann, N., Bunck, A., Schmidt, M., Buettner, R. & Merkelbach-Bruse, S. Alterations in ERBB2 and BRCA and microsatellite instability as new personalized treatment options in small bowel carcinoma. *BMC Gastroenterol.* **19,** 21 (2019).
308. Adam, L., San Lucas, F. A., Fowler, R., Yu, Y., Wu, W., Liu, Y., Wang, H., Menter, D., Tetzlaff, M. T., Ensor, J., Jr, Manyam, G., Arold, S. T., Huff, C., Kopetz, S., Scheet, P. & Overman, M. J. DNA Sequencing of Small Bowel Adenocarcinomas Identifies Targetable Recurrent Mutations in the ERBB2 Signaling Pathway. *Clin. Cancer Res.* **25,** 641–651 (2019).
309. Benson, A. B., Venook, A. P., Al-Hawary, M. M., Arain, M. A., Chen, Y.-J., Ciombor, K. K., Cohen, S. A., Cooper, H. S., Deming, D. A., Garrido-Laguna, I., Grem, J. L., Hoffe, S. E., Hubbard, J., Hunt, S., Kamel, A., Kirilcuk, N., Krishnamurthi, S., Messersmith, W. A., Meyerhardt, J., Miller, E. D., Mulcahy, M. F., Nurkin, S., Overman, M. J., Parikh, A., Patel, H., Pedersen, K. S., Saltz, L. B., Schneider, C., Shibata, D., Skibber, J. M., Sofocleous, C. T., Stoffel, E. M., Stotsky-Himelfarb, E., Willett, C. G., Johnson-Chilla, A., Gregory, K. M. & Gurski, L. A. Small Bowel Adenocarcinoma, Version 1.2020, NCCN Clinical Practice Guidelines in Oncology. *J Natl Compr Canc Netw.* **17,** 1109–1133 (2019).
310. Valencia, A. M. & Kadoch, C. Chromatin regulatory mechanisms and therapeutic opportunities in cancer. *Nat. Cell Biol.* **21,** 152–161 (2019).
311. Evans, D. J., Wasinger, A. M., Brey, R. N., Dunleavey, J. M., St Croix, B. & Bann, J. G. Seneca Valley Virus Exploits TEM8, a Collagen Receptor Implicated in Tumor Growth. *Front. Oncol.* **8,** 506 (2018).
312. Farmer, H., McCabe, N., Lord, C. J., Tutt, A. N. J., Johnson, D. A., Richardson, T. B., Santarosa, M., Dillon, K. J., Hickson, I., Knights, C., Martin, N. M. B., Jackson, S. P., Smith, G. C. M. & Ashworth, A. Targeting the DNA repair defect in BRCA mutant cells as a therapeutic strategy. *Nature* **434,** 917–921 (2005).
313. Huang, A., Garraway, L. A., Ashworth, A. & Weber, B. Synthetic lethality as an engine for cancer drug target discovery. *Nat. Rev. Drug Discov.* **19,** 23–38 (2020).
314. Franke, A. J., Skelton, W. P., Starr, J. S., Parekh, H., Lee, J. J., Overman, M. J., Allegra, C. & George, T. J. Immunotherapy for Colorectal Cancer: A Review of Current and Novel Therapeutic Approaches. *J. Natl. Cancer Inst.* **111,** 1131–1141 (2019).
315. Kopetz, S., Desai, J., Chan, E., Hecht, J. R., O'Dwyer, P. J., Maru, D., Morris, V., Janku, F., Dasari, A., Chung, W., Issa, J.-P. J., Gibbs, P., James, B., Powis, G., Nolop, K. B., Bhattacharya, S. & Saltz, L. Phase II Pilot Study of Vemurafenib in Patients With Metastatic BRAF-Mutated Colorectal Cancer. *J. Clin. Oncol.* **33,** 4032–4038 (2015).

316. Hoadley, K. A., Yau, C., Hinoue, T., Wolf, D. M., Lazar, A. J., Drill, E., Shen, R., Taylor, A. M., Cherniack, A. D., Thorsson, V., Akbani, R., Bowlby, R., Wong, C. K., Wiznerowicz, M., Sanchez-Vega, F., Robertson, A. G., Schneider, B. G., Lawrence, M. S., Noushmehr, H., Malta, T. M., Cancer Genome Atlas Network, Stuart, J. M., Benz, C. C. & Laird, P. W. Cell-of-Origin Patterns Dominate the Molecular Classification of 10,000 Tumors from 33 Types of Cancer. *Cell* **173,** 291–304.e6 (2018).