Influence of inherited genetic variants on risk, biology and prognosis of common cancers

Lingyun Xiong

Hertford College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Hilary 2021

Abstract

Cancer is a complex disease affected by both inherited risk variants and somatic alterations. Recent studies showed that inherited genetic variants and somatic driver mutations converge in cancer hallmark pathways, and that inherited genetic variants were associated with specific somatic driver mutations in tumours. However, it is unclear whether common inherited genetic variants interact with somatic driver mutations to influence cancer risk and prognosis.

In this study, I analyse multi-cancer datasets to investigate germline by somatic interaction in cancer and to evaluate the role of inherited genetic variants on risk, prognosis, nearby gene expression, response to anti-cancer therapy and adaptive immune response in common cancers. I describe how the p53 poly(A) SNP, rs78378222 (17p13.1; *TP53*), interacts with p53 mutational status in tumours to influence cancer risk, response to therapy and prognosis in common cancers. I find that a known colorectal cancer risk locus at 10p14 (rs10795668; *ATP5C1*) interacts with KRAS mutational status in tumours to associate with colorectal cancer risk. I identify two loci at 15q26.2 (rs72767781; *NR2F2-AS1*) and 15q22.2 (rs11447843; *VPS13C*) associated with prognosis in early-stage colorectal cancer at genome-wide significance, together with numerous suggestive loci. In particular, rs184423256 (8q22.3; *RRM2B*) and rs114887409 (4p16.1; *ABLIM2*) interact with KRAS mutational status in colorectal cancer to associate with clinical outcomes. Furthermore, I identify 22 loci associated with infiltrating CD8+ T cell density in colorectal cancer, which are used to infer the causal relationship between adaptive immune trait and prognosis in early-stage colorectal cancer.

In summary, this study constitutes an integrative analysis of germline by somatic interaction on risk, biology, and prognosis of common cancers. The findings provide supportive evidence that inherited genetic variants can play an active role during tumour development, thus offering insight into the genetic basis of cancer susceptibility and practical guidance for personalised disease prediction.

Acknowledgements

First and foremost, I would like to thank my supervisors, Drs. Gareth Bond and David Church. I cannot stress enough how much I value their mentorship in guiding me through the challenges and discoveries. I am very grateful that the Hui Pun Hing Foundation and the Ludwig Institute for Cancer Research provided funding for my study.

I would also like to thank Profs. Ahmed Ahmed and David Wedge, who acted as the assessors for my transfer of status and confirmation of status examinations, for their valuable time and constructive feedback on my work.

It was my great pleasure working with all the members in the Bond and Church groups as well as my collaborators across the departments and beyond. Especially, I would like to thank Drs. Ping Zhang, Enric Domingo, and Claire Palles for their analytical support and the many inspiring conversations we had together. I thank Oxford Research Computing Core and Biomedical Research Computing facility for their technical support.

Last but not least, I would like to express my sincere gratitude to my parents, Pei and Yanfen, for their sacrifice and devotion over the years. I thank Profs. Amina Eladdadi and Alan Garfinkel for their generous advice throughout my DPhil. I thank Profs. Philip Maini, Ramon Grima and Wenyi Wang for their encouragement. I also thank Yunju Yang, Chan Phetsouphanh, Dominic Saadi, Karrar Ali, Ting-Ting Hu, Stanny Goffin, Shyam Tailor, Billy Schoenberg, Renee Dale and Razvan Tanasescu for their friendship and company.

Declaration

I declare that no part of this thesis has been accepted or submitted for any degree, diploma, certificate or other qualification in this university or elsewhere. The results presented here are of my own.

The work included in Chapter 3 was published in *Cancer Research* in 2021 (listed below). The results presented in this thesis were from the original analyses and all figures shown here differ from those that were published.

Publication

Ping Zhang, Isaac Kitchen-Smith, **Lingyun Xiong**, Giovanni Stracquadanio, Katherine Brown, Philipp Richter, Marsha Wallace, Elisabeth Bond, Natasha Sahgal, Samantha Moore, Svanhild Nornes, Sarah De Val, Mirvat Surakhy, David Sims, Xuting Wang, Douglas A Bell, Jorge Zeron-Medina, Yanyan Jiang, Anderson Ryan, Joanna Selfe, Janet Shipley, Siddhartha Kar, Paul Pharoah, Chey Loveday, Rick Jansen, Lukasz F. Grochola, Claire Palles, Andrew Protheroe, Val Millar, Daniel Ebner, Meghana Pagadala, Sarah P Blagden, Tim Maughan, Enric Domingo, Ian Tomlinson, Clare Turnbull, Hannah Carter, Gareth Bond. Germline and somatic genetic variants in the p53 pathway interact to affect cancer risk, progression, and drug response. *Cancer Res.* 2021 Feb 8;canres.0177.2020. doi: 10.1158/0008-5472.CAN-20-0177.

Contents

List of Abbreviations

APC: Adenomatous Polyposis Coli
ASI: Allele-Specific Imbalance
BMP: Bone Morphogenetic Protein
CCOC: Clear Cell Ovarian Cancer
CI: Confidence Interval
CIMP: CpG Island Methylator Phenotype
CIN: Chromosomal Instability
CMS: Consensus Molecular Subtype
CN: Copy Number
CRC: Colorectal Cancer
CRE: *cis*-Regulatory Element
DNA: Deoxyribonucleic Acid
DSS: Disease-Specific Survival
EGFR: Epidermal Growth Factor Receptor
ENOC: Endometrioid Ovarian Cancer
EOC: Epithelial Ovarian Cancer
ER: Oestrogen Receptor
ER-BC: ER-negative Breast Cancer
ER+BC: ER-positive Breast Cancer
FAP: Familial Adenomatous Polyposis
FFPE: Formalin-Fixed Paraffin-Embedded
GAP: GTPase Activating Proteins
GDP: Guanosine Diphosphate
GEF: Guanosine Exchange Factor
GPCR: G-Protein Coupled Receptor
GRS: Genomic Risk Score
GTP: Guanosine Triphosphate
GWAS: Genome-Wide Association Study
HBOC: Hereditary Breast and Ovarian Cancer
HGSOC: High-Grade Serous Ovarian Cancer
HLA: Human Leukocyte Antigen

HNPCC: Hereditary Non-Polyposis Colorectal Cancer
HR: Hazard Ratio
HWE: Hardy-Weinberg Equilibrium
ICGC: International Cancer Genome Consortium
IHC: Immunohistochemistry
LD: Linkage Disequilibrium
LFS: Li-Fraumeni Syndrome
LGSOC: Low-Grade Serous Ovarian Cancer
LOH: Loss of Heterozygosity
LS: Lynch Syndrome
MAF: Minor Allele Frequency
MAPK: Mitogen-Activated Protein Kinase
MEK: Mitogen-activated protein kinase kinase
MHC: Major Histocompatibility Complex
MI: Mutual Information
MMR: Mismatch Repair
MOC: Mucinous Ovarian Cancer
MR: Mendelian Randomisation
MSI: Microsatellite Instability
MSS: Microsatellite Stable
MUT: Mutant
OR: Odds Ratio
OS: Overall Survival
PFI: Progression-Free Interval
PR: Progesterone Receptor
PRS: Polygenic Risk Score
QTLs: Quantitative Trait Loci
RFS: Recurrence-Free Survival
SNP: Single Nucleotide Polymorphism
TCF: T Cell Factor
TCGA: The Cancer Genome Atlas
TGF-β: Transforming Growth Factor-β
TIL: Tumour Infiltrating Lymphocyte
TMA: Tissue Microarray
TME: Tumour Microenvironment
TTR: Time to Recurrence
VHL: Von Hippel-Lindau disease
WT: Wild Type

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Inherited genetic variants in cancer

1.1.1 Overview

There is great genetic diversity in humans, which in part explains the phenotypic differences in the population [1]. Elucidating the inherited basis of genetic variation in human health and disease is one of the most pressing scientific challenges of this century. On average, the inherited genomes of two individuals differ at 4 to 5 million sites [2]. These genetic changes can be rare germline mutations or common genetic variants (here I refer to single-nucleotide polymorphisms, or SNPs; defined as a substitution of a single nucleotide at a specific position in the genome that is present in more than 1% of a population). So far, more than 335 million SNPs have been found across human populations [2].

Some of these inherited genetic changes are associated with diseases, including cancer. Germline mutations can cause Mendelian diseases, which are genetic disorders primarily due to alterations in one gene [3]. In particular, they play a major role in hereditary cancer cases, which account for 5 to 10% of all cancers [4]. A large proportion of germline mutation carriers develop cancers, indicating that the cancer-causing mutation has a high disease penetrance [5]. For instance, females with germline mutations in the *BRCA1* gene have an 80% lifetime risk of developing breast cancer; that is, the penetrance of *BRCA1* germline mutations is 80%

[6].

Most cancers, especially common adulthood cancers, are sporadic and predominantly caused by acquired (somatic) changes in the genome [7], where inherited genetic component account for 10-40% of variation in cancer susceptibility (discussed further below, see section 1.1.3). Similar to many other complex traits like height or skin colour, susceptibility to common cancers is influenced by more than one gene; that is, cancer is a polygenic disease [8, 9]. The genetic component of cancer susceptibility has been attributed to common inherited genetic variants that are associated with differential cancer risk [7, 10, 11]. In comparison to rare germline mutations of high penetrance, these common cancer risk variants typically have low disease penetrance (less than 1%; [12]). Collectively, these cancer risk variants contribute to the polygenic architecture of cancer aetiology [9].

In this section, I first review the role of germline mutations in hereditary cancer syndromes and the role of inherited genetic variants in common cancers. Then I discuss their importance in understanding cancer biology as well as their utility in research and the clinics.

1.1.2 Germline mutations and hereditary cancer syndromes

Germline mutations in more than 50 genes have been associated with hereditary cancer syndromes (Table 1.1), which genetically predispose carriers to substantially increased risk of developing certain types of cancer [13, 14]. Most hereditary cancer syndromes exhibit autosomal dominant inheritance: with a heterozygous germline mutation in a tumour suppressor gene (such as *TP53* in Li-Fraumeni syndrome or *RB* in retinoblastoma), tumours arise by losing the wild type allele through somatic mutations, deletions (loss of heterozygosity, or LOH), or chromosomal rearrangements (i.e., the "two-hit hypothesis"; [15, 16]).

The most common inherited cancer syndromes are hereditary breast and ovarian cancer (HBOC) and Lynch syndrome (LS) [17–19]. Below I review the genetic and clinical features of these hereditary cancer syndromes.

HBOC is an inherited genetic condition with increased risk for breast, ovarian

Syndrome	Common Features	Gene(s)	Ref.
Hereditary breast and ovarian cancer syndrome (HBOC)	Female carriers: breast, ovarian, pancreatic, melanoma, and other cancers. Male carriers: breast and prostate cancer.	*BRCA1* *BRCA2*	[17] [18]
Lynch syndrome (LS)	Colorectal cancer without extensive polyposis; other cancers include endometrial, ovarian and stomach cancers, and occasionally urothelial, hepatobiliary, and brain tumours.	*MLH1* *MSH2* *MSH6* *PMS2* *EPCAM*	[19] [20]
Familial adenomatous polyposis (FAP)	Multiple adenomatous polyps (> 100) and carcinomas of the colon and rectum; duodenal polyps and carcinomas; fundic gland polyps in the stomach; congenital hypertrophy of retinal epithelium.	*APC*	[21] [22] [20]
Li-Fraumeni syndrome (LFS)	Breast cancer, soft tissue sarcomas, osteosarcoma, and brain cancer. Rare early childhood cancers: embryonal rhabdomysarcoma, adrenal cortical cancer, and choroid plexus cancers.	*TP53*	[23] [24]
Von Hippel-Lindau disease (VHL)	Multi-system disorder characterised by abnormal growth of blood vessels in the retina, brain, spinal cord. Tumours in the adrenal gland, kidney and pancreas.	*VHL*	[25]
Cowden syndrome	Multiple tumour-like growths in the skin, intestinal tract or brain. Cancers in the breast, uterus and thyroid.	*PTEN*	[26]

Table 1.1. Genetics of representative hereditary cancer syndromes.

and other cancers. It is caused by pathogenic germline mutations in *BRCA1* or *BRCA2* (Table 1.1; [17]). BRCA1 and BRCA2 are involved in the repair of DNA damage; or trigger cell death if DNA cannot be repair [27]. Germline mutations in *BRCA1/2* thus impair DNA repair and increase cancer risk [6]. Less than 1% of the general population has pathogenic germline mutations in *BRCA1/2*; whereas up to 10% of women diagnosed with breast cancer have a germline mutation in one of the two genes, so do about 15% of women diagnosed with ovarian cancer [6]. Interestingly, somatic mutation in *BRCA1/2* were also detected in about 6% of breast cancer and 5% of ovarian cancer [28, 29]. Individuals with HBOC have a high risk for breast (40-87% lifetime risk; the range of odds ratio [OR] estimate: [6, 53]) and ovarian (10-44% lifetime risk; OR range: [5, 42]) cancers, as well as an increased risk for other cancers, such as prostate cancer, pancreatic cancer, and melanoma [6, 30–32]. Moreover, individuals with HBOC have a high risk of breast cancer before the age of 50 or before menopause, and they have an increased risk for

bilateral breast cancer (10-60% risk), a second primary tumour in a different tissue, and cancer recurrence [18].

Lynch syndrome (LS), also known as hereditary non-polyposis colorectal cancer (HNPCC), is an inherited genetic condition for colorectal, endometrial and other cancers. It is caused by germline mutations in *MLH1, MSH2, MSH6, PMS2* and *EPCAM* (Table 1.1; [33–35]). MLH1, MSH2, MSH6 and PMS2 are parts of the DNA mismatch repair (MMR) system that initiate repair of DNA mismatches in humans [36]. EpCAM is a trans-membrane glycoprotein mediating cell-cell adhesion in the epithelium [37]. DNA mismatches occur when one base is improperly paired with another base, or when there is a short addition or deletion in one strand of DNA that is not matched with the other strand. These mismatches can occur due to DNA replication errors or due to genetic recombination. Failure to recognise and repair these mismatches results in microsatellite instability (MSI) and elevated mutation rate [38], where a second hit, via mutation or LOH in the DNA MMR genes, abrogate their functions and lead to MMR deficiency (MMRd; [33]). Estimates suggest that as many as 1 in every 300 people carries pathogenic germline mutations in one of the five genes. Individuals with LS have a higher than usual risk of developing colorectal cancer (20-80% lifetime risk; tends to occur before age 50), and an increased risk of developing other types of cancers, such as endometrial (15-60%), stomach, breast, ovarian, small bowel, pancreatic, prostate, urinary tract, liver, kidney, and bile duct cancers [33].

Another important hereditary cancer syndrome is Li-Fraumeni syndrome (LFS) that is caused by germline mutations in *TP53* (Table 1.1; [23, 24]). This is a rare genetic condition: 1 in every 4500 people carries germline mutations in *TP53*. Individuals with LFS are characterised by their increased risk of development several types of cancer, including early-onset female breast cancer (before 31 years old), soft-tissue sarcomas, osteosarcomas, adrenocortical carcinomas, central nervous system tumours (as high as 90-95% lifetime risk; [24]). The cancers of LFS individuals may be diagnosed during childhood, adolescence or adulthood. In particular, individuals with LFS often develop cancer at a younger age (typically before the age of 50) than the general population and they usually development multiple types of cancer [23].

1.1.3 Inherited cancer risk variants in common cancers

As discussed above, germline mutations cause a small but important fraction of cancer, but the risk to common cancers is influenced by a multitude of factors, which can be genetic, environmental, or stochastic [39]. Cancer risk has long been recognised to have a familial component, and a sizeable fraction of cancer susceptibility is attributable to common inherited genetic variants (Fig 1.1). For example, males with a first-degree relative with prostate cancer are at approximately 2.5-fold increased lifetime risk of prostate cancer [40]. Family history is also a risk factor for other common malignancies such as breast cancer [41] and colon cancer [42]. Monozygotic twins have higher concordance rate of cancer diagnosis at the same anatomical site than dizygotic twins (Fig 1.1A; [4, 43, 44]). Twins' studies thus have estimated the extent of inherited genetic component that contributes to common cancers. As such, common inherited genetic variants were estimated to account for 42% of prostate cancer risk, 27% of breast cancer risk and 35% of colon cancer risk (i.e., genetic heritability of cancer; Fig 1.1B; [4, 43, 44]). Identifying variants underlying the genetic heritability of cancer is key to uncover the genetic basis of cancer susceptibility.

Genome-wide association studies (GWASs) take an unbiased approach to examine the genome for inherited genetic variants (Fig 1.1C) that are associated with differential cancer risk [45]. In particular, cancer GWASs compare variant allele frequencies among cancer patients (cases) to those among unaffected individuals (controls) [46, 47]. Alleles that have significantly higher frequency in cases than in controls are deemed as associated with increased cancer risk (Fig 1.1D). Because GWAS tests a large number of variants simultaneously, a large number of false-positive can arise due to chance findings. Therefore, strict statistical thresholds are usually in place to select true positive discoveries. It is a common practice to use 5e-08 as a stringent P-value cutoff in cancer GWASs [48]. To achieve this level of statistical power, large sample cohorts, usually thousands of cases and controls, are necessary.

So far, hundreds of inherited genetic variants have been associated with cancer risk, many of which were validated in independent cohorts (see `https://www.ebi.ac.uk/gwas/` for the full catalogue). Of note, the effect sizes of common inherited

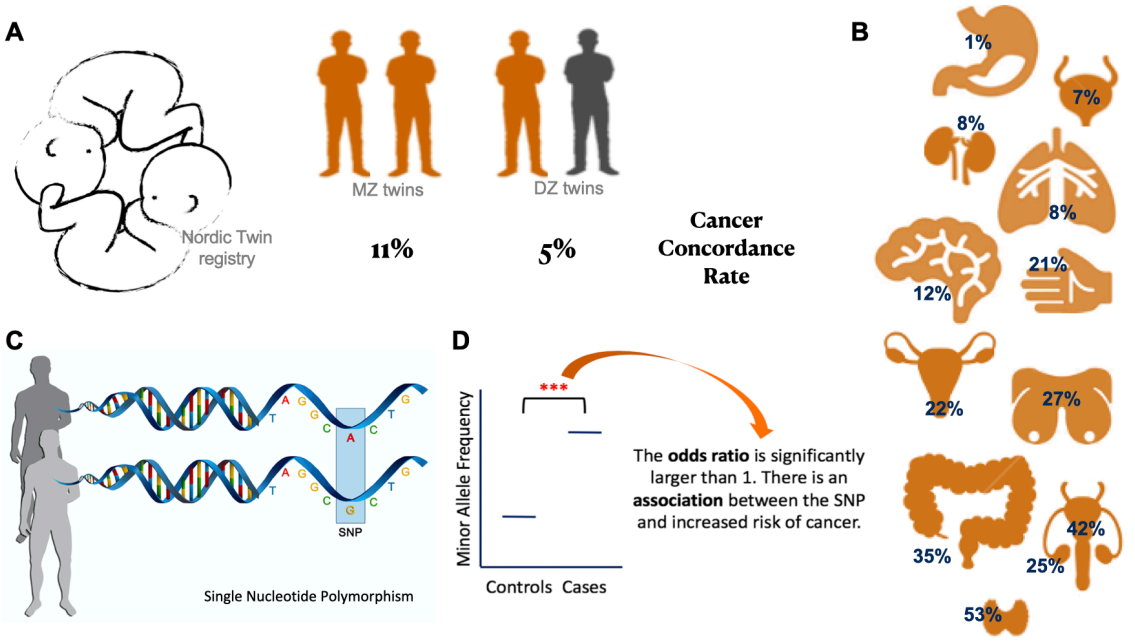

Figure 1.1. Genetic heritability of cancer and inherited cancer risk variants.

(A) A schematic of twins' studies estimating genetic heritability of cancer by measuring the concordance rates between monozygotic and dizygotic twins. (B) A list of genetic heritability estimated for different cancer types by the anatomical location. (C) A schematic of single nucleotide polymorphism (SNP), a typical example of common inherited genetic variants. (D) A schematic showing the underlying principle of cancer GWASs.

cancer risk variants are small: in general less than 1.5 [49], as opposed to the large ORs for germline mutations mentioned in section 1.1.2. Another notable feature of the inherited cancer risk variants identified thus far is that about 90% of them reside in the non-coding regions of the genome, and more than 40% are in intergenic regions [48]. As a result, interpreting the functional consequences of inherited cancer risk variants remains a challenge (discussed further below, see section 1.1.5). In this regard, insight into the biological mechanisms underlying genotype-risk associations will advance our understanding of the genes and pathways that promote cancer.

1.1.4 The missing heritability of cancer

Although a number of inherited cancer risk variants have been reported and validated, they explain a small fraction of the estimated genetic heritability of cancer [50]. For instance, breast cancer risk loci identified so far explain 18.3% of the estimated genetic heritability of breast cancer [51]; while known colorectal cancer risk loci explain only 15% of the estimated genetic heritability of colorectal cancer [52].

A number of explanations have been proposed to account for the missing heritability of cancer, such as the less explored rare and structural variants, gene-gene or gene-environment interactions. Meanwhile, it is also possible that some cancer risk variants could have been overlooked in standard cancer GWASs due to limitations in the study design.

1.1.5 Utilities of inherited genetic variants in cancer

Inherited genetic variants can influence cellular phenotype in several ways: by altering gene transcript and amino acid sequence, by disrupting transcription of non-coding RNAs, or by regulating gene expression such as influencing transcript abundance or mRNA splicing [9, 45]. Since the majority of cancer risk variants reside in the non-coding regions, research has been focused on examining the role of inherited genetic variants in regulating gene expression. In particular, certain variants can affect transcription activity locally (*cis*-expression quantitative trait loci, or *cis*-eQTLs; Fig 1.2A), which could help explain how inherited genetic variants confer cancer susceptibility [53]. Examples include a breast cancer risk allele located in an intron of *FGFR2*, which was associated with increased FGFR2 expression levels [54–56], and a prostate cancer risk allele residing in the promoter region of *MSMB*, which was associated with decreased expression level of MSMB, a biomarker for prostate cancer [48].

Inherited cancer risk variants can not only help elucidate the polygenic nature of cancer risk and uncover biological mechanisms of cancer-relevant genes or pathways (Fig 1.2A), but also hold great promise for personalised cancer risk prediction (Fig 1.2B), precision cancer prevention, as well as therapeutic development or repurposing existing anti-cancer drugs (Fig 1.2C; [45, 57]).

Although individual inherited cancer risk variants have modest effect sizes, they can be combined into polygenic risk scores (PRSs; Fig 1.2B), which can explain considerable proportion of variation in cancer risk [59] and can integrate with modifiable risk factors to improve disease prediction [60]. For instance, PRSs were used to enhance standard age-based screening strategy for colorectal cancer and to define individuals likely to gain maximal benefit from chemo-prevention [61]. The use of

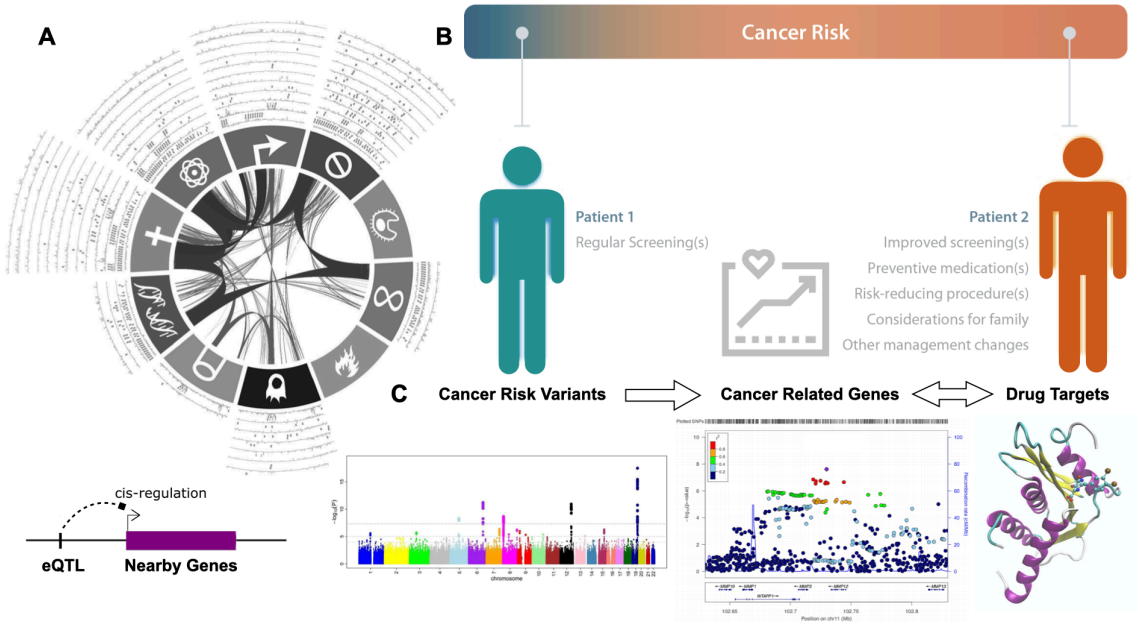

Figure 1.2. Utilities of inherited genetic variants in cancer.

(A) Inherited cancer risk variants mapped to genes involved in cancer hallmark pathways can help elucidate the polygenic architecture of cancer risk; adapted from [58]. Inherited genetic variants that regulate nearby gene expression are often termed as cis-eQTLs, which play important roles in modulating cellular phenotypes. (B) Information of inherited cancer risk variants can be combined to construct polygenic risk scores (PRSs) for personalised cancer risk predictions and to inform cancer screening and prevention programs. (C) Inherited cancer risk variants can also assist rational design of anti-cancer drugs; adapted from [45].

PRSs for refining cancer risk stratification is particularly pronounced among high-risk individuals such as *BRCA1*- and *BRCA2*-mutation carriers given that even small relative risk modifications to their baseline elevated risk translate into large differences in their absolute cancer risk [62].

In addition, a study suggested that selecting drug targets with genetic support from either rare, high-penetrance Mendelian diseases or GWASs of common complex diseases could double the success rates in clinical trials [63].

1.2 Somatic cancer driver mutations

1.2.1 Overview

As discussed above, while germline mutations and inherited genetic variants may predispose cancer risk, most cancers result from genetic changes acquired during

one's lifetime [64]. These changes are the consequence of multiple mutational processes, including the intrinsic slight infidelity of the DNA replication machinery [39], or from exposure to carcinogenic substances that damage DNA, such as chemicals in tobacco smoke and ultraviolet radiation [65–67]. Different mutational processes generate unique combinations of mutation types (i.e., mutational signatures). Note that such acquired alterations are somatic, present in only parts of the body. Somatic alterations present in cancer cells include mutations, copy number variations, chromosomal rearrangements and epigenetic silencing [65, 66, 68]. The most characterised somatic alterations in cancer are somatic mutations, which affect nucleotides in the DNA: one nucleotide may be replaced by another (point mutations), or a short stretch of nucleotides may be inserted or deleted (small indels).

It is increasingly clear that 'cancer' is a collective term for a set of diseases featuring the autonomous expansion and spread of a somatic clone [69]. The cancer clone achieves this by altering cellular processes to circumvent normal constraints on proliferation and apoptosis, to modify the local microenvironment to favour local invasion and metastasis, and to evade immune surveillance (i.e., the 'hallmarks' of cancer; [58]). No single event directs these phenotypic changes; rather, individual tumours usually exhibit diverse combinations of somatic mutations in multiple cancer hallmark pathways [70]. Some of these mutations alter cellular phenotypes, and an even smaller fraction confer a selective advantage for the cell to break away from the tight physiological constraints on growth and expansion. Mutations that provide such an advantage to the cancer clone are defined as driver mutations, as opposed to selectively neutral passenger mutations [67].

The prevalent method for identifying cancer driver mutations is based on mutation frequency [71, 72], assuming that cancer driver genes are more frequently mutated compared to the background mutation rate, which is usually adjusted for patient-specific and gene-specific mutational heterogeneity [73]. As such, a compendium of 568 significantly mutated genes have been curated [67, 74].

Many cancer driver mutations have been shown to drive tumorigenesis. Typically, missense mutations activate oncoproteins (e.g., RAS family GTPases and MYC; [75, 76]); truncating mutations inactivate tumour suppressors (e.g., APC and

NF1; [77, 78]). The best-known cancer driver genes are oncogenes *MYC, KRAS, RET, MET, KIT, EGFR* and *BRAF*, and tumour suppressor genes *TP53, RB1, PTEN, CHEK2, CDKN2A, BRCA1, BRAC2* and *APC* [67, 69]. When these cancer driver genes were considered in the context of cancer hallmark pathways, the EGFR/RAS-MAPK signalling pathway was the most frequently altered across all cancer types, and the p53 and cell cycle pathways were frequently found co-altered in multiple cancer types [79].

Although many cancer driver mutations are shared across cancer types, distinct mutational patterns exist in individual cancer types. In the following sections, I highlight features of somatic mutations in three cancer types to be examined in depth in this study: breast, ovarian and colorectal cancers (Fig 1.3), to set the stage for the discussion on potential germline by somatic interaction in common cancers in the next section.

1.2.2 Breast cancer

Breast cancer is one of the most common cancers worldwide: in 2020, there were 2.3 million cases (11.7% of total cancer cases) and 684,996 deaths (6.9% of total cancer deaths) [80]. A list of significantly mutated genes in breast cancer is provided in Table 1.2. Breast cancer is often categorised into three groups on the basis of therapies: the oestrogen receptor or progesterone receptor (ER/PR) positive group, the HER2 amplified group, and the triple-negative breast cancer group, which lack expression of ER, PR and HER2 [81]. Together with traditional clinicopathological variables including tumour size, tumour grade and nodal involvement, they were conventionally used for patient prognosis and management [81].

Meanwhile, mRNA expression profiling-based clustering has established the molecular subtypes of breast cancer: luminal/ER+, HER2-enriched and basal-like breast cancer subtypes [82, 83]. Gene expression is intimately linked to cellular phenotype and tumour behaviour. This method has also been used to identify biologically homogeneous cancer subtypes [83]. Transcriptomic and genomic analysis of breast tumours revealed a strong association between somatic mutations and these molecular subtypes, and that many mutations are specific to one subtype but not

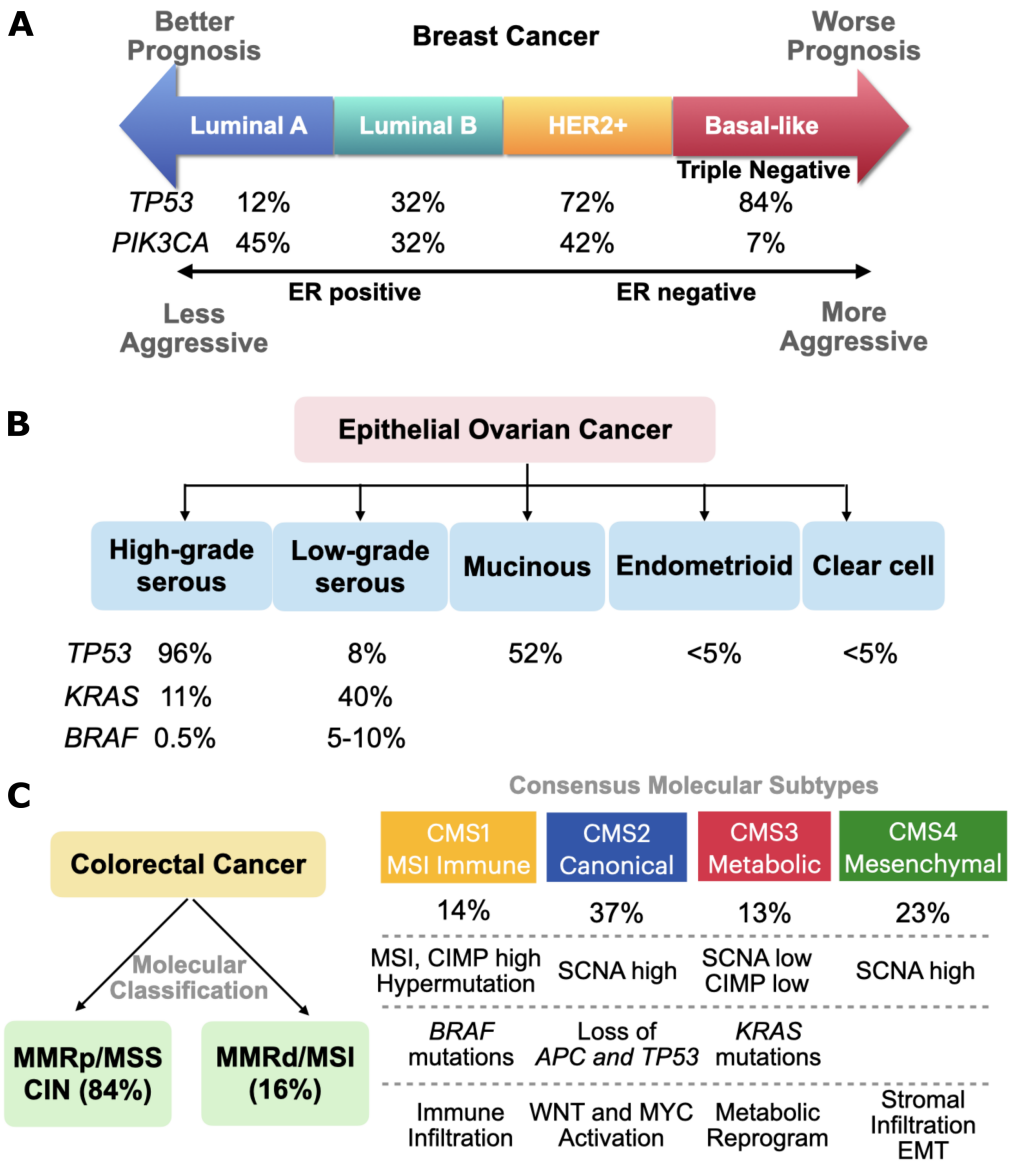

Figure 1.3. An overview of molecular subtypes of breast (A), ovarian (B) and colorectal (C) cancers.

others [28]. Overall, the mutation burden is the lowest in luminal A subtype and the highest in the basal-like subtype; the spectrum of significantly mutated genes is considerably wider in luminal tumours than in basal-like and HER2-enriched subtypes (Fig 1.3A). Luminal A tumours frequently harbour mutations in *PIK3CA* (45%), and less commonly in *TP53*(12%); luminal B tumours have mutations in *PIK3CA* and *TP53* with similar frequencies (32% each) [28]. In contrast, *TP53* mutations occurr in 84% of basal-like tumours, and 7% have mutations in *PIK3CA*, while the HER2-enriched subtype shows a mixed pattern with a high frequency of mutations in *TP53* (72%) and *PIK3CA* (42%) [28]. Intriguingly, *TP53* mutations in luminal

Rank	Breast Cancer	Ovarian Cancer	Colorectal Cancer
1	*PIK3CA*	*TP53*	*APC*
2	*TP53*	*RB1*	*TP53*
3	*GATA3*	*BRCA1*	*KRAS*
4	*MAP3K1*	*NF1*	*PIK3CA*
5	*CDH1*	*CDK12*	*ARRID1A*
6	*CBFB*	*KRAS*	*SOX9*
7	*PTEN*	*HNF1B*	*FAM123B*
8	*RUNX1*	*PTEN*	*BRAF*
9	*MLL3*	*LARP1*	*SMAD4*
10	*CDKN1B*	*BRCA2*	*B2M*

Table 1.2. Top 10 significantly mutated genes in breast, ovarian and colorectal cancers.

This list is obtained by analysing the TCGA exome sequencing data from all available tumour samples using the MutSig2CV software [72]. Genes were ranked by the false discovery rate and the frequency of non-silent mutations detected in respective cancer cohort.

tumours are mostly missense mutations, whereas those in basal-like tumours are predominantly truncating mutations [28]. Consistent with the differences in p53 mutational frequency, the p53 pathway is largely intact in luminal A tumours but often inactivated in basal-like tumours [28, 67].

Histopathological and molecular subtypes are also correlated with clinical outcomes (Fig 1.3A; [84]). For example, luminal A tumours have the best prognosis among all the subtypes. ER-negative (triple-negative and basal-like) tumours (ER-BC) are more aggressive and have poorer prognoses compared to ER-positive (luminal-A and -B) subtypes (ER+BC) [85, 86].

Considering that breast cancer subtypes differ greatly in the histopathological and molecular features, especially their profile of somatic driver mutations, they may represent cancers of distinct underlying pathophysiology. The polygenic architecture of susceptibility to each of the subtypes might be different as well. Indeed, some inherited genetic variants were associated with risk for one breast cancer subtype but not others (discussed further below, see section 1.3.1; [51]). Therefore, association studies stratified by molecular subtypes could represent a new method to uncover cancer risk variants that have been overlooked in standard cancer GWASs (see also section 1.1.4) and to refine cancer risk association. In addition, this could potentially provide insight into the biological basis of genotype-risk association.

1.2.3 Ovarian cancer

Epithelial ovarian cancer (EOC) is the one of the leading causes of cancer incidence and mortality among women: in 2020, there were 313,959 cases (1.6% of total cancer cases) and 207,252 deaths (2.1% of total cancer deaths) worldwide [80]. There are five major histopathological subtypes of EOC : high-grade serous (HGSOC), low-grade serous (LGSOC), mucinous (MOC), endometrioid (ENOC) and clear cell (CCOC) carcinoma [87]. These histological subtypes differ in lifestyle and genetic risk factors, precursor lesions, patterns of spread, somatic driver events, response to chemotherapy and prognosis [88]. Of note, serous ovarian cancers (both HGSOC and LGSOC) are thought to be derived from the Fallopian tube secretory epithelial cells [89, 90]; CCOC, ENOC carcinomas are developed from endometriosis [91, 92]; whereas MOC resembles adenocarcinoma of gastrointestinal origin [93].

HGSOC accounts for 67% of all ovarian cancers and is the most aggressive subtype [94]. LGSOC accounts for approximately 10% of total cases, occurs at a younger age than HGSOC and exhibits relative chemo-resistance but prolonged overall survival [95]. About 70% of EOC deaths occur in patients with advanced-stage HGSOC [94]. Germline mutations in *BRCA1/2* contribute to approximately 10% of HGSOC, while the majority of cases were affected by somatic mutations in cancer driver genes (Table 1.2; [94]). In particular, HGSOC is characterised by mutations in *TP53*, present in almost all HGSOC cases (96%; Fig 1.3B; [94, 96]). In contrast, mutations in *TP53* are much less frequent in LGSOC cases (8%; Fig 1.3B; [97, 98]); while p53 mutation frequency vary among MOC, ENOC and CCOC cases (Fig 1.3B; [99–101]). Besides, RB1 and PI3K/RAS-MAPK pathways were dysregulated in 67% and 45% of HGSOC cases, respectively. LGSOC tumours are also characterised by activating mutations in the RAS-MAPK pathway; KRAS and BRAF mutation frequencies in LGSOC are much higher than those in HGSOC (Fig 1.3B; [95, 102]).

Again, we have seen that ovarian cancer subtypes differ greatly in their mutational frequency of cancer driver genes. Therefore, like in breast cancer (see also section 1.2.2), association studies stratified by ovarian cancer subtypes could reveal cancer risk variants that have been overlooked by standard GWASs (see also sec-

tion 1.1.4) and to refine cancer risk association in ovarian cancer. Of note, no such stratified association study has been carried out in ovarian cancer.

1.2.4 Colorectal cancer

Colorectal cancer (CRC) is a leading cause of cancer morbidity and mortality worldwide: in 2020, there were 1.88 million cases (9.7% of total cancer cases) and 915,880 deaths (9.2% of total cancer deaths) [80]. Known risk factors for CRC include age (about 90% of CRC cases occur in people who are above 50 years old), inflammatory bowel disease (e.g., Crohn's disease or ulcerative colitis), a personal or family history of CRC or colorectal polyps, and a genetic syndrome, such as familial adenomatous polyposis (FAP) or Lynch syndrome (LS) (2-3% total CRC cases) (Table 1.1; [103, 104]). In addition, lifestyle factors may contribute to an increased risk for CRC, including lack of regular physical activity, a diet that is low in fruit/vegetable/fibre or high in fat/processed meat, overweight and obesity, alcohol consumption, as well as tobacco use [103, 104].

In addition to the small fraction of CRC with MMRd due to LS, a further 10-15% cases display MMRd due to somatic alterations, most commonly *MLH1* promoter methylation. MMRd/MSI tumours (see also section 1.1.2) are primarily located in the right colon and they are frequently associated with CpG island methylator phenotype (CIMP) and hypermutation (Fig 1.3C; [105]). The remaining 85% of CRCs retain MMR function (MMR proficient, or MMRp) and are microsatellite stable (MSS), but are characterised by the alternative mechanism of genomic instability through chromosomal instability (CIN) [105].

Similar to breast and ovarian cancers, dozens of cancer driver genes have been identified in CRC (Table 1.2), such as *APC, TP53, KRAS, PIK3CA* and *SMAD4*, notably affecting the WNT, p53, RAS-MAPK, PI3K, and the BMP/TGF-β pathways (discussed further in Chapter 4, see section 4.1.2; [20]). The most common target is the WNT pathway, which is altered in 93% of CRCs, often by biallelic inactivation of *APC* (70-80% of cases) or activating mutations in *CTNNB1* [21, 22]. Genetic alterations in the RAS-MAPK pathway are also common in CRC, both in non-hypermutated (80%) and in hypermutated (59%) tumours [105]. Specifi-

cally, about 55% of non-hypermutated tumours have somatic mutations in *KRAS* and *NRAS*, but *BRAF* mutations are enriched in hypermutated tumours with MSI and CIMPhigh status [67, 105]. A strong tendency for mutual exclusivity between mutations in *KRAS* and *BRAF* is a notable feature in both MMRp and MMRd subgroups [105]. The molecular function of KRAS and the pathophysiology of mutated KRAS will be discussed in detail in Chapter 4 (see section 4.1.3). Alterations in the p53 pathway are also frequent in CRC. For example, biallelic inactivation of *TP53* occurs in 59% of non-hypermutated cases [105]. Loss of p53 and mutations in the cell cycle checkpoint and DNA damage response kinase ATM also have a mutually exclusive pattern [67, 105].

The application of transcriptomic profiling to define molecular subtypes of breast cancer was discussed in section 1.2.2. It was also applied to CRC. As such, four consensus molecular subtypes (CMSs) have been defined, each was characterised by distinct genomic and epigenomic features (Fig 1.3C; [106]). Most tumours with MSI cluster in the CMS1 group (MSI immune subtype, 14% of early-stage tumours), which is characterised by hypermutation, hypermethylation, enrichment of *BRAF-V600E* mutations and strong infiltration of immune cells in the tumour microenvironment (TME). Tumours with CIN can be sub-classified into three groups based on gene expression patterns: CMS2 (canonical subtype, 37% of early-stage tumours); CMS3 (metabolic subtype, 13% of early-stage tumours); and CMS4 (mesenchymal subtype, 23% of early-stage tumours). WNT and MYC activation due to loss of *APC* is typical of CMS2 tumours; mutations in *KRAS* characterise CMS3 tumours, which have been linked to extensive metabolic adaptation; while CMS4 tumours typically activate pathways related to epithelial-mesenchymal transition and stemness. Notably, patients diagnosed with CMS4 tumours have significantly higher risk of distant relapse and death when compared to those with CMS2 tumours, despite that their somatic copy number (CN) variation patterns and mutation spectrum are highly comparable [107].

Among the CRC molecular subtypes, tumours in the CMS1 group and MMRd tumours feature an increased mutational load due to impaired DNA MMR and are associated with high levels of tumour infiltrating lymphocytes (TILs) [106].

This subtype is associated with decreased risk of metastasis, favourable clinical outcomes, and better prognosis [108]. More generally, adaptive immune response plays an important role in CRC development (discussed further in Chapter 5, see section 5.1.2), such as selecting immune evasive clones, inducing immunosuppression, or even mediating metastasis [109, 110]. The presence and extent of anti-tumour adaptive immune response is a key prognostic factor in CRC: immunological profile or 'immunoscore' (i.e., a summary of the types, density, and location of adaptive immune cells) of tumour samples predicted patient survival in multiple studies [111–114].

Given that molecular subtypes of CRC are characterised by distinct somatic mutations (Fig 1.3C) and they behave very differently in terms of clinical outcomes and prognosis, association studies stratified by CRC subtypes or specific somatic driver mutations could potentially reveal cancer risk variants that have been overlooked by standard GWASs (see also section 1.1.4) and refine cancer risk association, as in breast cancer (see also section 1.2.2; [51]). Furthermore, there is much unexplained variation in anti-tumour adaptive immune response and prognosis in CRC. Therefore, an evaluation of the role of inherited genetic variants on anti-tumour adaptive immune response and prognosis in CRC is also warranted.

1.3 Germline by somatic interactions in cancer

The previous sections have considered the role of inherited genetic variants and somatic driver mutations in cancer separately. However, emerging evidence indicates that inherited genetic variants could interact with somatic driver events to influence tumour development. These 'germline by somatic interactions' can manifest in cancer subtypes, tumour molecular profiles, tumours' response to treatment as well as clinical outcomes [115, 116]. In this section, I review literature on germline by somatic interactions in cancer (Fig 1.4).

1.3.1 Inherited genetic variants and cancer subtypes

As discussed in section 1.1.2, individuals with hereditary cancer syndromes often develop tumours with characteristic histopathological or molecular features compared

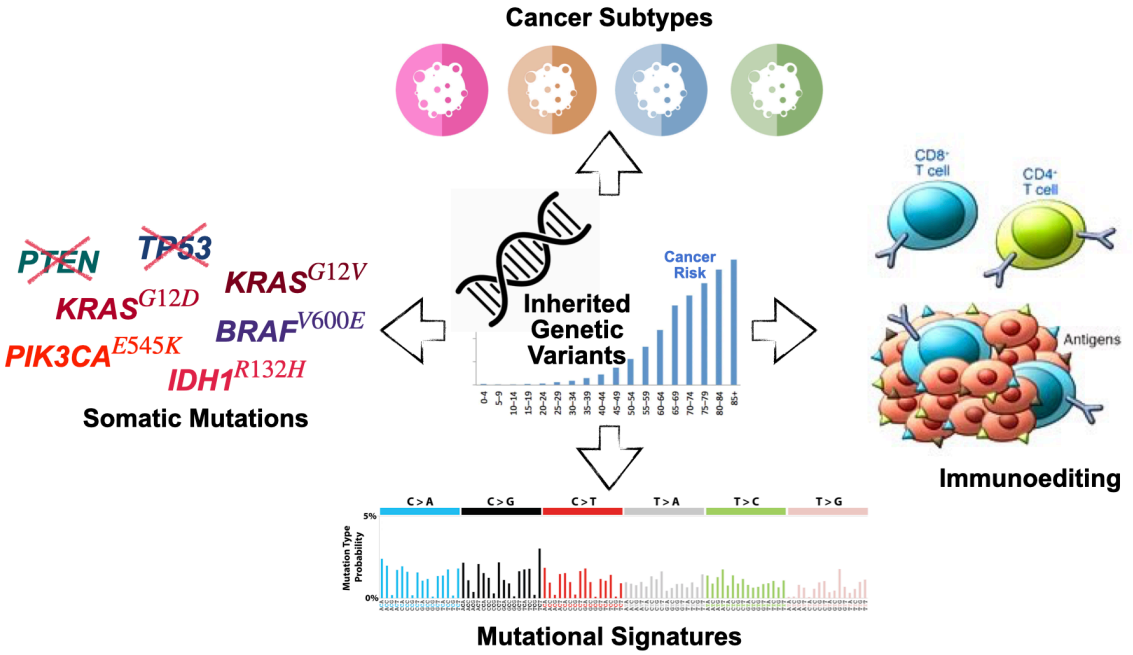

Figure 1.4. An overview of germline by somatic interactions in cancer.

to those arising in the general population. For example, *BRCA1*-related breast tumours differ from non-*BRCA1* tumours in terms of histological features: tumours of germline *BRCA1* mutation carriers are more likely to be high-grade triple-negative cases of basal-like subtype [117–119]. As for cases with Lynch syndrome, germline mutation in DNA MMR genes is closely linked to the MSI-high phenotype [35], and CMS1 subtype in CRC [19].

Common inherited genetic variants were also associated with specific cancer subtypes. For instance, a dozen of inherited genetic variants have been associated with breast cancer subtypes differing in their ER status (Table 1.3; [30, 120]). There were also variants specifically associated with HER2-enriched subtype [121] or triple-negative breast cancer (Table 1.3; [122]). A recent genome-wide association study on breast cancer used both standard and novel methodologies accounting for tumour heterogeneity by ER, PR and HER2 status, as well as tumour grade [51]. Four loci showed association with risk in opposite directions for luminal A-like and triple-negative subtypes (Table 1.3), which may modulate enhancer activity in a cell type-specific manner [51]. Interestingly, this study revealed that luminal A-like and triple-negative breast cancer subtypes have a genetic correlation of 0.46 (standard error = 0.05), indicating a low level of similarity in the genetic basis between

these subtypes. Moreover, this study also showed that the proportions of genetic heritability explained by the 210 known breast cancer risk loci differ by subtypes: 54.22% for luminal A-like versus 37.63% for triple-negative breast cancers [51].

RSID	Nearby Gene(s)	Cancer Subtype	OR (95% CI)	Ref.
rs3732568	*EPHB1*	ER-positive	**OR=0.71 (0.55-0.91)**	[120]
		ER-negative	OR=0.80 (0.58-1.11)	
rs4707795	*EPHA7*	ER-positive	**OR=2.15 (1.06-4.37)**	[120]
		ER-negative	OR=0.43 (0.10-1.96)	
rs12765929	*BMPR1A*	ER-positive	OR=0.93 (0.59-1.48)	[120]
		ER-negative	**OR=0.79 (0.62-1.00)**	
rs9836340	*EPHA3*	ER-positive	OR=0.88 (0.61-1.26)	[120]
		ER-negative	**OR=1.53 (1.02-2.31)**	
rs2981582	*FGFR2*	ER-positive	**OR=1.35 (1.17-1.56)**	[30]
		ER-negative	**OR=0.91 (0.85-0.98)**	
rs3803662	*TOX3*	ER-positive	**OR=1.25 (1.06-1.46)**	[30]
	TNRC9	ER-negative	OR=1.05 (0.97-1.13)	
rs4973768	*SLC4A7*	ER-positive	**OR=1.17 (1.03-1.33)**	[30]
	NEK10	ER-negative	OR=0.99 (0.93-1.06)	
rs3817198	*LSP1*	ER-positive	OR=1.07 (0.93-1.22)	[30]
		ER-negative	**OR=1.07 (1.00-1.15)**	
rs2046210	*ESR1*	Triple-negative	**OR=1.29 (1.17-1.42)**	[122]
rs12662670	*ESR1*	Triple-negative	**OR=1.33 (1.15-1.53)**	[122]
rs999737	*RAD51L1*	Triple-negative	**OR=0.86 (0.80-0.93)**	[122]
rs3803662	*TOX3*	Triple-negative	**OR=1.17 (1.09-1.26)**	[122]
rs8170	*C19orf62*	Triple-negative	**OR=1.27 (1.17-1.38)**	[122]
rs8100241	*C19orf62*	Triple-negative	**OR=0.84 (0.78-0.90)**	[122]
rs78378222	*TP53*	Luminal A-like	**OR=1.13 (1.05-1.20)**	[51]
		Triple-negative	**OR=0.67 (0.57-0.80)**	
rs206435	*AP005209.1*	Luminal A-like	**OR=1.03 (1.01-1.05)**	[51]
		Triple-negative	**OR=0.95 (0.92-0.98)**	
rs141526427	*RPS11P1*	Luminal A-like	**OR=0.96 (0.94-0.98)**	[51]
		Triple-negative	**OR=1.04 (1.01-1.08)**	
rs6065254	*MAFB*	Luminal A-like	**OR=0.96 (0.94-0.97)**	[51]
		Triple-negative	**OR=1.04 (1.01-1.07)**	

Table 1.3. Common inherited genetic variants reported to associate with breast cancer subtype-specific risk.

Of note, miRNA variants were marginally associated with CRC subtypes differing in MMR status [123]; a variant in the *MGMT* promoter region was associated with increased risk of developing *MGMT*-methylated CRC but decreased risk for *MGMT*-unmethylated CRC [124].

1.3.2 Inherited genetic variants and mutational signatures

Inherited genetic variants were also associated with specific mutational patterns in tumours. Notably, inherited genetic variants affecting genes involved in DNA repair or genome integrity maintenance were correlated with hypermutability in tumours [125] and characteristic mutational signatures (see section 1.2; [115]). For example, germline mutations in *BRCA1* were associated with COSMIC mutational signature 3 and tandem duplication in tumours; germline mutations in *BRCA2* were associated with small structural deletions [126, 127]. Variants in the melanocortin 1 receptor gene (*MC1R*) were associated with somatic C>T mutations in melanoma [128]. This mutational signature is linked to sun exposure, which relates to MC1R's function in regulating skin pigmentation. A recent pan-cancer study revealed that multiple variants at 22q13.1 were associated with APOBEC mutagenesis, including a known risk variant for breast cancer [69]. These findings suggest that the link between inherited genetic variants and mutational signatures in tumours could be an additional biological explanation for their association with cancer risk.

1.3.3 Inherited genetic variants and somatic driver mutations

In addition to the association with cancer subtypes and mutational signatures, some inherited genetic variants were associated with specific somatic events in tumours (Table 1.4), either locally (within 2Mb of the same chromosome; *in cis*) or distally (more than 2Mb away on the same chromosome, or on different chromosomes; *in trans*).

Local impact of inherited genetic variants on somatic events in tumours includes loss of heterozygosity (LOH), allele-specific imbalance (ASI), and somatic mutations occurring at the same locus as the inherited lesion (see also section 1.2.1). As for loci harbouring oncogenes, tumours often exhibit CN gains of the activating mutant alleles [129–131]. In addition, some inherited genetic variants were associated with somatic mutations in the same gene. One example is rs12343867 and somatic mutation in *JAK2* in myeloproliferative neoplasms (rare blood cancers), where the risk haplotype co-occurred preferentially with *JAK2-V617F* mutation [132]. The T

allele of rs712829 in *EGFR* promoter region was associated with increased cancer risk and favoured *EGFR* exon 19 microdeletion in non-small cell lung cancer [133]. Likewise, the risk allele C of rs10932384 in *ERBB4* was associated with HER4 somatic mutations in clear cell renal cell carcinomas [134].

Cancer Type	**RSID**	**Nearby Gene(s)**	**Mutated Gene**	**Association**	**Ref.**
Breast	rs252913	*MAP3K1*	*PIK3CA*	OR=2.97 (1.17-7.75)	[135]
	rs331499	*SETD9*		OR=1.75 (1.11-2.77)	
Gastric	rs2285947	*DNAH11*	*PDGFB*	OR=0.26 (0.12-0.55)	[136]
Lung	rs712829	*EGFR*	*EGFR*	OR=4.28 (1.41-12.98)	[133]
	rs36600	*MTMR3*	*ARID1A*	OR=2.45 (1.47-4.08)	[137]
	rs2395185	*HLA-DRB9*	Cell cycle pathway genes	OR=1.56 (1.11-2.95)	[137]
	rs3817963	*BTNL2*	Cell cycle pathway genes	OR=1.58 (1.23-2.04)	[137]
		TSBP1-AS1	MAPK signalling pathway genes	OR=1.54 (1.19-1.98)	[137]
Melanoma	rs12203592	*IRF4*	*BRAF V600E*	OR=0.59 (0.43-0.79)	[138]
	rs132985	*PLA2G6*	*BRAF V600E*	OR=1.32 (1.05-1.67)	
			BRAF (others)	OR=1.82 (1.11-2.98)	
	rs738322	*PLA2G6*	*BRAF V600E*	OR=1.28 (1.02-1.60)	
			BRAF (others)	OR=1.66 (1.01-2.72)	
Multiple	rs8051518	*RBFOX1*	*SF3B1*	8-fold increase in incidence	[139]
	rs25673	*AP3D1*	*PTEN*	4-fold increase in incidence	
Kidney	rs10932384	*ERBB4*	*ERBB4*	P=0.003	[134]

Table 1.4. Inherited genetic variants that have been reported to associate with somatic driver mutations in common cancers.

A number of inherited genetic variants were also associated with distal somatic events. One of the first systematic analyses of inherited genetic variants and somatic events identified dozens of loci associated with specific somatic mutations in 20 cancer-related genes across cancer types [139]. For example, the risk allele of rs25673 (19p13.13) was associated with 4-fold increased likelihood of inactivating *PTEN* (10q23.31) somatically. The same variant was associated with the expression of nearby genes *GNA11* and *STK11*, both of which are involved in the PIK3CA/mTOR pathway and inhibited by PTEN. Interactions between inherited genetic variants and somatic mutations have also been reported in individual cancer types. In ER+BC, rs252913 and rs331499 (5q11.2), located in the *MAP3K1/SETD9*

gene body, were associated with somatic PIK3CA mutational status (Table 1.4; [135]). In lung cancer, rs36600 (22q12.2) was associated with somatic mutations in *ARID1A* (1p36.11), and multiple variants were associated with somatic mutations in the cell cycle and MAPK pathway (Table 1.4; [137]). In melanoma, variants in *MC1R* were associated with *BRAF V600E* cases among individuals with darker phenotype, but inversely associated with *BRAF V600K* cases [140]; rs12203592 in *IRF4* was associated with both *BRAF V600E* and *BRAF V600K* mutations, and rs132985 in *PLA2G6* was associated with *BRAF V600E* mutation but not with *BRAF V600K* (Table 1.4; [138]). Furthermore, inherited genetic variants on chromosomes 6 and 19 were associated with the occurrence of specific mutations within *KRAS* codon 61 in mouse lung tumours [141].

1.3.4 Inherited genetic variants and immune response

The importance of the anti-tumour adaptive immune response during tumour development has been discussed above in section 1.2.4. Perhaps unsurprisingly, given the proven importance of inherited genetics on autoimmune disease [142–150], inherited genetic variants are also likely to associate with differential anti-tumour adaptive immune response. Though the number of studies is at present limited, emerging evidence suggests that inherited genetic variants may influence the expression of genes involved in immune pathways and play a role in immunoediting in tumours (Fig 1.4): patient-specific HLA genotypes were found to restrict antigen presentation of oncogenic mutations, thus influencing clonal selection for mutations that go undetected by immune surveillance in the TME (discussed further in chapter 5, see section 5.1.3; [151, 152]). Recent studies also suggested that inherited genetic variants are associated with immune gene expression in cancer cells, which in turn predicted immune cell abundance in the tumour microenvironment (discussed further in Chapter 5, see section 5.1.4; [153–155]).

On the other hand, inherited genetic variants can also contribute to tumour-extrinsic immune modulation. For instance, common inherited genetic variants explained up to 21% of variance in leukocyte indices [156]; variants were associated with immune infiltration in normal tissues [157]. Intriguingly, a variant in the *FGFR4* gene (CD334) was shown to enhance signal transducer and activator of

transcription 3 (STAT3) signalling and impede the tumour infiltration of CD8+ T cells *in vivo* [158]; a breast cancer risk variant rs3903072 was associated with CTSW expression levels in tumour infiltrating lymphocytes but not in cancer cells; CTSW expression level is positively correlated with breast cancer prognosis [159].

Despite recent development in elucidating the role of inherited genetic variants on anti-tumour adaptive immune response, much of the topic remains unexplored. For example, are inherited genetic variants directly associated with the abundance of tumour infiltrating immune cells? Furthermore, it is potentially important to investigate how the relationship of inherited genetic variants and anti-tumour adaptive immune response might influence disease progression and patient survival outcomes.

1.4 Aims of the study

Taken together, accumulating evidence suggests that inherited genetic variants could play an actively role during tumour development, potentially influencing somatic driver mutations and shaping the molecular and phenotypic profile of the tumour. Considering that inherited genetic variants can predispose carriers to somatic driver mutations that are characteristic of specific cancer subtypes, it is plausible that inherited genetic variants can be associated with risk for one cancer subtype, but not others. Because conventional cancer GWASs mixed cancer cases of different molecular subtypes, inherited risk variants that are subtype-specific were likely to have gone undetected in prior studies. This could potentially explain part of the missing heritability of cancer [50], a hypothesis that warrants a thorough investigation.

Therefore, the aims of this study are

- to define inherited genetic variants that render carriers susceptible to cancers of specific molecular profile,
- to further investigate the interactions between inherited genetic variants and somatic driver mutations in common cancers, and
- broadly to evaluate the role of inherited genetic variants on cancer risk, prognosis, and adaptive immune response in tumours.

The thesis is therefore structured as follows:

Chapter 2 details the data and methods used in the analyses presented in this thesis.

Chapter 3 describes how I explore public multi-cancer datasets to define inherited genetic variants that track with p53 somatic mutations in tumours, to demonstrate that common inherited genetic variants interact with p53 mutational status in tumours to influence cancer risk, response to therapy, and prognosis in common cancers.

Chapter 4 extends the germline by somatic interaction to the RAS-MAPK oncogenic pathway, and in particular, to test whether inherited genetic variants interact with KRAS hotspot mutations to influence CRC risk and prognosis.

Chapter 5 reports the investigation of inherited genetic variants predictive of patient prognosis and anti-tumour adaptive immune response, the potential interactions of these variants with somatic driver mutations, and the causal relationship between adaptive immune traits and clinical outcomes in CRC.

Chapter 2

Methods

2.1 Study cohorts

2.1.1 The TCGA pan-cancer cohort

2.1.1.1 Genotype imputation

Genotype data was obtained and filtered as described in [139]. Briefly, genotype calls from the Birdsuite-processed Affymetrix 6.0 SNP arrays for matched normal samples [160] were downloaded from the TCGA data portal (`https://gdc-portal.nci.nih.gov/`); low confidence SNP calls (error rate $> 10\%$) were set to missing; individuals and SNPs with $< 95\%$ call rate and SNPs with MAF $< 1\%$ were filtered; and untyped genotypes were imputed using the 1000 Genomes Project phase 1 v3 data as the reference panel on the secure Michigan Imputation Server [161]. Principal component analysis was performed with the combined TCGA and HapMap Phase III populations to identify hidden population substructure that could give rise to spurious associations. Samples that did not cluster tightly with Europeans populations were removed.

2.1.1.2 Somatic mutations in TCGA tumours

The mutation profiles of *TP53* in TCGA primary tumours were downloaded from the TCGA data portal (`https://gdc-portal.nci.nih.gov/`). The mutation calls (1,245 unique mutations in 3,956 tumours) were classified into pathogenic (1,097

unique mutations in 3,895 tumours), benign (143 unique mutations in 148 tumours), or unclear (5 unique mutations in 5 tumours) based on previously curated datasets [162, 163]. The pathogenic missense mutations in *TP53* were further annotated as loss of function (LOF), or oncogenic (either dominant negative effect [DNE] or gain of function [GOF]) as described in section 2.2.1. Tumours without p53 mutations were assigned as wild type (WT); tumours with at least one pathogenic p53 mutations were assigned as mutant (MUT). Furthermore, tumours with only benign and/or unclear p53 mutations were assigned as benign/unclear; tumours with only pathogenic p53 missense mutations were assigned as missense mutant; tumours with only oncogenic p53 missense mutations were assigned as oncogenic missense mutant. The copy number (CN) profiles of *TP53* in TCGA primary tumours were retrieved from the Broad GDAC Firehose (`https://gdac.broadinstitute.org/`) through the fbget tool (v0.1.11; released on Oct 31, 2017). Similarly, somatic mutations and CN profiles in *KRAS* were also obtained from the TCGA data portal and the Broad GDAC Firehose, respectively.

2.1.2 Cohorts for the KRAS-stratified association studies

The VQ2 cohort in the Phase I study comprised 1,884 CRC cases (58% male, mean age 63 year, SD=10) from the UK-based VICTOR and QUASAR2 clinical trials and 2,673 individuals from the 1958 Birth Cohort in Wellcome Case Control Consortium 2 (WTCCC2). In brief, QUASAR2 was an open-label, randomised, controlled trial investigating the addition of the anti-angiogenic agent, bevacizumab, to capecitabine in the adjuvant treatment of high-risk stage II or III CRC [164]. In the VICTOR trial, patients with histologically proven stage II or III CRC were randomised to the COX-2 inhibitor rofecoxib or placebo, after completion of anti-cancer therapy (including surgery, radiotherapy and chemotherapy) [165].

The COIN cohort in the Phase II study included 2,244 cases (64% male, mean age 61 years, SD=10) participating in two independent Medical Research Council clinical trials of advanced/metastatic CRC: COIN and COIN-B [166, 167]. For controls, genotype information of 2,800 individuals from the UK Blood Service Control Group in WTCCC2 were utilised.

2.1.2.1 Genotyping and quality controls

Cases in the VQ2 cohort were genotyped using Illumina Hap300, Hap370, 610K or Omni2.5M arrays according to the manufacturer's recommendations (Illumina, San Diego, CA, USA). Genotyping of controls from the 1958 Birth Cohort was performed as part of the WTCCC2 study on Illumina 660W array. SNPs probed in all 5 genotyping platforms were used for genotype imputation. Cases in the COIN cohort were genotyped using Affymetrix Axiom Array according to the manufacturer's recommendations (Affymetrix, Santa Clara, CD 95051, USA), using duplicated samples. Sequencing of selected SNPs in a subset of samples were performed to confirm genotyping accuracy. Genotype concordance was $> 99\%$ for all SNPs probed. Controls in the Blood Service Group were genotyped as part of the WTCCC2 study on Affymetrix 6.0 array. SNPs probed in both Affymetrix platforms were used for genotype imputation.

Stringent quality control measures were applied to each sample cohort following standard protocols. Individuals with low SNP call rate ($< 98\%$) were removed. For apparent first-degree relative pairs (IBD > 0.185), I excluded the control from a case-control pair; otherwise, I excluded the individual with the lower call rate. Population stratification was examined using EIGENSTRAT [168], and individuals evaluated to be non-European ancestry (using the HapMap release 23a CEU, JPT/CHB and YRI populations as a reference) were removed from analysis. SNPs with a call rate $< 98\%$ were excluded as well as those with a MAF $< 1\%$ or displaying significant deviation from Hardy-Weinberg equilibrium ($P <$ 1e-20). SNPs in regions of long-range linkage disequilibrium, highly polymorphic regions, and sex chromosomes were also excluded. All genotype quality control steps were conducted in PLINK v1.9 [169].

2.1.2.2 Genotype Imputation

Prediction of the untyped SNPs in each individuals was carried out using SHAPEIT v2.837 [170] and IMPUTE v2.3.2 [171], utilising a merged reference panel combining data from 1000 Genomes Project (phase 3, December 2013 release; [2]) and UK10K (April 2014 release; [172]). Post-imputation quality controls were performed using

qctool v2 (`https://www.well.ox.ac.uk/~gav/qctool_v2/`). Rare variants (MAF $< 1\%$) and poorly imputed SNPs (info score < 0.8) were excluded. For the imputed SNPs, genotype probabilities were converted to discrete genotypes using gtool v0.7.5 (`https://www.well.ox.ac.uk/~cfreeman/software/gwas/gtool.html`) with the default settings. A minimum probability threshold of 0.9 was required for specifying sample genotypes. Otherwise, genotype would be annotated as missing. Genetic linkage of SNPs was determined by calculating D' ad R^2 using 1000 Genomes Project data (phase 3, December 2013 release, non-Finnish Europeans only) via web server LDlink [173].

I found little evidence of genomic inflation in the datasets (λ_{GC} range: [1.00, 1.01]; Fig 2.1), indicating that there were no hidden population structures in the datasets that could give rise to spurious associations.

2.1.2.3 Somatic profiling of tumours

In the VQ2 cohort, tumour DNA was extracted from Formalin-Fixed Paraffin-Embedded (FFPE) tumour blocks when available and screened for mutations in *KRAS* codons 12 and 13, and *BRAF V600E* via direct Sanger sequencing as described in [174]. A customised gene panel of 82 genes, including *KRAS, BRAF* and other major CRC driver genes, were also screened for somatic mutations using the Ion Torrent PGM sequencer. Tumours from the COIN cohort were screened for mutations in *KRAS* codons 12, 13, and 61 and *BRAF* codon 600 by pyrosequencing [175]. Additionally, *KRAS* (all three codons), *BRAF V600E* (codons 594 and 600), and *NRAS* (codons 12 and 61) were screened for mutations by MALDI-TOF mass array (Sequenom, San Diego, CA, USA). In both cohorts, KRAS mutational status in tumours was assigned based on the presence of hotspot missense mutations in codon 12, 13, and 61. An unknown status was assigned to cases whose somatic information was not available.

2.1.3 Cohorts for the survival analyses

Details of the QUASAR2 and SCOT trials have been reported previously [164, 176]. As mentioned in 2.1.2, QUASAR2 was an open-label, randomised, controlled trial

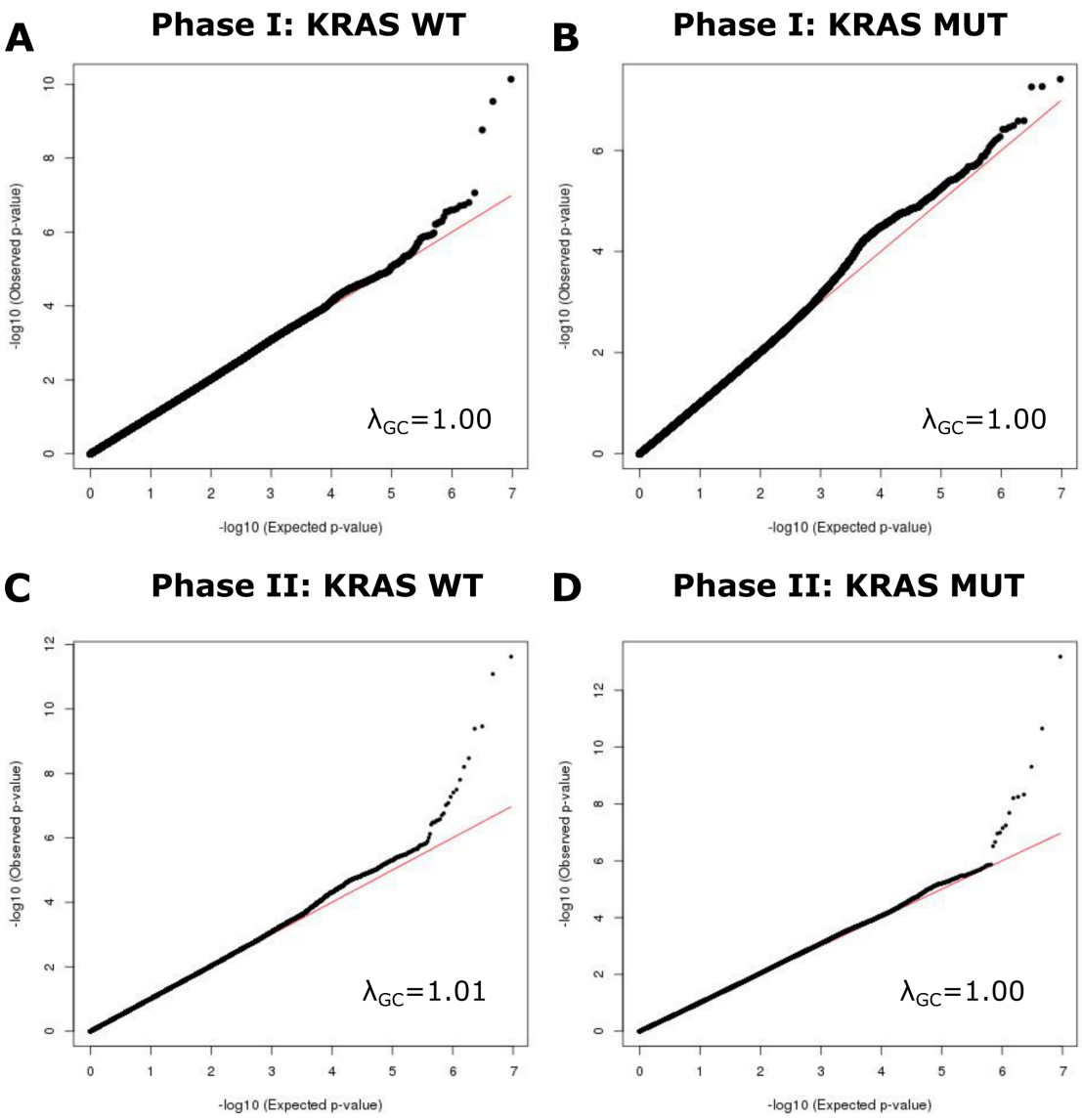

Figure 2.1. Quantile-quantile plots of stratified GWASs by KRAS mutational status.

(A, B) Phase I study for cases with KRAS WT (A) or MUT (B) tumours. (C, D) Phase II study for cases with KRAS WT (C) or MUT (D) tumours.

investigating the addition of the anti-angiogenic agent, bevacizumab, to capecitabine in the adjuvant treatment of stage II/III colorectal cancer [164]. The final intention-to-treat population comprised 1941 patients of whom 1715 (88%) had colonic tumours and 226 (12%) had rectal tumours. The SCOT trial was an international, randomised non-inferiority trial investigating 3 versus 6 months of adjuvant chemotherapy in high-risk stage II and stage III colorectal cancer [176]. The final intention-to-treat population consists of 6065 patients of whom 4987 (82%) had colonic tumours and 1101 (18%) had rectal tumours. Individuals from the QUASAR2 and SCOT tri-

als were identified for inclusion in this study based on the availability of genotyping, tumour tissue microarrays (TMA; for QUASAR2 only) and clinical outcome data. All tumours were either stage II or III and had undergone confirmed R0 resection. The primary and secondary research objectives were the association of variant genotype with time to CRC recurrence (TTR) (defined as the time from randomisation to CRC relapse, with censoring at last contact or death in case of no recurrence), and with overall survival (OS), respectively.

2.1.3.1 Genotyping and quality controls

DNA was extracted from EDTA-venous bloods using standard methods. Samples that failed DNA extraction were excluded. In the QUASAR2 cohort, 929 cases were successfully genotyped using Illumina Hap370, 610K and Omin2.5 arrays. Only SNPs typed in all three genotyping platforms were extracted and utilised. 3067 blood DNA samples from the SCOT cohort were genotyped using the Illumina Global Screening Array, and 2939 cases remained after quality evaluation.

Stringent quality control measures were applied to each sample cohort following standard protocols. Individuals with low SNP call rate ($< 95\%$) were removed. For apparent first-degree relative pairs (IBD > 0.185), I excluded the control from a case-control pair; otherwise, I excluded the individual with the lower call rate. Population stratification was examined using EIGENSTRAT [168], and individuals evaluated to be non-European ancestry (using the HapMap release 23a CEU, JPT/CHB and YRI populations as a reference) were removed from analysis. SNPs with a call rate $< 95\%$ were excluded as well as those with a MAF $< 1\%$ or displaying significant deviation from Hardy-Weinberg equilibrium ($P <$ 1e-05). SNPs in regions of long-range linkage disequilibrium, highly polymorphic regions, and sex chromosomes were also excluded.

2.1.3.2 Genotype Imputation

Prediction of the untyped SNPs was carried out using SHAPEIT v2.837 [170] and IMPUTE v2.3.2 [171], utilising a merged reference panel combining data from 1000 Genomes Project (phase 3, December 2013 release; [2]) and UK10K (April 2014 release; [172]). Post-imputation quality controls were performed using qctool v2

(`https://www.well.ox.ac.uk/~gav/qctool_v2/`). Rare variants (MAF $< 1\%$) and poorly imputed SNPs defined by an info score measure < 0.8 were excluded. For the imputed SNPs, genotype probabilities were converted to discrete genotypes using gtool v0.7.5 (`https://www.well.ox.ac.uk/~cfreeman/software/gwas/gtool.html`) with the default settings. A minimum probability threshold of 0.9 was required for specifying sample genotypes. Genetic linkage of SNPs was determined by calculating D' ad R^2 using 1000 Genomes Project data (phase 3, December 2013 release) via web server LDlink [173]. All genotype quality control steps were conducted in PLINK v1.9 [169].

I found little evidence of genomic inflation in any of the datasets (λ_{GC} range: [1.00, 1.02]; Fig 2.2), indicating that there were no hidden population structures in the datasets that could give rise to spurious associations.

2.1.3.3 HLA imputation, homozygosity and supertyping

The HLA regions of all cases in QUASAR2 and SCOT trials were typed using SNP arrays. To determine whether specific coding variants within HLA genes contributing to the patient survival outcomes, I imputed the alleles in the classical HLA genes (*-A, -B, -C, -DPA1, -DPB1, -DQA1, -DQB1* and *-DRB1*) and coding variants across the HLA regions using SNP2HLA v1.0.3 [177]. The pre-phasing and imputation were performed using Beagle v3.0.4 [178] based on a reference panel from the Type 1 Diabetes Genetics Consortium (T1DGC) consisting of genotype data from 5225 individuals of European descent with genotyping data of 8961 common SNPs and indel polymorphisms across the HLA region. Imputed allele dosage of four-digit typing of the HLA class I and II molecules were utilised for all downstream analyses.

Homozygosity in HLA-I genes of an individual was determined if both alleles for an HLA-I gene (*HLA-A, -B,* or *-C*) are of the same four-digit type. An individual can be homozygous or heterozygous in all three HLA-I genes, or any combination in between. Homozygosity in the five HLA-II genes (*HLA-DPA1, -DPB1, -DQA1, -DQB1* and *-DRB1*) were classified in a similar fashion.

Individual alleles in *HLA-A* and *HLA-B* genes were classified into 12 discrete supertypes on the basis of similar peptide-anchor-binding specificity [143, 179, 180].

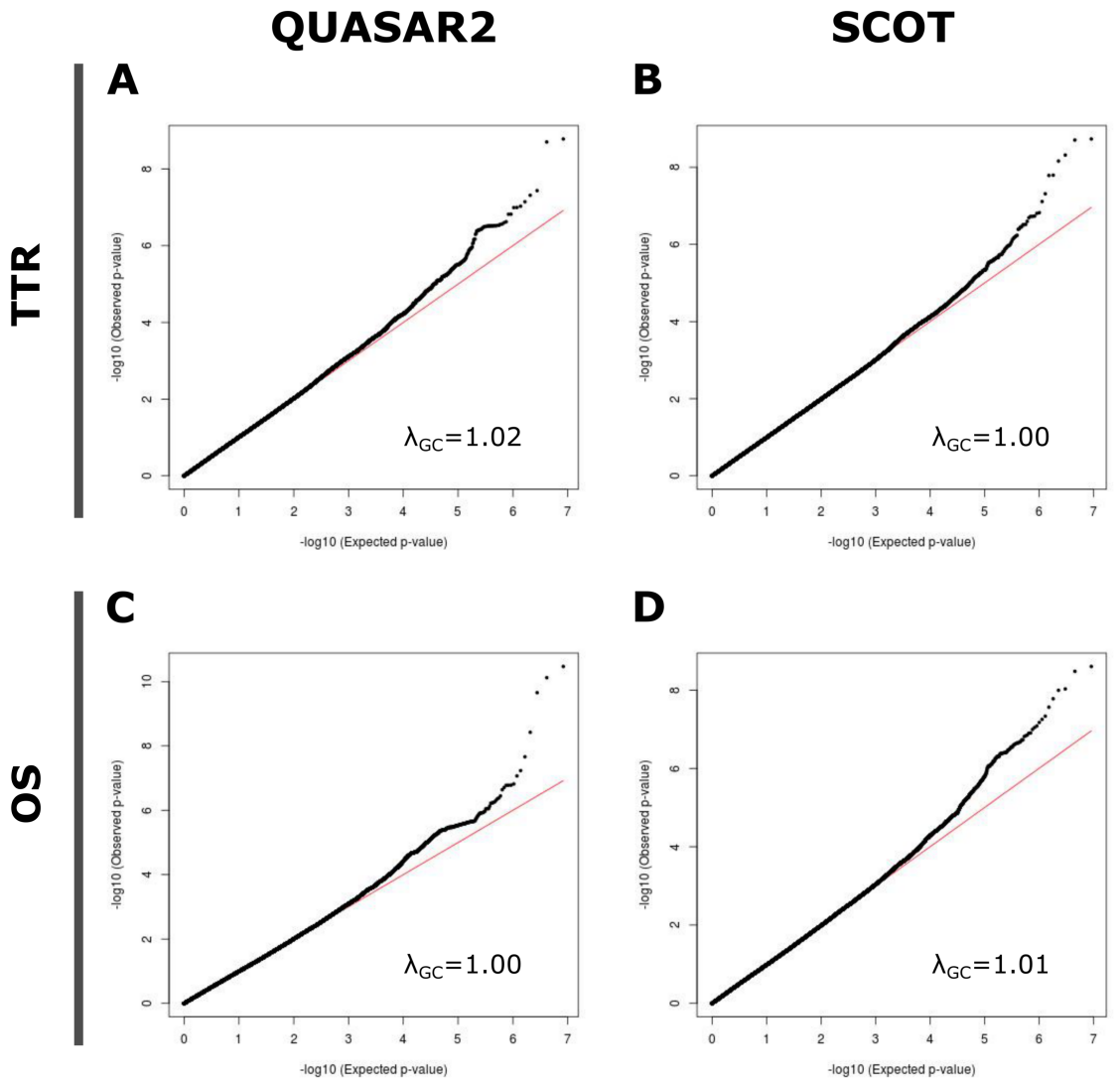

Figure 2.2. Quantile-quantile plots of genome-wide survival analyses. *(A, B) Time to CRC recurrence (TTR). (C, D) Overall survival (OS).*

Both HLA homozygosity and HLA-I supertype were tested against patient survival outcome in this study, assuming an additive genetic model.

2.2 Data Curation

2.2.1 Curation of *TP53* missense mutations in cancers

I curated pathogenic missense mutations in *TP53* by integrating experimental evidence from both literature and databases. Specifically, I combined *TP53* driver mutations identified in human tumours from two *in silico* studies to obtain a list of 323 *TP53* pathogenic mutations [67, 162]. To determine which of these 323 *TP53*

diver mutations are oncogenic (defined as those exerting dominant negative effect [DNE] or gain-of-function activities [GOF]), I relied on two sources of annotations: 188 missense mutations were curated to be oncogenic via promoter-functional assays in IARC *TP53* Database (release 18; [163]); 1101 missense mutations were ascertained by human cancer cell-based saturation mutagenesis screen [181] with the following filtering criteria:

- A549p53WTNutlin-3Z-score > 1
- A549p53NULLNutlin-3Z-score > 1
- A549p53NULLEtoposide_Z-score < -1

In total, 218 of the 323 *TP53* pathogenic mutations are oncogenic (Supplementary Data Table 2.1).

2.2.2 Known cancer risk variants

The GWAS catalogue was downloaded on 2018-02-28 (`https://www.ebi.ac.uk/gwas/`). I filtered the genome-wide significant lead SNPs ($P <$ 5e-08) in Europeans, and identified their proxy SNPs in linkage disequilibrium (LD) using the 1000 Genomes phase 3 data through the web server rAggr (`http://raggr.usc.edu`). In particular, I selected the proxies that met the following criteria:

- Population: EUR
- Minimal minor allele frequency (MAF): 0.01
- R^2 range: [0.8, 1]
- Maximal distance: 500KB
- Maximal Number of Mendel error: 1
- Hardy-Weinberg equilibrium (HWE) P-value: 1e-6
- Minimal genotype coverage: 95%

In total, I obtained a list of 283,240 variants. Among those, I identified 28,592 cancer GWAS SNPs, including 1,225 lead SNPs and 27,367 proxies, by mapping the GWAS SNPs to 106 unique cancer traits that are distributed into 27 distinct cancer types (Supplementary Data Table 2.2). All variants were mapped to the human genome Ensembl Release 91 (dbSNP build 150) to retrieve their hg38 genomic

coordinates using the *biomart* R package.

2.2.3 eQTLs in normal and cancerous tissues

Data for the eQTL analysis in normal human tissue are from: (1) GTEx v8 release [53], and (2) the Netherlands Study of Depression and Anxiety (NESDA) and the Netherlands Twin Register (NTR) that consisted of 4,896 blood samples with European ancestry [182]. Data for the eQTL analysis in human cancers were obtained from TCGA data portal. The gene expression profiles in TCGA primary tumours were retrieved from the Broad GDAC Firehose (`https://gdac.broadinstitute.org/`) through the fbget tool (v0.1.11, released on Oct 31, 2017). eQTL effects were determined using a linear model approach with mRNA expression level as the dependent variable and variant genotypes as the independent variable.

2.2.4 Variant annotations

Functional annotation of genetic variants was carried out using ANNOVAR (2019Oct24 release; [183]) and GWAVA [184]. This comprised of annotating the cytoband where the variants locate, nearby reference genes, variant frequency in diverse populations, and putative functional consequences. Candidate cancer risk loci were than checked against five types of functional data for biological interpretation (listed below). Chromatin and regulatory data were obtained from ENCODE database (phase 3 release; [185]).

- The presence of non-synonymous coding change
- The presence of an associated *cis*-eQTL
- The presence of a regulatory epigenetic state
- Evidence of TF binding
- The presence of a Hi-C contact linking to a gene promoter

2.3 Quantitative analyses

2.3.1 Analysis of oncogenic *TP53* missense mutations in breast and ovarian cancers

2,262 *TP53* mutations in 2,201 unique breast cancer samples (from 12 studies; exclude 737 duplicate mutations in samples sequenced by multiple studies) and 492 TP53 mutations in 471 unique ovarian cancer samples (from 3 studies; exclude 477 duplicate mutations in samples sequenced by multiple studies) were downloaded from cBioPortal (`http://www.cbioportal.org`) on 2018-09-14. All *TP53* missense mutations were extracted and matched with the curated lists of pathogenic and oncogenic *TP53* missense mutations as described in section 2.2.1. Then cancers with pathogenic missense mutations and oncogenic missense mutations were counted. Specifically, 1113 out of 2262 (49.2%) *TP53* mutations in breast cancer are pathogenic missense mutations, of which 1012 (90.9%) are oncogenic. Similarly, 260 out of 492 (52.8%) *TP53* mutations in ovarian cancer are pathogenic missense mutations, of which 228 (87.7%) are oncogenic.

2.3.2 Analysis for subtype heterogeneity SNPs in breast and ovarian cancers

Summary statistics of GWASs for breast cancer susceptibility were downloaded on 2018-03-12 (`http://bcac.ccge.medschl.cam.ac.uk/bcacdata/oncoarray/gwas-icogs-and-oncoarray-summary-results/`), which included summary statistics from case-control association analyses for ER-positive breast cancer cases (ER+BC) and ER-negative breast cancer cases (ER-BC) compared against disease-free controls. Summary statistics of GWASs for ovarian cancer susceptibility were downloaded on 2018-04-16 (`https://www.ebi.ac.uk/gwas/downloads/summary-statistics`), which included summary statistics for SNP association with low grade serous ovarian cancer (LGSOC), and with high grade serous ovarian cancer (HGSOC). Estimates of effect sizes [log(OR)s] for subtype-specific case-control studies and their corresponding standard errors were utilised for meta- and heterogeneity-analyses using METAL (2011-03-25 release; [186]), under an inverse-variance fixed-effect model.

Cochran's Q statistic was calculated to test for heterogeneity and the I^2 statistic to quantify the proportion of the total variation that was caused by heterogeneity.

2.3.3 Pathway enrichment analysis

The pathway gene sets of KEGG and Hallmark were downloaded from the Molecular Signatures Database (`http://software.broadinstitute.org/gsea/msigdb/index.jsp`). The known p53 direct target genes were obtained from [187]. *cis*-eQTL datasets were obtained from GTEX (`https://gtexportal.org/home/datasets`; V7 and q-value ≤ 0.05), NESDA/NTR (`https://eqtl.onderzoek.io/index.php?page=download`) and PancanQTL (`http://bioinfo.life.hust.edu.cn/PancanQTL/download`). The hypergeometric distribution enrichment analysis was performed as described in [188]. Significance was determined using the *PHYPER* function in R (v3.4.3; `http://www.r-project.org`) and multiple hypotheses testing by Benjamini-Hochberg correction.

2.3.4 TCGA pan-cancer analysis

KRAS mutational status in TCGA primary tumours was assigned based on hotspot mutations at codons 12, 13 and 61, as described in section 2.1.2.3. The CN profiles and mRNA levels of *TP53, KRAS, ATP5C1, GATA3, RRM2B* in TCGA primary tumours were downloaded from the Broad GDAC Firehose (`https://gdac.broadinstitute.org/`) using the fbget tool (v0.1.11 released Oct 31 2017). Since CN is strongly correlated with gene expression levels in general, I evaluated the relationship between variant genotypes and nearby gene expression among diploid tumours (i.e., GISTIC score = 0) exclusively. The association testing of allelic difference in tumours with distinct KRAS mutational status was performed using two-sided Fisher exact test and logistic regression assuming an additive genetic model. All statistical analyses were carried out in the R Statistical Computing environment (v3.4.3; `http://www.r-project.org`).

TCGA clinical data was downloaded from Pan-Cancer Clinical Data Resource (TCGA-CDR; [189]). Overall survival (OS) and progression-free interval (PFI), the two most accurate clinical outcomes using the current TCGA data, were merge with

the genotype data of primary tumours. Of the 7,021 TCGA patients of European ancestry, OS and PFI data was available for 6,979 and 6,977 patients, respectively. A multivariate Cox proportional hazards regression model was used to compare survival probabilities between groups differ in genotype, while accounting for confounding factors such as gender and age at diagnosis. Log-rank tests were performed to calculate the hazard ratio, the 95% confidence interval and P-values. All statistical tests were two-sided. TCGA clinical radiation data was retrieved using the *TCGAbiolinks* R package (version 2.16.1). The patients with "Radiographic Progressive Disease" were defined as radiation non-responders, and with "Complete Response" or "Partial Response" were defined as responders. All analyses were performed in the R Statistical Computing environment (v3.4.3; `http://www.r-project.org`) and implemented using the *survival* R package (`https://cran.r-project.org/package=survival`). The between-group difference in survival probability was visualised by plotting Kaplan-Meier curves.

2.3.5 GDSC drug sensitivity analysis

p53 somatic mutation, CN, mRNA expression data, and drug IC50 values for cancer cell lines were downloaded from Genomics of Drug Sensitivity in Cancer (GDSC; release-8.1). Specifically, a list of the mutated genes (file titled "mutations_20191101.csv"), the processed CN variation data (file titled "cnv_gistic_20191101.csv") and RNAseq gene expression data (file titled "rnaseq_read_count_20191101.csv") were downloaded from `https://cellmodelpassports.sanger.ac.uk/downloads`. The drug response data (folder titled "GDSC1_fitted_dose_response_15Oct19") was downloaded from `https://www.cancerrxgene.org/downloads/bulk_download`. Cell lines without mutations in *TP53* were assigned as p53 WT; Cell lines with *TP53* mutations ("cancer_driver" defined by GDSC) and copy-number alterations (GISTIC score < 0) were assigned as p53 MUT and CN loss; The cell lines were further grouped based on the gene transcript levels: low ($\leq$ 1st quartile), intermediate ($>$ 1st quartile and $<$ 3rd quartile), high ($\geq$ 3rd quartile). The effects of p53 mutational status or transcript levels on drug sensitivity were then determined using linear regression with log2 transformed IC50 values as dependent variable and CN status or transcript levels as independent variable.

2.3.6 KRAS-stratified association studies

Test of association between individual SNP genotypes and case-control status were undertaken using SNPTEST v2.5.2 [190] assuming an additive model. Stratified studies were carried out by including cases with either wild type or mutated KRAS in the association testing within respective study cohorts. The adequacy of the association studies was evaluated by quantile-quantile plots of test statistics in individual cohorts, and association signals were represented in Manhattan plots. Meta-analysis was performed using the fixed-effects inverse-variance method using METAL [186]. Cochran's Q-statistic testing for cohort heterogeneity and the I^2 statistic that quantifies the proportion of the total variation due to heterogeneity were calculated. Estimates of genomic inflation due to population stratification were calculated using the regression method implemented in *GenABEL* R package.

Statistical power of identifying genetic variants of various minor allele frequency associated with per allele CRC risk at varying significance levels was evaluated using Genetic Power Calculator with default settings (`http://zzz.bwh.harvard.edu/gpc/cc2.html`; [191]). Association signals of imputed markers in prioritised loci were visualised using the LocusZoom software [192]. For LocusZoom plot construction, I used the default combination rates, gene locations, the NHGRI catalogue data. LD values were calculated using the 1000 Genomes Project phase 3 data of the non-Finnish European population.

2.3.7 Survival analysis in the KRAS-stratified studies

A multivariate Cox proportional hazards regression model was used to compare survival probabilities between groups differ in genotype, while accounting for potential confounding factors. Proportionality of hazards was confirmed by Schoenfeld coefficients. For the multivariate analyses, adjustment was made for baseline demographic variables (age and gender), clinicopathological covariates of known prognostic value where available, and treatment type and schedule depending on the cohort. In the QUASAR2 trial, these comprised age, gender, tumour location (colon vs. rectum), primary tumour stage (pT4 vs. pT1-3), nodal stage (N0 vs. N1/2), treatment regimen (bevacizumab and capecitabine vs. capecitabine). In

the VICTOR trial, these included age, gender, tumour location (colon vs. rectum), disease stage (II vs. III), chemo-radiation history, treatment regimen (Vioxx vs. placebo). In both cohorts, covariates were prespecified and no selection procedure was performed. Log-rank tests were performed to calculate the hazard ratio, the 95% confidence interval and *P*-values. All statistical tests were two-sided. All analyses were performed in the R Statistical Computing environment (v3.4.3; `http://www.r-project.org`) and implemented using the *survival* R package (`https://cran.r-project.org/package=survival`). The between-group difference in survival probability was visualised by plotting Kaplan-Meier curves.

2.3.8 Survival analysis in the QUASAR2 and SCOT cohorts

The primary and secondary objectives (see also section 2.1.3) were evaluated by univariate and multivariate Cox proportional hazard regression models, assuming an additive or dominant genetic effect. The log-rank test was used to determine statistical significance of the survival distributions between patients with a specific genotype and those without. Proportionality of hazards was confirmed by Schoenfeld coefficients. For the multivariate analyses, adjustment was made for baseline demographic variables (age and gender), clinicopathological covariates of known prognostic value where available, and treatment type and schedule depending on the cohort. In the QUASAR2 study, these comprised age, sex (male vs. female), tumour location (rectum vs. colon), primary tumour stage (pT4 vs. pT1-3), nodal stage (N1/2 vs. N0), treatment regimen (bevacizumab and capecitabine vs. capecitabine). In the SCOT study, age, sex (male vs. female), tumour location (rectum vs. colon), primary tumour stage (pT1-2 vs. pT3 vs. pT4), nodal stage (N0 vs. N1 vs. N2), treatment regimen (FOLFOX vs. CAPOX), and treatment duration (24 vs. 12 weeks) were included. In both studies, covariates were prespecified and no selection procedure was performed. All statistical tests were two-sided.

All analyses were performed in the R Statistical Computing environment (v3.4.3; `http://www.r-project.org`) and implemented using the *survival* R package (`https://cran.r-project.org/package=survival`). The *gwasurvivr* R package [193] was used to scale up computation in the genome-wide survival analysis, where the backbone algorithms are consistent with those implemented in the *survival* R

package. The adequacy of the association studies was evaluated by quantile-quantile plots of test statistics in individual cohorts, and association signals were represented in Manhattan plots (generated using the *qqman* R package). Meta-analysis was performed using the fixed-effects inverse-variance method using METAL [186]. Cochran's Q-statistic testing for cohort heterogeneity and the I^2 statistic that quantifies the proportion of the total variation due to heterogeneity were calculated. Genomic inflation were assessed using the regression method implemented in the *GenABEL* R package.

2.3.9 Tumour immunohistochemistry and iQTL mapping

QUASAR2 tissue microarrays (TMAs) were constructed using punches taken from the centre of the tumour in FFPE tumour blocks following identification by the study pathology. Immunohistochemistry (IHC) for CD8 were performed as described in [194]. Marker positive cells were quantified by computerised analyses using ImmunoPath 1.3.9.0 and expressed as the proportion of CD8+ cells in the total number of cell nuclei across all TMA cores for each sample. CD8+ cell density was log2 transformed prior to inclusion in the immune quantitative trait loci (iQTL) analysis. Tests of association between individual SNP genotypes and quantitative trait were evaluated by Cochran-Armitage test under an additive genetic model in SNPTEST v2.5.2 [190]. P-values were generated for each variant testing against the alternative hypothesis that the slope of a linear regression model between genotype and CD8+ cell density in TMAs deviates from 0.

2.3.10 Two-stage Mendelian randomisation

Candidate variants for the genetic instrument was selected from the iQTL analysis using the QUASAR2 data. Amongst the 80 iQTLs associated with tumour-infiltrating CD8+ T cell density at suggestive significance ($P <$ 1e-05), 22 lead SNPs were selected after LD pruning ($R^2 < 0.2$; Non-Finnish European populations) to construct the instrumental variable. First, I extracted the summary statistics of these 22 variants, including beta estimates for alleles associated with increasing tumour-infiltrating CD8+ T cell density (z_k for SNP k), standard error and P-value. Next, I constructed a weighted genomic risk score (GRS) for each individual

in the QUASAR2 cohort using the extracted information. Implemented in PRSice v2.3.3 [195], these variants were weighted by their effect size of association with tumour-infiltrating CD8+ T cell density per allele.

The first-stage regression analysis evaluated the IV-exposure relationship, i.e., testing the association between the GRS and log2-transformed tumour-infiltrating CD8+ T cell density in a univariate linear regression model. Denote the exposure variable x_i for individual i, the form of relationship is summarised below:

$$x_i = \alpha_0 + \alpha_1 GRS + \epsilon_{X_i},$$

where $GRS = \sum_k \alpha_k z_{ik}$ and ϵ_{X_i} is the individual-specific error term.

F-statistic of the first-stage regression analysis was evaluated to assess the strength of the genetic instrument in the linear model. An F-statistic < 10 indicates a weak instrument [196]. Potential association between our instrumental GRS of tumour-infiltrating CD8+ T cell density and common confounders including age, gender, tumour location, pT and N stage, treatment was also examined to ensure the second MR assumption. I also searched the NHGRI-EBI GWAS Catalogue for any association between the 22 candidate iQTLs and potential confounding prognostic variants.

The second-stage regression analysis evaluated the association between GRS and patient survival outcomes by a multivariate Cox regression model, adjusting for known prognostic factors. I measured the causal effect using a likelihood-based method equivalent to the coefficient ratio method widely used in MR studies [196]. That is, the fitted values $\widehat{x_i} = \widehat{\alpha_0} + \widehat{\alpha_1} GRS$ were passed into Cox regression to obtain an estimate of the effect of the exposure on the patient survival outcome (hazard exponent y_i for individual i):

$$y_i = \beta_0 + \beta_1 \widehat{x_i} + \epsilon_{Y_i}.$$

Here, β_1 is the causal parameter of interest, ϵ_{X_i} and ϵ_{Y_i} are independent error terms.

2.3.11 Mutual information

The theoretic concept of mutual information (MI) provides a general framework to evaluate dependencies between variables. In the context of causal inference, it has been suggested as a general measure of the presence of nonlinear relationship as opposed to commonly used linear methods. MI is defined in terms of discrete variables, its application to continuous data thus requires binning procedures.

In this study, I calculated MI between tumour-infiltrating CD8+ T cell density (log2 transformed) and patient survival outcome data using the *infotheo* R package (`https://cran.r-project.org/package=infotheo`), by specifying equal-frequency discretisation method and empirical probability distribution for entropy computation. To investigate the extend of numerical uncertainty in MI estimate, I tested a range of bins up to the recommended upper bound $N^{0.6}$ [197], where N is the number of samples. I found that MI estimate increases as the number of bins increases (Fig 2.3). Therefore, I fixed the number of bins to be 53 (the upper bound) for all downstream MI calculation. To account for effect of right censoring in patient survival outcome data, I simulated various scenarios on mutual information estimate by assuming a time delay τ in real observation amongst censored cases, where $\tau \sim U(0, D)$ and $D \in \{100, 500, 1000\}$ (in days). Each scenario was simulated for 10,000 iterations. The empirical distribution of mutual information estimate was constructed via bootstrapping, where the tumour-infiltrating CD8+ T cell density data were sampled with replacement for 50,000 iterations. The empirical p-value was calculated as the probability of obtaining an estimate larger than the observed mutual information estimate. Maximal information coefficients were computed using *minerva* R package (`https://CRAN.R-project.org/package=minerva`) using default settings.

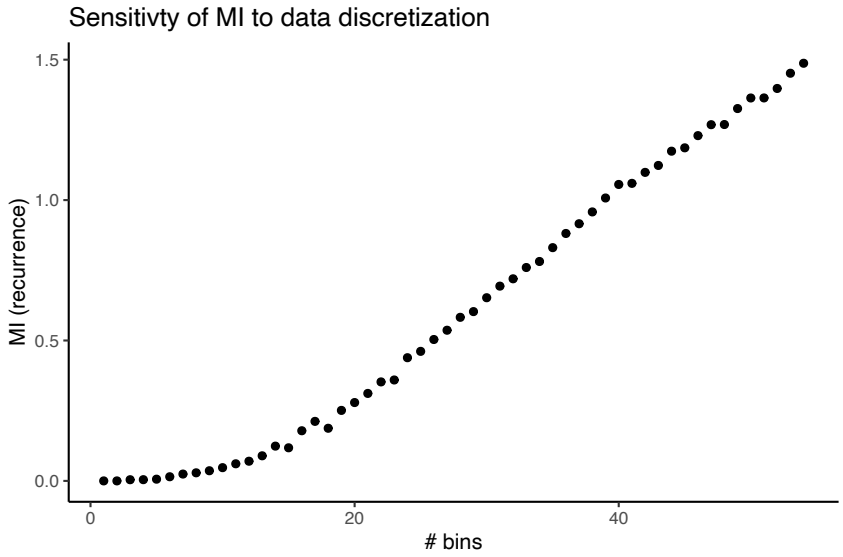

Figure 2.3. Sensitivity of MI estimate to the granularity of discretisation.

Chapter 3

Germline by somatic interaction in the p53 pathway

3.1 Introduction

3.1.1 The p53 tumour suppressor pathway

Research over the last forty years has established the notion that the transcription factor p53 safeguards the integrity of the genome, thus preventing the cells from neoplastic transformation and suppressing tumour development [198–200]. In response to cellular stresses such as DNA damage, oncogenic activation, oxidative and proteotoxic stress, p53 is activated and induces target gene expression to coordinate cellular outcomes, notably cell cycle arrest and apoptosis (Fig 3.1). In particular, upon moderate DNA damage, activated ATM or ATR trigger transient accumulation and oscillations of p53 in the nucleus, inducing cell-cycle arrest and facilitating DNA damage repair via transactivating genes involved in the process of cell cycle control (e.g., *CDKN1A*) and those in the DNA repair machinery (e.g., *XPC*), respectively [201]. In other cases, prolonged or extensive DNA damage sustains p53 activation in the nucleus, resulting in apoptosis (via selectively inducing the expression of genes such as *BAX* and *PUMA*) or cellular senescence [201]. p53 target genes also include those involved in metabolic remodelling, angiogenesis and cellular motility (Supplementary Data Table 3.1; [187, 202]). Together with p53's upstream regulators and

its diverse downstream effectors in the pathway, the tumour suppressor p53 plays an essential role in normal physiology and acts against uncontrolled cell proliferation [199, 203, 204].

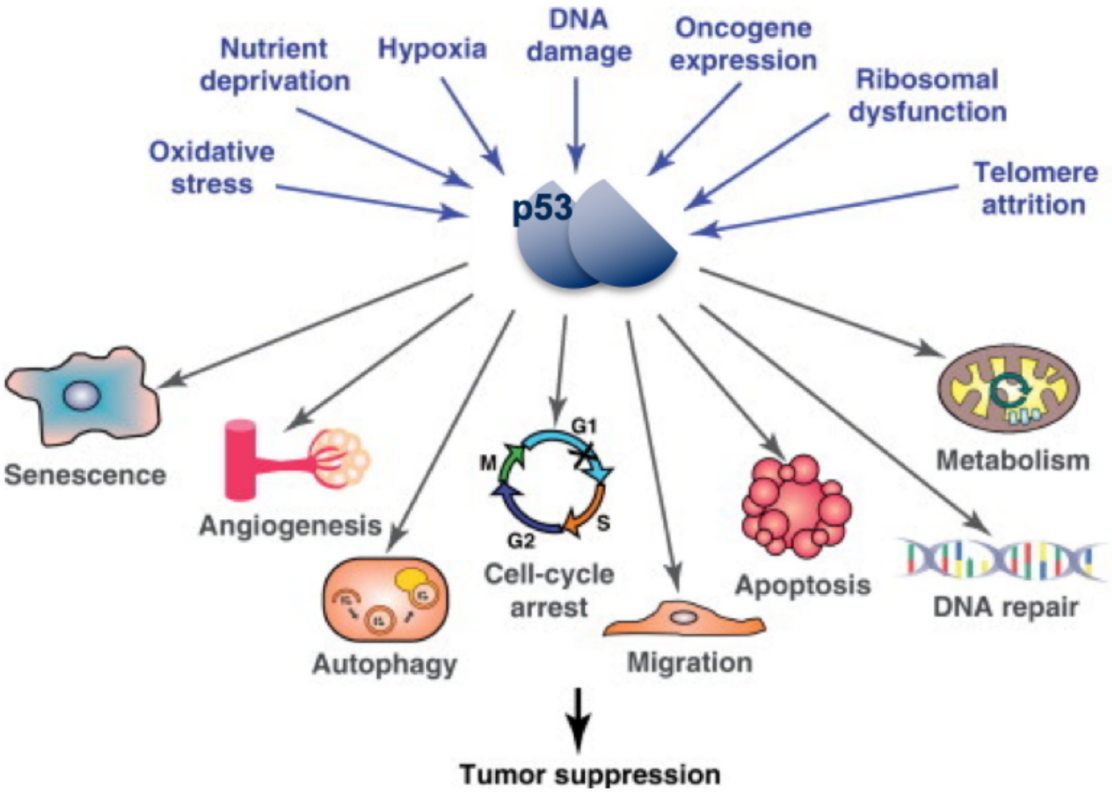

Figure 3.1. An overview of the p53 tumour suppressor pathway.

3.1.2 Somatic mutations in the p53 pathway

The p53 protein is encoded by the *TP53* gene, which is mutated in 36% of all human cancers [205, 206]. Somatic mutations detected in tumours affect all parts of the gene body, with four residues in the DNA-binding domain the most frequently perturbed: R273, R248, R175, and R282 (i.e., 'mutation hotspot'; Fig 3.2A). Unlike other tumour suppressor genes that are mostly mutated by truncating mutations such as nonsense and frame-shift mutations, 64% of somatic mutations in *TP53* are missense mutations (Fig 3.2B). In general, p53 loses its tumour suppressive function via three types of driver mutations [181, 207, 208]: (1) typical loss-of-function (LOF) mutations, including all truncating mutations; (2) missense mutations resulting in mutated p53 that is structurally intact but exerts dominant-negative (DNE) effect by binding to wild type p53 and inhibiting its transactivation function; or (3) missense

mutations resulting in mutated p53 that exhibit neomorphic gain-of-function (GOF) activities that promote cancer instead of suppressing it (Fig 3.2C). Moreover, 91% of tumours with *TP53* mutations lose both wild type copies of p53, due to loss of heterozygosity (LOH) or mutations affecting both alleles [206].

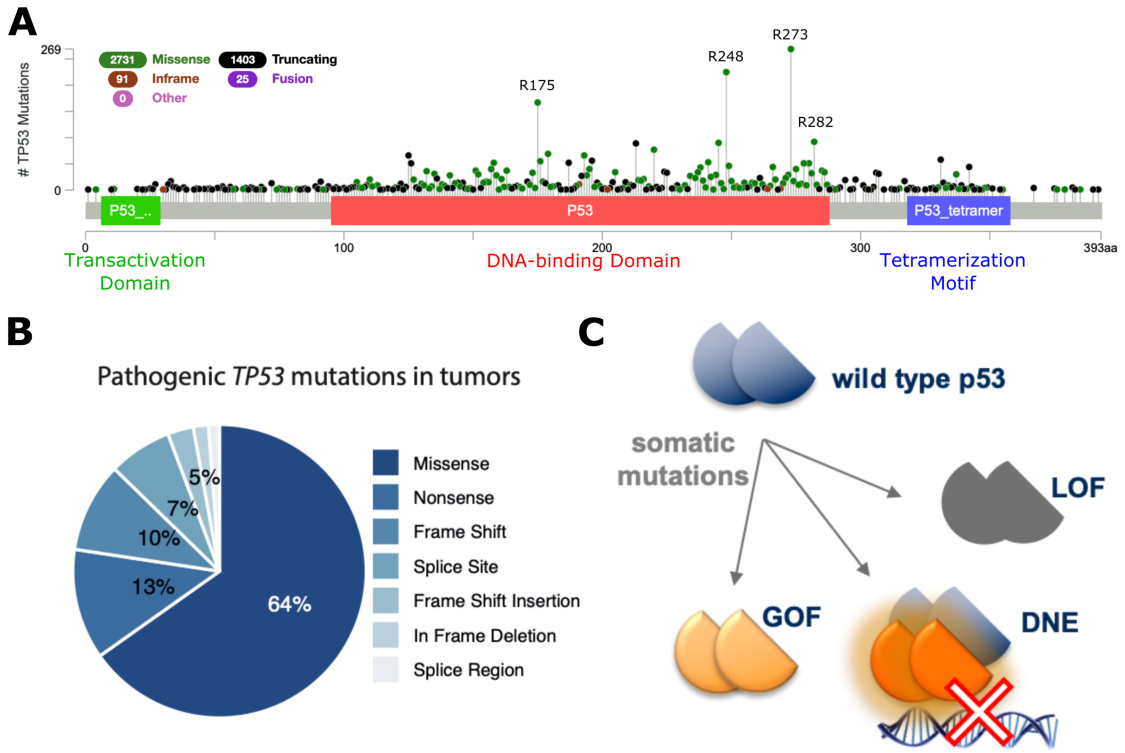

Figure 3.2. Somatic mutations in TP53 and their functional consequences.
(A) A survey of the p53 mutational spectrum in human cancers. 3,943 of the 10,967 tumour samples (36%) in the Cancer Genome Atlas (TCGA) pan-cancer cohort harbour somatic mutations in TP53. The lollipop plot shows the cumulative incidence of mutations observed in the samples. Hotspot residues are annotated. (B) The distribution of different mutation types amongst all pathogenic mutations that have deleterious effect on p53 function (see Methods, section 2.2.1). (C) Consequences of somatic driver mutations that affect p53 functions: (i) loss of wild-type function (LOF); (ii) dominant-negative effect (DNE), where mutated p53 binds to wild-type p53 and inhibits its DNA-binding activity; and (iii) gain-of-function activities (GOF), such as transactivating novel target genes.

Interestingly, I found that 67% of all pathogenic mutations affecting p53 in tumours are functionally DNE or GOF mutations, and that over 85% of tumours with mutated p53 harbour DNE or GOF mutations (Supplementary Data Table 2.1; see Methods, sections 2.2.1 and 2.3.1). This suggests that mutated p53 in tumours behaves very differently from its wild type counterpart. In other words, wild type p53 is a tumour suppressor, whereas mutated p53 tends to act as an oncoprotein in tumours. Considering the pivotal role of p53 in the cellular signalling network,

such functional disparity between the wild type and mutated p53 could plausibly contribute to divergent disease phenotypes seen in cancer subtypes that differ in p53 mutational status (see also section 1.2).

Apart from somatic mutations affecting p53, other members of the p53 pathway are also altered somatically in tumours. For instance, inactivation mutations or copy number (CN) loss in ATM were observed in 10% of all human cancers [209], and CN amplification in MDM2 (3%) and MDM4 (1%) were also detected in tumours [67, 210]. Overall, impaired p53 pathway function is a common feature in human cancers [206, 209].

3.1.3 Inherited cancer risk variants in the p53 pathway

As discussed in section 1.1.2, pathogenic germline mutations in *TP53* predispose LFS patients to a wide spectrum of early onset cancers. Mostly missense mutations affecting key amino acid residues in *TP53* (R158H, G245S, R248W, R248Q, E282W), these germline mutations occur in 70% of LFS, in 20-40% of Li-Fraumeni like families [24]. In addition, up to 20% of pathogenic germline *TP53* mutation carriers were found to have acquired the mutations *de novo*. For individuals with these pathogenic germline mutations in *TP53*, the overall cancer risk is 70% in men and almost 100% in women [24].

In addition to somatic and rare germline mutations, common inherited genetic variants were also found to modify the activity of the p53 pathway and associated with cancer [211]: P47R (rs1800371), P72R (rs1042522), and PAS (rs78378222) SNPs in *TP53*, SNP309G (rs2279744) and SNP285C (rs117039649) SNPs in *MDM2*, and SNP34091 (rs4245739) in *MDM4*. Interestingly, the p53 pathway is enriched for cancer risk variants compared to other cellular signalling pathways, and cancer risk variants in the p53 pathway share genetic characteristics with well-characterised somatic driver mutations in the pathway [188]. This suggests that inherited genetic variants in the p53 pathway could potentially impair pathway activity to confer cancer risk, just like the somatic driver mutations in the pathway to promote cancer, although with a modest effect size (see also section 1.1.3).

As discussed in section 1.3.3, a number of inherited genetic variants were as-

sociated with specific somatic mutations in tumours (Table 1.4). This concurrence of inherited genetic variants and somatic driver mutations in the same cancer hallmark pathway poses the question whether they interact to influence cancer risk and prognosis. If true, disease prediction leveraging inherited cancer risk variants (see also section 1.1.5) might also need to accommodate for molecular information about tumours to achieve an accurate personalised profile for cancer susceptibility and prognosis. However, the interaction between inherited genetic variants and somatic driver mutations on cancer risk and prognosis is yet to be defined. In this chapter, I aim to explore germline by somatic interaction in the p53 tumour suppressor pathway in common cancers.

3.1.4 A hypothesis of TP53-mSNPs

While 25 cancer risk loci have been associated with multiple cancers (e.g., 8q24 and 5p15; [212–216]), most cancer risk loci are specific to individual cancer type. As shown in section 1.3.1, inherited genetic variants can be associated with specific cancer subtypes, and the associations vary drastically across cancer subtypes (Table 1.3). Moreover, as mentioned above, inherited genetic variants were also shown to associate with specific somatic driver mutations in tumours (see also section 1.3.3; Table 1.4). In light of these associations, if certain inherited genetic variants track with p53 somatic driver mutations in tumours, their association with cancer risk would differ between cancer subtypes with contrasting p53 mutational status (TP53-related subtype heterogeneity SNPs, or TP53-shSNPs).

Considering that the wild type p53 (p53 WT) is a tumour suppressor but the mutated p53 (p53 MUT) tends to act as a oncoprotein, I reason that inherited genetic variants that track with p53 mutational status in tumours could associate with inverted risk between the p53 WT and p53 MUT tumour strata. For example, an allele could be depleted in p53 WT cases when compared to healthy controls, but the same allele could be enriched in p53 MUT cases (Fig 3.3A). This pattern of association would result in inverted risk profile between p53 WT and p53 MUT strata (Fig 3.3B). In this scenario, standard association studies would overlook these candidate loci, due to the off-setting effect of mixing cases (dashed lines in Fig 3.3A). In this chapter, my objective is to test this hypothesis by searching for cancer risk

variants that track with p53 mutational status in tumours (TP53 mutation-specific SNPs, or TP53-mSNPs).

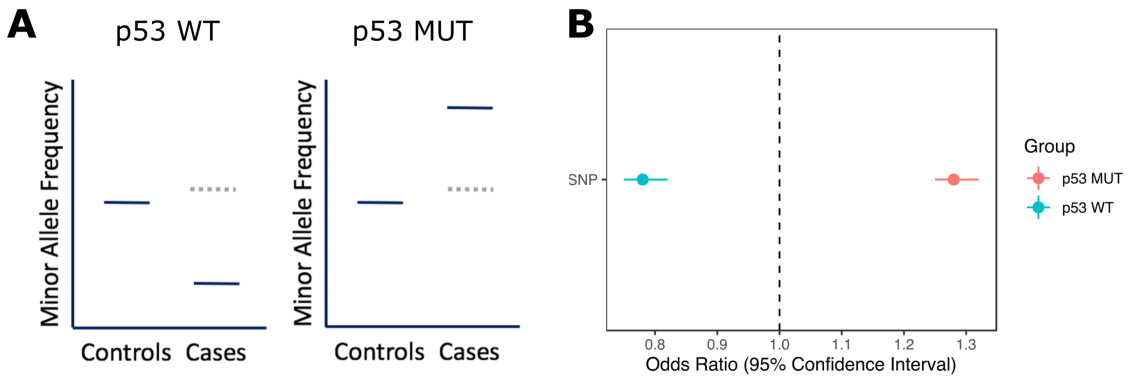

Figure 3.3. A hypothesis of TP53-mSNPs.

(A) A schematic demonstrating inverted association between inherited genetic variants and cancer risk by comparing minor allele frequency (MAF) between cancer cases and healthy controls, where cases are stratified by p53 mutational status. (B) A forest plot demonstrating the inverted risk profile of a TP53-mSNP as shown in (A).

3.2 Results

3.2.1 Cancer subtype-stratified association studies uncover TP53-shSNPs

3.2.1.1 p53 mutational pattern in breast and ovarian cancer subtypes

As discussed in 1.1.3, large population cohorts with both germline and somatic data are the ideal for identifying potential TP53-mSNPs. However, these cohorts are rare. To circumvent this issue, I decided to leverage large cohorts with germline and molecular subtype information to first identify TP53-shSNPs, which could potentially be candidates for TP53-mSNPs.

I start by selecting cancer cohorts with subtypes that differ greatly in p53 mutation frequencies, for which genotype data are available. I found two potentially informative cancer types (see also sections 1.2.2 and 1.2.3). The first was breast cancer: 18% of ER positive breast cancer cases (ER+BC) have mutated p53, in contrast to 76% of ER negative breast cancer (ER-BC) cases [217]. The second was ovarian cancer: less than 10% of low-grade serous ovarian cancer (LGSOC) cases have mutated p53 [218], in contrast to 96% of high grade serous ovarian cancer

(HGSOC) cases [94]. As in section 3.1.2, I observed that over 85% of all pathogenic missense mutations in *TP53* render the mutated p53 oncogenic (DNE or GOF) in breast and ovarian cancers (Fig 3.4B).

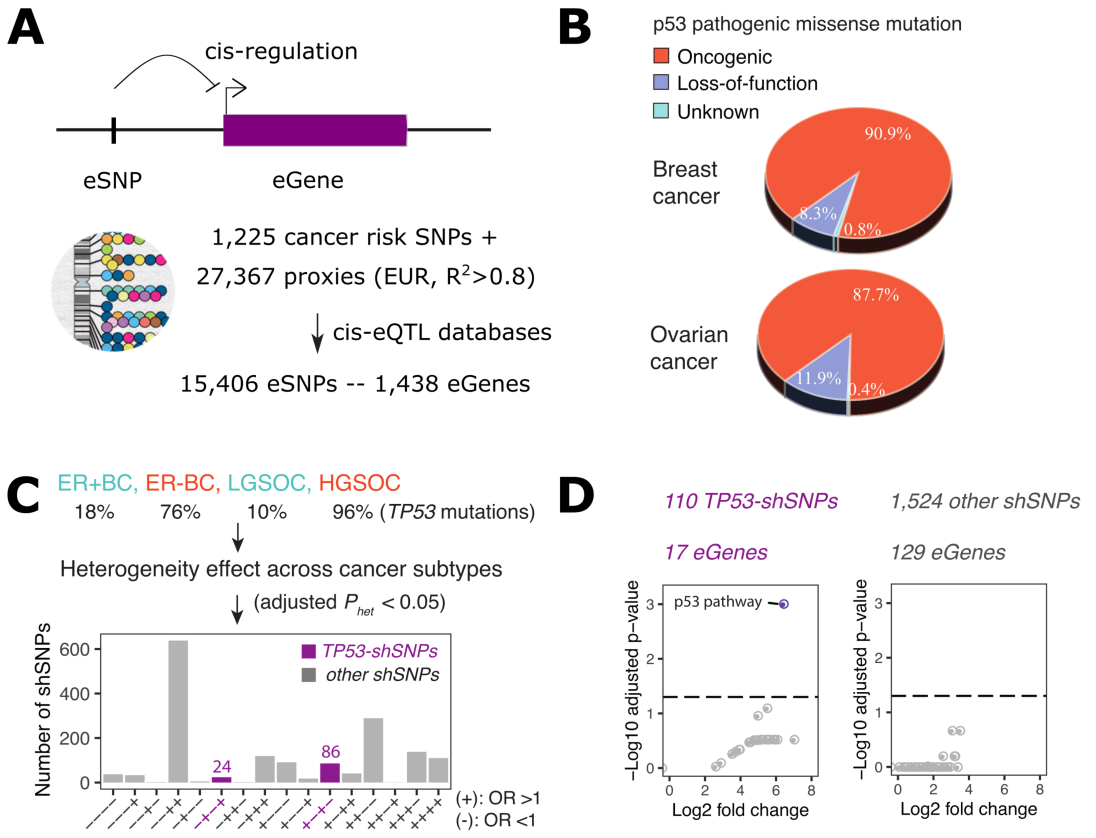

Figure 3.4. Cancer risk variants in the p53 pathway show subtype heterogeneous associations.

(A) A schematic demonstrating candidate loci selection. (B) Pie charts of the percentages of tumours with oncogenic, loss-of-function p53 mutations amongst all tumours bearing p53 pathogenic mutation in breast and ovarian cancers. (C) A bar plot demonstrates number of SNPs with significant heterogeneity P-values (P_{het} < 5e-08) across both breast and ovarian cancer subtypes (shSNPs), including the subtype heterogeneity SNPs (i.e., TP53-shSNPs with -+-+ and +-+- directions in estimated effect sizes; purple bars) that have consistent effects in subtypes with different p53 mutation frequencies. (D) A scatter plot of the fold enrichment of subtype heterogeneity eGenes on the x-axis (log2 scale), and the adjusted P-value on the y-axis (-log10 scale), amongst each pathway relative to all eGenes of the genome. The enrichment of TP53-shSNPs in p53 pathway genes is marked in purple while the other 185 annotated KEGG pathways are marked in grey. The horizontal dashed lines mark the false discovery rate of 0.05.

My next objective was to identify cancer risk variants that show heterogeneous association between these subtypes (i.e., TP53-shSNPs) with sufficient statistical power. Here, I utilised data generated in an analysis of 90,969 breast cancer cases (69,501 ER+BC, 21,468 ER-BC) and 105,974 controls [219] and an analysis of 14,049

ovarian cancer cases (1,012 LGSOC, 13,037 HGSOC) and 40,941 controls [220] (see Methods, section 2.3.2).

3.2.1.2 TP53-shSNPs in breast and ovarian cancers

To identify TP53-shSNPs, I first curated known cancer risk loci from previous cancer GWASs (see Methods, section 2.2.2). By Feb 2018, there were 1,225 cancer GWAS lead SNPs ($P < 5e\text{-}08$) in the GWAS catalogue, which are in linkage disequilibrium (LD) with 27,367 variants (i.e., proxies; $r^2 > 0.8$ in European populations). Among these cancer risk variants (both lead SNPs and proxies), 15,406 were associated with differential expression levels of 1,438 genes in at least one tissue or cell type (Fig 3.4A; see Methods, section 2.2.3; [53, 182, 221]. In this section, I aimed to test whether these expression cancer risk variants (eSNPs) show heterogeneous association between the breast and ovarian cancer subtypes. Of note, the 1,438 eSNP-mapped genes (eGenes) were enriched for p53 pathway genes compared to genes in other annotated KEGG pathways (adjusted $P = 7.8e\text{-}05$; see Methods, section 2.3.3), consistent with a previous report [188]. The p53 pathway eGenes include *TP53*, its key regulators (*MDM4, ATM, CHEK2, CDKN2A*) as well as its key effectors (*CASP8, CDKN1A, FAS, PIDD, CCNE1, CCND1, SESN1, PMAIP1*).

Next, I examined the large-cohort datasets for TP53-shSNPs. Of the 15,406 eSNPs, 1,634 showed significant subtype heterogeneity after correction for multiple hypothesis testing (Bonferonni adjusted $P_{het} < 0.05$; Supplementary Data Table 3.2) across the four subtypes (ER+BC, ER-BC, LGSOC, HGSOC; Fig. 3.4C). Intriguingly, of these 1,634 shSNPs, I identified 110 TP53-shSNPs, whose pattern of risk association is inversely correlated with p53 mutation frequencies in the breast and ovarian cancer subtypes (purple bars in Fig. 3.4C). That is, the alleles of these SNPs that are associated with increased cancer risk (OR > 1) in the subtypes with low p53 mutation frequencies (ER+BC and LGSOC), are associated with decreased cancer risk (OR < 1) in the subtypes with high p53 mutation frequencies (ER-BC and HGSOC), and vice versa.

The 110 TP53-shSNPs were mapped to 17 eGenes, and the remaining 1,524 shSNPs were mapped to 129 eGenes. The 17 eGenes mapped by the 110 TP53-

shSNPs were significantly enriched in the p53 pathway but not in any other annotated KEGG pathways (87.0-fold, adjusted P = 9.9e-04; see Methods, section 2.3.3; left panel in Fig. 3.4D). In contrast, I didn't observe such enrichment in any annotated pathways for the 129 eGenes mapped by the other shSNPs (right panel in Fig. 3.4D). The 17 eGenes mapped by the 110 TP53-shSNPs include *TP53* (3 eSNPs) and two of its key regulators (*ATM*, 44 eSNPs; *MDM4*, 33 eSNPs). All variants in each gene are in LD ($r^2 > 0.9$ in European populations; Supplementary Data Table 3.3). These results illustrates that key p53 pathway genes harbour TP53-shSNPs, which show heterogeneous association that is inversely correlated with p53 mutation frequencies in breast and ovarian cancer subtypes.

3.2.2 Pan-cancer analysis reveals rs78378222 is a TP53-mSNP

The above cancer subtype-stratified association analysis shortlisted 110 cancer risk variants as potential candidates for TP53-mSNPs. For example, the minor allele C of the lead SNP in *TP53*, rs78378222, was associated with increased risk for ER+BC and LGSOC (OR = 1.12, P = 9.98e-04 and OR = 1.59, P = 0.016, respectively), but with decreased risk for ER-BC and HGSOC (OR = 0.80, P = 2.30e-04 and OR = 0.75, P = 3.68e-04, respectively) (Fig 3.5A). Together, the minor allele C show heterogeneous association with risk between the breast and ovarian cancer subtypes that is inversely correlated with p53 mutation frequency (adjusted P_{het} = 3.8e-05). On the other hand, rs11374964 in *ATM* and rs4245739 in *MDM4* showed association specific to ER-BC and HGSOC (Fig 3.5A).

As mentioned above, the three leading SNPs in *TP53*, *ATM*, and *MDM4* were previously associated with differential risk for cancer. In particular, rs78378222 in *TP53* was associated with differential risk for brain malignancies, skin cancer, breast cancer, leukaemia, lymphoma, leiomyoma, and soft-tissue sarcoma [51, 216, 222–230]. Meanwhile, rs11374964 in *ATM* was exclusively reported to associate with risk for breast cancer [231]; rs4245739 in *MDM4* was associated with risk for breast and prostate cancer [231–233]. Despite that subtype-specific risk associations of these three loci have been reported in breast cancer [51, 231], no study has examined the interaction between these variants and p53 mutational status in tumours.

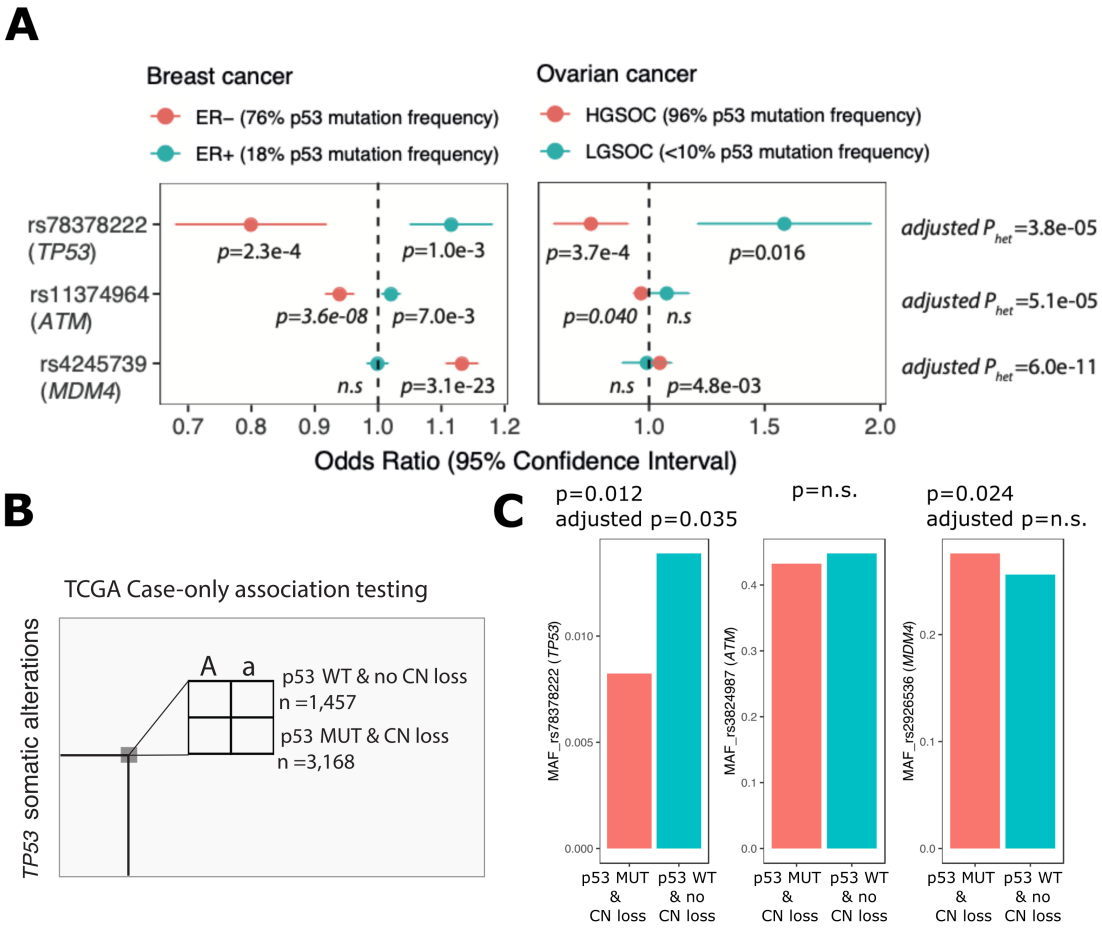

Figure 3.5. TP53-shSNPs track with p53 mutational status.

(A) Forest plot illustrating the patterns of cancer risk associations of the top candidate variants in p53 pathway amongst breast and ovarian cancer subtypes. (B) A schematic overview of the association testing between shortlisted candidate variants and p53 mutational status in primary TCGA tumours. (C) Bar graphs of the allelic difference of the candidate variants in (A) between tumours with wild type and mutant p53 in TCGA patient cohort.

To test whether these three loci are TP53-mSNPs, I performed similar analyses on these three loci in the Cancer Genome Atlas (TCGA) pan-cancer cohort, for which both germline and somatic information was available. To minimise population stratification, I selected 7,021 patients of European ancestry (see Methods, section 2.1.1.1), who have been diagnosed with 31 different cancers. Next, I determined the p53 mutational status of their tumours (see Methods, section 2.1.1.2).

In the TCGA cohort, 35.8% of patients have at least one pathogenic mutation in *TP53* (defined as LOF, DNE or GOF mutations) in their tumours, 37.8% have CN loss in *TP53*, and 20.8% have both pathogenic mutation and CN loss (see Methods, section 2.1.1.2; Supplementary Data Table 3.4). Given that p53 mutations are

significantly correlated with CN loss in *TP53* ([206]; Fig 7.1A), I stratified TCGA patients into two groups based on the presence of p53 somatic alterations (MUT & CN loss vs. WT & no CN loss) and performed association testing on the three loci against p53 somatic alterations using a one-sided Fisher's exact test (Fig 3.5B). I found that one of the three loci tracked with p53 mutational status in tumours (rs7837822 in *TP53*; adjusted $P = 0.035$; Fig 3.5C). As shown in Fig 3.5A, the minor allele C of rs78378222 was associated with increased frequency in ER+BC and LGSOC (low p53 mutation frequency) cases compared to healthy controls, but decreased frequency in ER-BC and HGSOC (high p53 mutation frequency) cases. The current finding indicates that the minor allele C is significantly depleted in p53 MUT & CN loss tumours than in p53 WT & no CN loss tumours, consistent with the results in cancer subtype-stratified association analysis in section 3.2.1.2 (Fig 3.5A). This finding indicates that rs78378222 in *TP53* is a TP53-mSNP as hypothesised.

3.2.3 Further analyses of candidate variant rs78378222

3.2.3.1 rs78378222 is associated with p53 expression levels

The variant, rs78378222, resides in the 3'-UTR of *TP53*. Specifically, the minor allele C disrupts the canonical *TP53* poly-adenylation signal sequence (PAS, p53 poly(A) SNP; Fig 3.6A). The minor allele C was previously associated with lower *TP53* mRNA levels in blood samples [234].

When I examined transcriptome-wide eQTL data from 4,896 peripheral blood samples (see Methods, section 2.2.3), I found that, indeed, the minor allele C was associated with 1.54-fold per-allele decrease in p53 mRNA levels (beta = -0.62, $P =$ 2.0e-25; Fig. 3.6B). To investigate the association of this variant with p53 expression levels in tumours, I analysed expression data from 3,248 tumours in the TCGA cohort. Since CN loss in *TP53* is associated with lower p53 expression levels in TCGA tumours (Fig 7.1B), I only considered diploid samples in this analysis (i.e., GISTIC score = 0). Like in normal tissues, the minor allele C was associated 1.3-fold per-allele decrease in *TP53* mRNA levels in tumours (beta = -0.37, $P =$ 1.7e-04; Fig. 3.6C). To test whether this association depends on p53 mutational status in tumours, I separated the tumours into three groups based on their p53 mutational

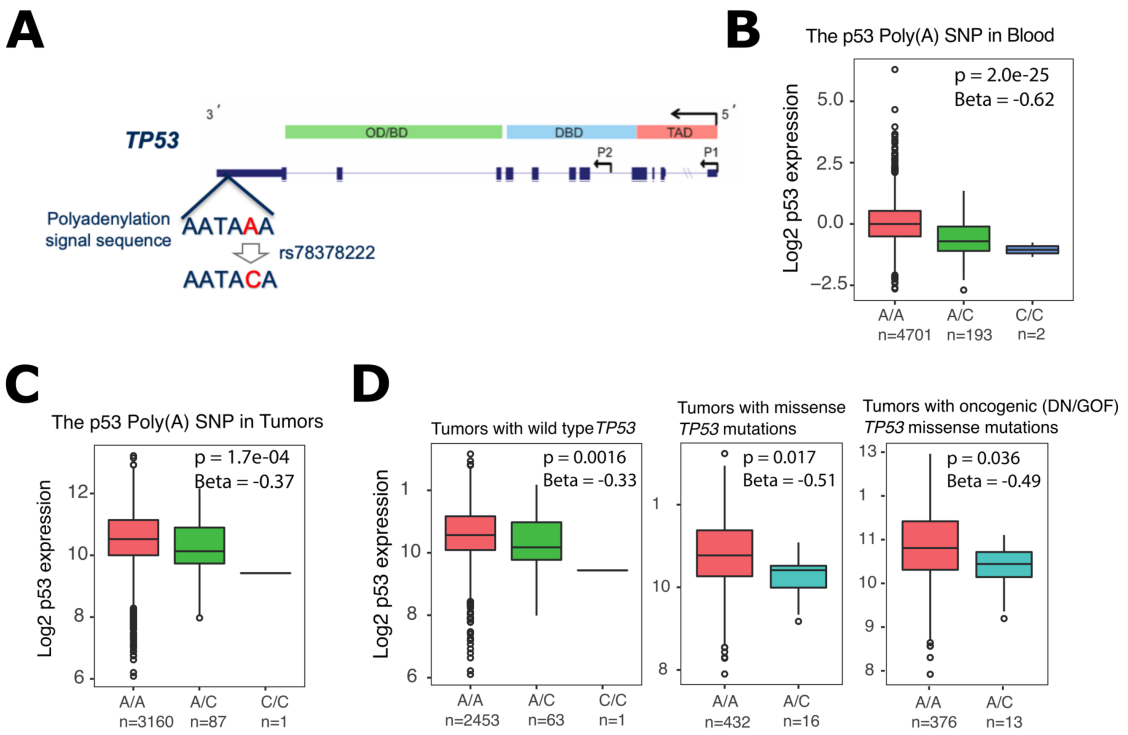

Figure 3.6. rs78378222 is associated with p53 expression levels in normal and cancerous tissues.

(A) A schematic illustrating the nature of the p53 poly(A) SNP in its gene context. (B) Box plot showing the relationship between p53 mRNA abundance distribution (log2 transformed) and the variant genotype in blood samples. The central horizontal line indicates the median value of each distribution, upper and lower boundaries of the boxes indicate the 3rd and 1st quartile. N = 4710 [A/A homozygote], 193 [A/C heterozygote] and 2 [C/C homozygote]. (C) Box plot showing the relationship between TP53 mRNA abundance distribution (Log2 transformed) and the variant genotype in tumours. (D) Box plots showing the relationship between TP53 mRNA abundance distribution (Log2 transformed) and the variant genotype in tumours stratified on p53 mutational status.

status: wild type p53, missense mutated p53, oncogenic p53 (Fig. 3.6D), and tested the relationship between variant genotype and p53 expression levels within the individual groups. Considering that truncating mutations in *TP53* typically reduce p53 expression levels in tumours [206, 235], I had excluded *TP53* LOF mutations in this analysis. I observed that the minor allele C was consistently associated with 1.26- to 1.42-fold per-allele decrease in p53 expression levels in tumours, regardless of p53 mutational status (beta range: [-0.51, -0.33], $P < 0.036$; Fig. 3.6D).

To further investigate the molecular basis of the p53 poly(A) SNP in regulating p53 expression, an isogenic cell model with the two alleles of the variant was developed (in collaboration with Dr. Ping Zhang and Mr. Isaac Kitchen-Smith in the Bond group). Of note, the Hap1 cells utilised in the experiments contain a DNE

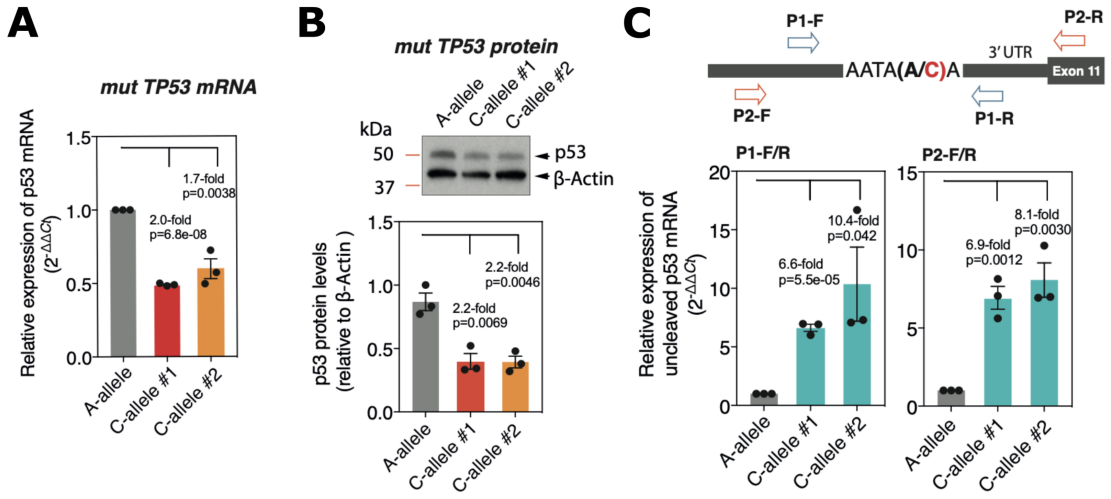

Figure 3.7. rs78378222 regulates p53 expression levels in isogenic cell models.

(A) A bar plot of p53 cDNA levels for each genotype in Hap1 cells, measured using qRT-PCR normalised to GAPDH. Error bars represent standard error of the mean (SEM) of 3 independent experiments. P-values were calculated using a two-tailed t-test. (B) A bar plot of p53 protein levels for each genotype in Hap1 cells, measured using densitometric analyses of results from Western blot analyses (upper pane) and normalised to β-actin. Error bars represent SEM of 3 independent experiments. P-values were calculated using a two-tailed t-test. (C) A schematic overview of the qRT-PCR strategy to measure the levels of unleaved TP53 mRNA in Hap1 cells of differing genotypes (upper). Two bar plots of uncleaved TP53 mRNA levels for each genotype in Hap1 cells, measured using qRT-PCR normalised to GAPDH (lower). Two sets of primers (P1-F/R and P2-F/R) were used to amplify the p53 pre-mRNAs.

missense mutation in *TP53* (p.S215G), which results in a mutated DNA-binding domain [236] and has been found in many cancer types (COSM43951), including breast cancer and glioma [226]. CRISPR/Cas9-mediated genome editing with homology-directed repair were used to generate clones with either the A-allele or the C-allele (Fig 3.6A). Consistent with above, cells with the C-allele were found to express lower amount of *TP53* mRNA relative to clones with the A-allele based on qRT-PCR analysis (2-fold decrease, P = 6.8e-08 for clone 1 and P = 0.0038 for clone 2; Fig 3.7A). Cells with the C-allele also had lower p53 protein levels (2-fold decrease, P = 0.0069 for clone 1 and P = 0.0046 for clone 2; Fig 3.7B). The impairment of 3'-end processing and subsequent transcription termination due to the C-allele have been proposed as a mechanism for the genotype-dependent p53 expression [234]. To investigate whether this is also the mechanism by which the C-allele reduces mutant p53 levels in cancer cells, the levels of *TP53* mRNA not cleaved at the con-

ical AAUAAA site (uncleaved) were measured relative to the cleaved transcripts. Up to 10-fold enrichment of uncleaved *TP53* mRNA in cells with the C-allele were observed compared to those with the A-allele (Fig 3.7C). Together, these findings demonstrate that the p53 poly(A) SNP regulates p53 expression levels in cancer cells.

3.2.3.2 rs78378222 and p53 mutational status interact to influence response to anti-cancer therapy

Mutated p53 in tumours is known to desensitise cells to radiotherapy [237], an important modality in anti-cancer treatment. Indeed, I found that the TCGA pan-cancer data supported this notion. Of the 7,021 patients of European ancestry (see also section 3.2.2), 848 patients had phenotype on radiation responses: 603 responders and 134 non-responders (see Methods, section 2.3.4). I found that p53 MUT was correlated with poorer response to radiation (OR = 1.6, $P = 0.021$; Fig 3.8A). This correlation was heightened when considering both somatic mutations and CN variation in *TP53*: p53 MUT & CN loss was also correlated with poorer response to radiation (OR = 2.2, $P = 0.0026$; Fig 3.8A). Furthermore, I observed that the minor allele C of the p53 poly(A) SNP, rs78378222, was correlated with poorer response to radiation among p53 WT tumours (OR = 5.6, $P = 0.011$; Fig 3.8B), but not among p53 MUT tumours (Fig 3.8B). Similarly, CN loss in *TP53* was also correlated with poorer response to radiation among p53 WT tumours (OR = 1.6, $P = 0.027$; Fig 3.8C), but not among p53 MUT tumours (Fig 3.8C). These findings suggest that the relative 2-fold reduction of wild type p53 levels in tumours due to the minor allele C of the p53 poly(A) SNP results in unfavourable response to radiotherapy.

Like in radiotherapy, mutated p53 in tumours also confers resistance to targeted DNA damaging chemotherapies [238]. As a result, therapeutic efforts have been focused on restoring wild type p53 activity to improve p53-mediated cell killing. The above analyses demonstrated how the p53 poly(A) SNP interacts with p53 mutational status in tumours on cancer risk and p53 expression levels. This points to another potential entry point for therapeutic interventions by manipulating p53 pathway activities based on commonly inherited genetic variants. I reason that p53

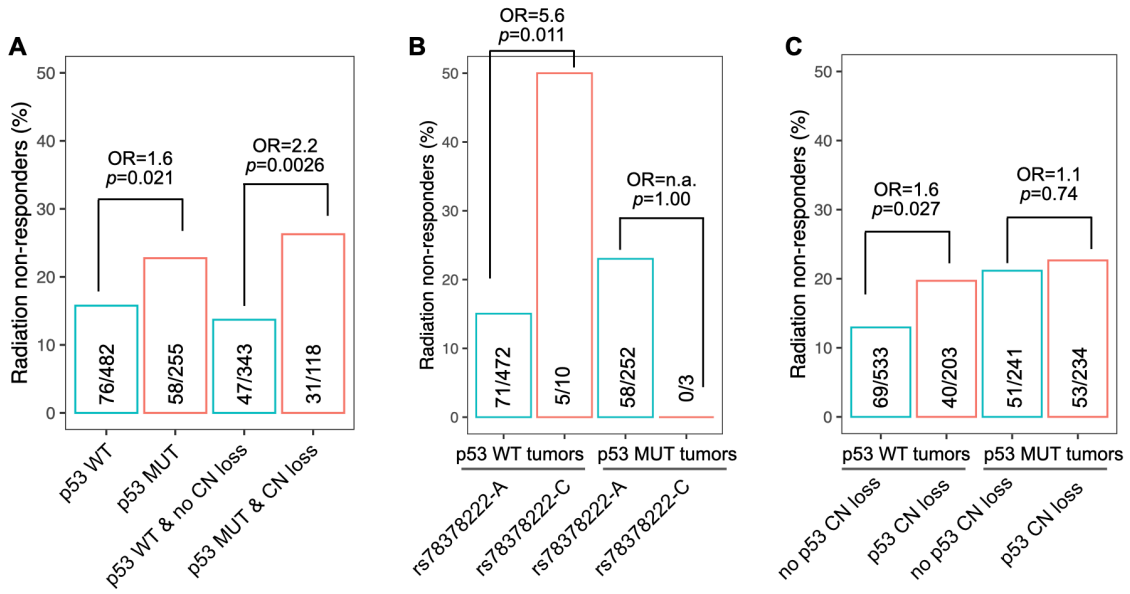

Figure 3.8. rs78378222 and p53 mutational status in tumours interact to influence cellular response to radiotherapy.

(A) A bar plot showing the percentage of radiation non-responders among patients differing in p53 mutational status. (B) A bar plot showing the percentage of radiation non-responders among patients with different variant genotype and p53 mutational status. (C) A bar plot showing the percentage of radiation non-responders among patients with different p53 mutational status and CN variation.

pathway genes that harbour TP53-shSNPs (see also section 3.2.1) could be utilised to achieve better p53-mediated cancer killing.

To test this idea, I used a drug screening dataset with both somatic genetic and gene expression data (Genomics of Drug Sensitivity in Cancer, or GDSC), where 304 compounds were tested across 1,001 cancer cell lines for their anti-cancer efficacy (Fig 3.9A). Of the 304 compounds analysed, 127 showed heightened sensitivity in cell lines with wild type p53 compared to those with mutated p53 (adjusted $P < 0.05$; Fig 3.9B). In particular, sensitivity to nutlin3-induced cell killing were was significantly associated with p53 WT status (fold change = 5.8; P = 2.6e-67). Nutlin3 activates p53 via blocking the inhibition of Mdm2 on p53. It has been shown that mutated p53 is refractory to Mdm2 inhibition, thus insensitive to nutlin3. As expected, cellular chemosensitivity to nutlin3 was correlated with Mdm2 expression levels in p53 WT cancer cells (P = 2.9E-09; Fig 3.9C), but not in p53 MUT cancer cells (P = 0.17; Fig 3.9D). This indicates that the higher the level of *MDM2* transcript in p53 WT cancer cells, the more sensitive the cells are to nutlin3-induced cell killing.

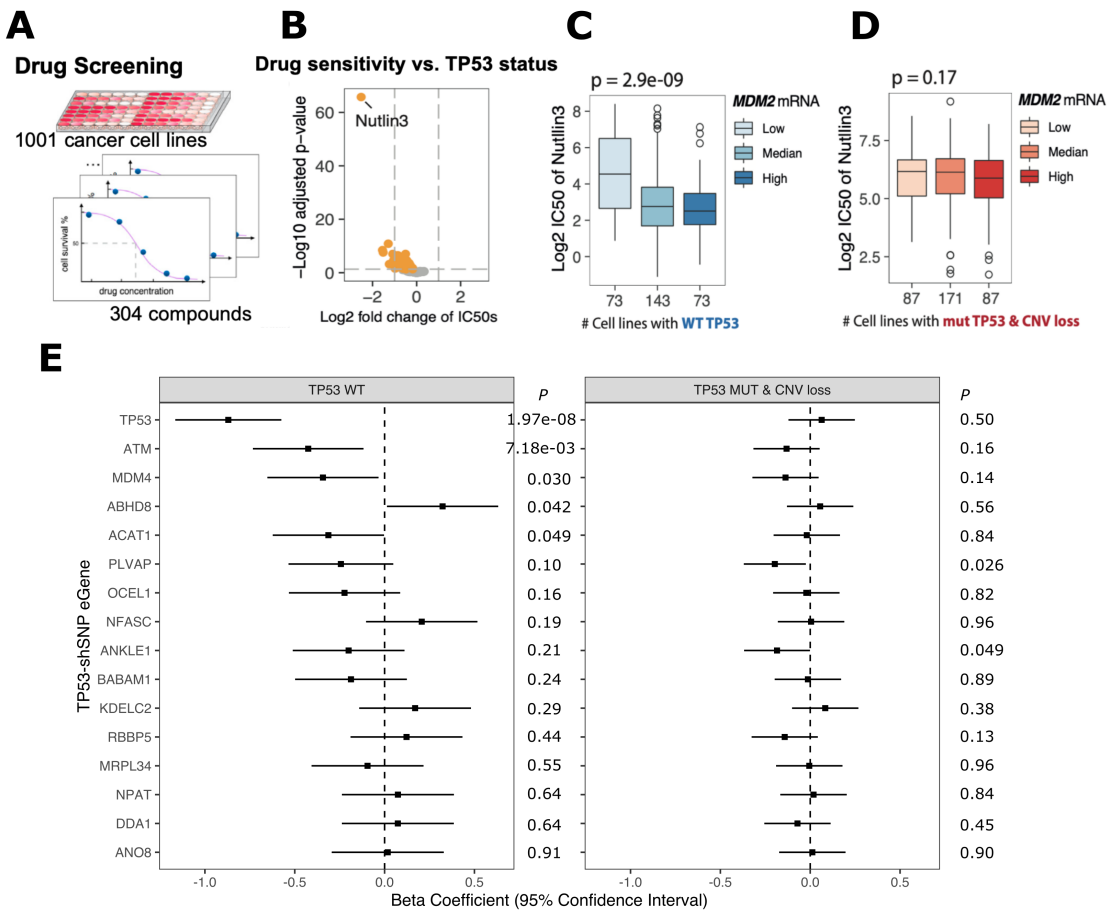

Figure 3.9. p53 pathway genes with TP53-shSNPs can aid in tuning cellular chemosensitivities to nutlin3.

(A) A schematic of the drug screen dataset (see Methods, section 2.3.5). (B) A volcano plot showing the correlation between cellular sensitivity to the 304 tested drugs and p53 mutational status in the 1,001 cancer cell lines, where the -log10 adjusted p-values (calculated from linear regression) were plotted against log2 fold change in IC50 values of each drug. The horizontal dashed line represents the adjusted p-value of 0.05. (C, D) Box plots showing the relationship between cellular sensitivity to nutlin3 treatment and Mdm2 expression levels in p53 WT (C) or in p53 MUT & CN loss (D) cancer cells. (E) Forest plots showing the relationship between eGene expression levels and cellular sensitivity to nutlin3 treatment in p53 WT (left panel) or in p53 MUT & CN loss (right panel) cancer cells.

Here, I am interested to examine whether TP53-shSNPs could provide information that can be utilised to further enhance cellular chemosensitivity to nutlin3. As mentioned in section 3.2.1, 110 TP53-shSNPs were mapped to 17 eGenes. Therefore, I compared the expression levels of these 17 eGenes in the cancer cells against the cells' sensitivity to nutlin3 treatment (see Methods, section 2.3.5). As shown in Fig 3.9E, I found that the expression of genes in the p53 pathway was associated with cellular chemosensitivity to nutlin3 treatment in p53 WT cells: higher p53, ATM, and Mdm4 expression levels are correlated with higher sensitivity to

nutlin3-induced cell-killing; but not in p53 MUT & CN loss cells. In addition, I observed that higher ABHD8 expression level was negatively associated with sensitivity to nutlin3 treatment in p53 WT cells, while higher ACAT1 expression level was positively associated with sensitivity (Fig 3.9E). ABHD8 and ACAT1 are involved in cellular metabolism, and *ABHD8* is a target gene of p53. These findings suggest that besides *TP53*, expression of other eGenes mapped by TP53-shSNPs also plays a role in influencing the cellular sensitivity to nutlin3 treatment. This implies that TP53-shSNPs highlight gene targets for modulating chemosensitivity to nutlin3-induced cell killing.

3.2.3.3 rs78378222 and p53 mutational status interact to influence cancer prognosis

Not only was mutated p53 in tumours associated with lack of response to therapy in human cancers, but it was also associated with worse clinical outcomes [239]. In the TCGA pan-cancer cohort, I found that patients with p53 MUT tumours had shorter progression free interval (PFI; OR = 1.36 [1.26-1.47], P = 2.92e-14) and poorer overall survival (OS; OR = 1.53 [1.41-1.67], P = 3.68e-24) when compared to those with p53 WT tumours (Fig 3.10); and that CN loss in *TP53* was associated with unfavourable outcomes (PFI: OR = 1.21 [1.12-1.31], P = 2.45e-06; OS: OR = 1.25 [1.15-1.36], P = 1.40e-07; Fig 3.10). Expectantly, p53 MUT & CN loss was associated with shorted PFI (OR = 1.45 [1.31-1.60], P = 8.60e-13) and poorer OS (OR = 1.64 [1.48-1.82], P = 1.95e-20) (Fig 3.10).

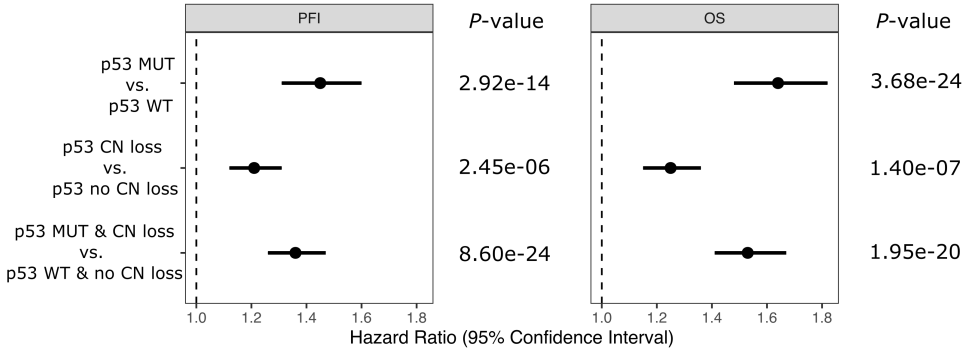

Figure 3.10. p53 mutational status in tumours is associated with patient survival outcomes in TCGA.

Forest plots show the relationship between p53 mutational status and progression free intervals (PFI) or overall survival (OS) among the TCGA patients. HR and log rank p values were calculated using a multivariate Cox regression model, adjusted for patient age, sex and tumour types.

To explore the prognostic value of the p53 poly(A) SNP, rs78378222, I studied whether this variant was associated with clinical outcomes in the TCGA cohort. I observed that the variant genotype was not associated with PFI when the cancer cases were mixed (pan-cancer: OR = 1.13 [0.89-1.44], $P = 0.32$; breast cancer: OR = 1.44 [0.35-5.87], $P = 0.61$; Fig 3.11A, C). However, when I stratified the cases by the p53 mutational status in tumours, I found that the minor allele C of rs78378222 was associated with shorter PFI in p53 WT tumours (pan-cancer: OR = 1.54 [1.12-2.12], $P = 7.4e\text{-}03$; breast cancer: OR = 6.11 [1.39-26.92], $P = 6.3e\text{-}03$; Fig 3.11B, D), but not in p53 MUT tumours (pan-cancer: OR = 0.68 [0.32-1.43], $P = 0.31$; breast cancer: OR = 0.81 [0.01-5.81], $P = 0.88$; Fig 3.11B, D).

These findings suggest that the variant genotype interacts with p53 mutational status in tumours to associate with patient survival outcomes ($P_{interaction} = 0.065$): p53 mutations and the minor allele C of rs78378222 was both associated with unfavourable clinical outcomes; meanwhile, the prognosis of individuals with the minor allele C and mutated p53 (the 'MUT-AC' curve in Fig 3.11B, D) did not differ significantly from those with the major allele A and wild type p53 (the 'WT-AA' curve in Fig 3.11B, D). Note that this is consistent with the germline by somatic interaction on cancer risk that the minor allele C was associated with increased risk for p53 WT tumours but with decreased risk for p53 MUT tumours (Fig 3.5A; see also section 3.2.2).

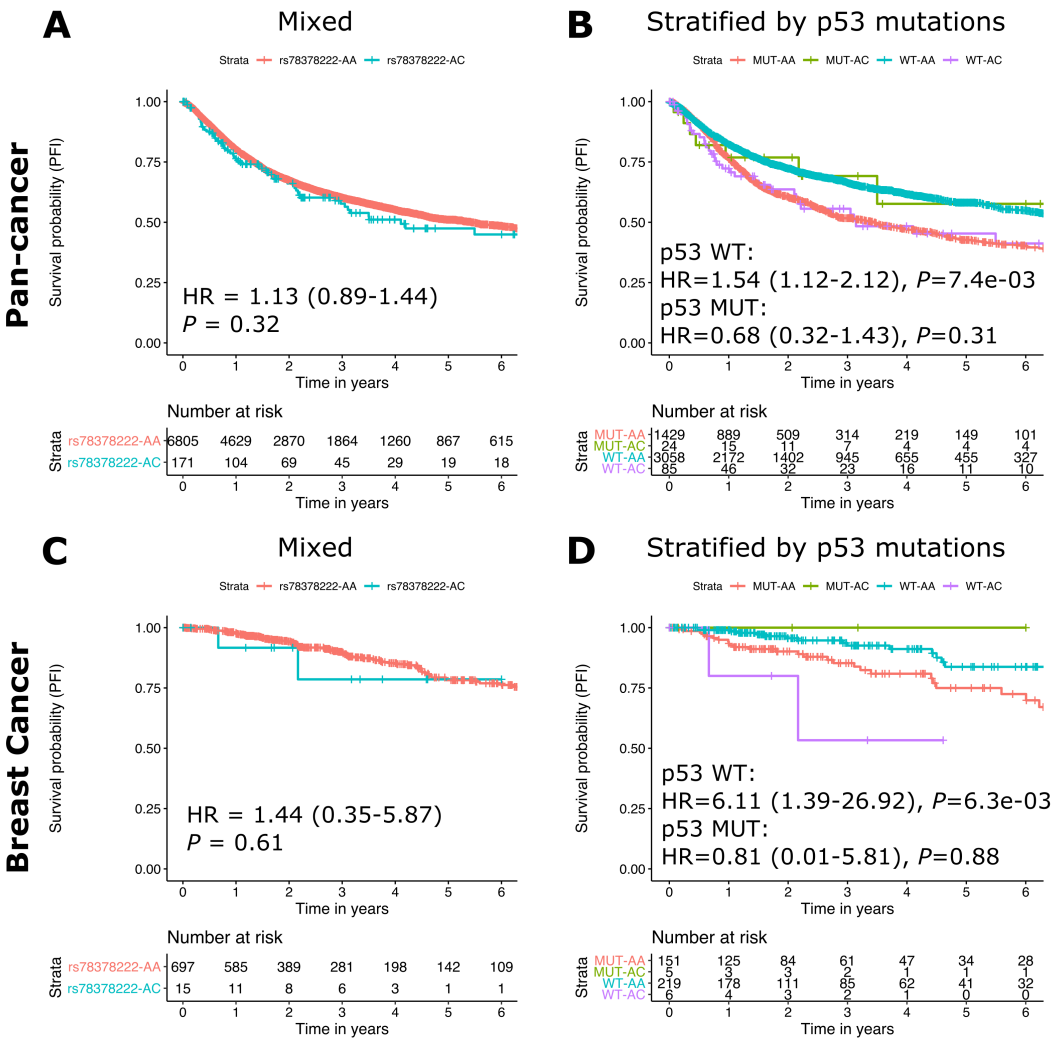

Figure 3.11. rs78378222 and p53 mutational status in tumours interact to influence cancer prognosis in TCGA.

(A, B) Kaplan-Meier survival curves of variant genotype and progression free interval (PFI) in the pan-cancer cohort, either the cases were mixed (A) or stratified by p53 mutational status (B). HR and log rank p values were calculated using a multivariate Cox regression model, adjusted for patient age, sex and tumour types. (C, D) Kaplan-Meier survival curves of variant genotype and PFI in the breast cancer cohort, either the cases were mixed (C) or stratified by p53 mutational status (D). HR and log rank p values were calculated using a multivariate Cox regression model, adjusted for patient age and sex.

3.3 Discussion

As discussed in section 1.1.5, inherited genetic variants associated with cancer can not only provide insight into cancer aetiology, but also be utilised for personalised cancer risk prediction and aid in clinical management. In the last two decades, standard cancer GWASs have been treating genotype-risk association as a relationship that is homogeneous in a population and independent of somatic driver events.

Recent studies showed that common inherited genetic variants were associated with somatic driver events and could influence tumour development (see section 1.3). In light of the emerging evidence, an in-depth study on inherited genetic variants that are predictive of susceptibility to cancers of specific molecular profile would be valuable, to further elucidate the influence of inherited genetic variants on risk, biology and prognosis of common cancers.

In this chapter, I investigated inherited cancer risk variants in the p53 tumour suppressor pathway to define germline by somatic interaction on cancer risk, response to therapy and prognosis. Using population-based multi-cancer datasets, I found that 110 variants were associated with subtype-specific risk, whose pattern was inversely correlated with p53 mutation frequencies in breast and ovarian cancer subtypes (TP53-shSNPs; Fig 3.4). These variants were mapped to genes that are enriched in the p53 pathway. I observed that the p53 poly(A) SNP, rs78378222, tracked with p53 mutational status in TCGA pan-cancer tumours (Fig 3.5). Therefore, this variant is a TP53-mSNP I hypothesised in section 3.1.3. Note that the minor allele of the p53 poly(A) SNP was associated with both increased and decreased cancer risk previously [51, 216, 222–230]. Considering that cancer types vary drastically in p53 mutation frequencies, the interaction between the p53 poly(A) SNP and p53 mutational status in tumours on cancer risk potentially provides an explanation for these conflicting findings.

By examining the relationship between variant genotype and gene expression levels, I showed that the minor allele C of the p53 poly(A) SNP was associated with lower p53 expression in both normal and cancerous tissues, regardless of p53 mutational status (Fig 3.6). An isogenic cell model developed in the Bond group demonstrated that compared to the major allele A, the minor allele C reduced p53 mRNA and protein levels by about two-fold (Fig 3.7). The molecular evidence supports a biological hypothesis as the following (Fig 3.12). The minor allele C disrupts the canonical PAS of *TP53* and hinders mRNA processing. As a result, cells carrying the C-allele produce less p53 than cells carrying the A-allele. In p53 WT tumours, this reduction in p53 protein levels compromises the tumour suppression function of wild type p53, thus conferring risk for cancer; in p53 MUT tumours, this

reduction in p53 protein levels mitigates the oncogenic effect of mutated p53 (see also section 3.1.2), thus reducing cancer risk. Alternatively, the relative depletion of the minor allele C in p53 MUT & CN loss tumours than in p53 WT & no CN loss tumours (Fig 3.5) suggests that carriers of the C-allele have a slightly lower likelihood to mutate *TP53* somatically during tumour development. It will be of interest to test both hypotheses in experimental models.

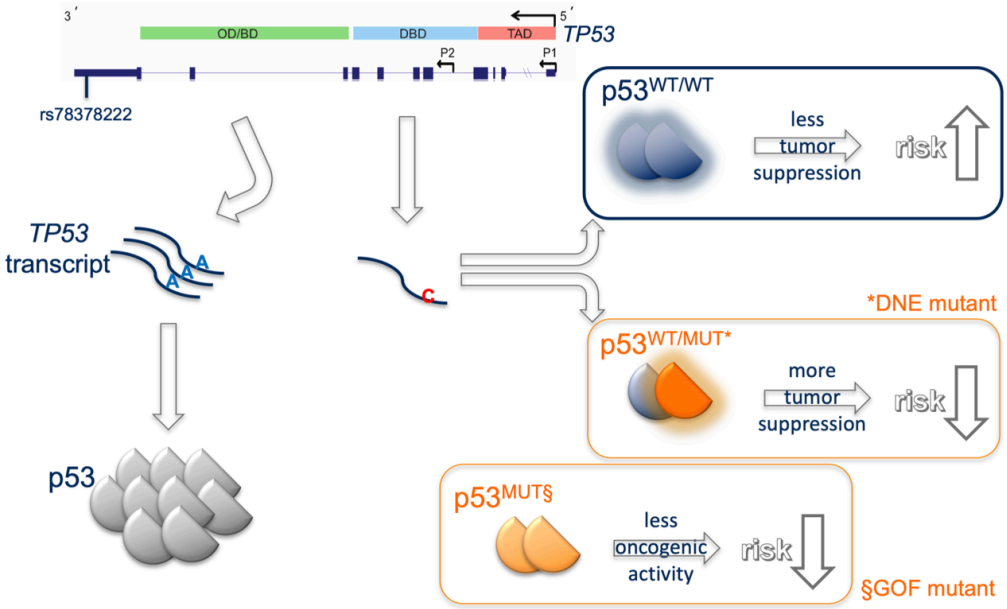

Figure 3.12. A hypothesis of the interaction between rs78378222 and p53 mutational status in tumours on cancer risk.

I showed that the two-fold reduction in p53 expression levels due to the minor allele C of the p53 poly(A) SNP was also associated with poor response to radiotherapy among patients with p53 WT tumours (Fig 3.8). This result has implications for anti-cancer therapies. Existing therapeutic strategies targeting somatic mutations in tumours usually have variable responses in the clinic, frequently high failure rate and eventually resistance to therapies. Somatic genetic heterogeneity in tumours is a major factor contributing to the differences in response to therapy and disease progression. In this chapter, using the p53 poly(A) SNP as an example, I demonstrated the inherited cancer risk variants could influence response to radiotherapy and potentially provide new entry point to enhance chemosensitivity. Therefore, understanding of such germline by somatic interaction could further inform personalised decision on anti-cancer therapy.

In addition, I found that rs7837822 genotype also interacts with p53 mutational status in tumours to associate with patient survival outcomes, consistent with the risk association patterns (Fig 3.11). In particular, the protective effect of the minor allele C in patients with p53 MUT tumours demonstrates that like somatic driver mutations, germline genotype can also be informative for cancer risk stratification and prognosis prediction. If such germline by somatic interaction is not exclusive to the p53 pathway, PRS-based cancer risk prediction might need to account for molecular information about the tumour towards a more personalised genetic profile.

Together, these results illustrate how an inherited genetic variant could play an active role during tumour development. The study in this chapter has been centred on the p53 tumour suppressor pathway. As a result, it is unclear whether the hypothesised germline by somatic interaction on cancer risk and prognosis exists in other cancer hallmark pathways. Another drawback of the subtype-stratified study design is that it requires data of large-cohort with both germline and somatic information for sufficient discovery power. These cohorts are rare, especially for systematic studies on cancer prognosis. Nevertheless, it would be of wide interest to study germline by somatic interaction in other cancer hallmark pathways, and to further investigate the influence of inherited genetic variants on risk, biology and prognosis of common cancers.

Chapter 4

Germline by somatic interaction in the RAS-MAPK pathway

4.1 Introduction

4.1.1 Genetic susceptibility to colorectal cancer

As introduced in section 1.2.4, colorectal cancer (CRC) is a leading cause of cancer morbidity and mortality worldwide. CRC has a strong genetic component: individuals with Lynch syndrome have a significantly increased risk of developing colorectal cancer (20-80% lifetime risk; see section 1.1.2), and common inherited genetic variants were associated with increased CRC risk (see section 1.1.3). Also, CRC cases tend to aggregate in families [42, 44]. Monozygotic twins have significantly higher concordance for CRC than dizygotic twins [4, 43]. It has been estimated that up to 40% of the variation in CRC risk is attributed to genetic factors, in the form of common inherited genetic variants [43, 52, 240].

The use of genome-wide association studies (GWASs) in cancer has been discussed in section 1.1.3. So far, CRC GWASs have reported 175 risk variants mapping to 138 distinct genomic loci in European populations [52, 241]. These CRC risk variants primarily target genes involved in three key biological processes: (1) Wnt (e.g., *WNT4, CTNNB1, TCG7L2, LRP1*) and BMP signalling pathways (e.g., *BMP2, BMP4, BMP5, BMP7, GREM1, GREM2, SMAD6, SMAD7, SMAD9, TBX3*); (2)

MYC activity (e.g., *MYC, TCF7L2*); and (3) genomic integrity maintenance (e.g., *TERT, RTEL1*) and DNA repair (e.g., *MLH1, POLD3*). Collectively, CRC risk variants identified thus far account for approximately 15% of the familial relative risk of CRC [52], leaving a large proportion unexplained. While increasing effective sample sizes could potentially identify additional inherited genetic variants associated with CRC risk, research strategies beyond conventional cancer GWAS are urged to address the conundrum of 'missing heritability of cancer' (see also section 1.1.4; [50]).

Existing GWASs have focused on identifying inherited risk loci for developing CRC as a single end point. In other words, CRC has been treated as a homogeneous binary disease trait (having CRC or not). However, what is less explored is genetic susceptibility to well-defined subtypes of CRC, in particular those defined by specific cancer driver mutations. If certain inherited genetic risk loci are specific for one molecular subtype of CRC but not others, their association signals with disease status would be dampened or diminished when cases are mixed. If this is true, such subtype-specific risk variants could account for part of the 'missing heritability' in CRC.

The concept of germline by somatic interaction on cancer risk has been exemplified in chapter 3. In addition, as discussed in section 1.3.1, four loci showed association with risk in opposite directions for luminal A-like and triple-negative breast cancer subtypes (Table 1.3; [51]). In light of these results, it would be of interest to investigate germline by somatic interaction on CRC subtype-specific risk.

4.1.2 Somatic mutations in colorectal cancer

Intestinal epithelium undergoes rapid tissue renewal [242]. This depends on the activity of stem cell populations that generate new epithelial cells to replace lost and damaged ones due to natural attrition or tissue injury [243]. This process of intestinal regeneration is tightly regulated by signalling pathways that balance cell proliferation and differentiation, such as the Wnt, EGFR, and BMP/TGF-β signalling pathways. Mutations that dysregulate the activities or the dynamics of these signalling pathways endow colorectal cancer cells with niche-independent

growth advantages during tumorigenesis (Fig 4.1). For instance, 70-80% of sporadic CRC cases have biallelic inactivation of the tumour suppressor gene *APC* due to somatic truncating mutations [21, 22]. This loss of APC disrupts the phosphorylation and degradation of β-catenin, which mimics the constitutive activation of Wnt ligand-mediated signalling, promoting cell proliferation [20, 242]. Besides mutations in *APC*, another characteristic of CRCs are somatic mutations in the oncogene *KRAS*, which result in hyperactivated EGFR signalling, promoting cell survival and tissue growth (discussed further in section 4.1.3; [244, 245]).

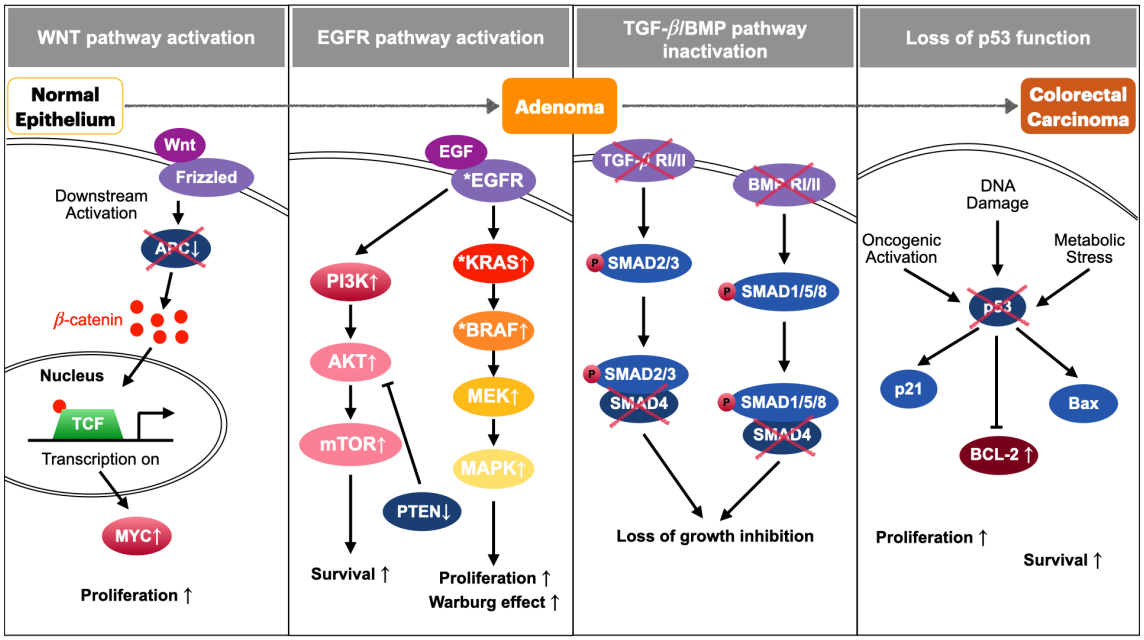

Figure 4.1. An overview of somatic mutations and pathway dysregulation in CRC tumorigenesis and progression.

The sequential scheme of tumour development is adapted from the genetic model for CRC proposed by Fearon and Vogelstein in 1990 [245]. Oncoproteins affected by activating mutations are marked with asterisks, and tumour suppressors affected by inactivating mutations are indicated by red crosses. Up- or down-arrows indicate altered protein abundance or activity in tumours. APC: adenomatous polyposis coli; BMP: bone morphogenetic protein; EGF: epidermal growth factor; MAPK: mitogen-activated protein kinase; MEK: mitogen-activated protein kinase kinase; mTOR: mammalian target of rapamycin; TCF: T cell factor; TGF-β: transforming growth factor β.

Genome sequencing surveys have provided a systematic overview of the patterns of somatic alterations in CRCs, emphasising the prevalence of somatic mutations in known cancer driver genes, including *APC, KRAS, TP53, PIK3CA* and *SMAD4* (see also section 1.2.1; Table 1.2; [105]). Non-hypermutated tumours (typically defined as those with $<$ 12 non-silent mutations per million exonic nucleotides) comprise 84% of all CRCs, among which *APC* (81%), *TP53* (60%), *KRAS* (43%),

NRAS (10%) and *BRAF* (3%) are frequent mutated somatically [105]. In contrast, hypermutated tumours (16% of all CRCs) exhibit genetic instability in simple tandem repeat sequences termed microsatellite DNA (microsatellite instability, or MSI; see also section 1.2.4; [125]), often associated with defective DNA mismatch repair (MMRd) or mutations in the exonuclease domain of the POLE replicative DNA repair polymerase [105]. Among the hypermutated tumours, somatic mutations frequently affect *ACVR2A* (63%), *APC* (53%), *BRAF* (46%; mostly V600E), *MSH3* (40%), *MSH6* (40%), *KRAS* (30%) and *NRAS* (10%).

As shown in Fig 4.1, the EGFR pathway integrates information from extracellular growth factors to activate intracellular growth-stimulating transcriptional program. Somatic mutations disrupting EGFR pathway activity were frequently observed in aberrant crypt foci [246], which are flat colonic epithelial lesions with altered glandular architecture but usually not dysplasia. Many driver mutations in the pathway have been shown to contributes to CRC development in human cell lines and mouse models [247–249]. Overall, 59% of non-hypermutated and 80% of hypermutated CRCs harbour somatic mutations in this pathway [105]. KRAS, NRAS and BRAF all function in the EGFR pathway. As listed above, somatic mutations in *KRAS* and *BRAF* show mutual exclusivity in CRC tumours: 43% non-hypermutated CRCs mutate *KRAS* and 3% mutate *BRAF*, whereas 30% hypermutated CRCs mutate *KRAS* and 46% mutate *BRAF* [105]. Somatic activating mutations or amplifications in ERBB family genes, which encode receptor tyrosine kinases including EGFR, are also detected in CRC tumours (5% in non-hypermutated tumours and 20% in hypermutated tumours) [105].

Considering the importance of the EGFR pathway in CRC, I reasoned that a study of genetic susceptibility to molecular subtypes of CRC that differ in KRAS mutational status or EGFR pathway mutational status would be of wide scientific interest. A further pragmatic consideration is that, given the prevalence of somatic mutations in *KRAS* among CRCs (i.e., 60% cases with wild type KRAS vs. 40% with mutated KRAS), a stratified study on KRAS mutational status will also be the most statistically powered, compared to mutational frequencies of other cancer driver genes in CRC.

4.1.3 The KRAS protein in colorectal cancer

The RAS family of small-G proteins, notably KRAS, HRAS, and NRAS, are intracellular signal transducers, which play essential roles in mediating signalling cascades downstream of growth factor receptors, regulating cell survival and proliferation [250]. KRAS, in particular, is a critical hub in the cell signalling circuitry, whose activity is under tight control [251, 252]. Upon EGF stimulation on the cell membrane, KRAS transduces activating signals to the MAPK pathway, which in turn phosphorylates and activates ERK, one of the best-known MAPKs (Fig 4.1). Activated ERK then induces the expression of target genes and promote cellular survival and proliferation [253]. In normal physiology, EGF stimulation activates ERK only transiently due to active receptor degradation and negative pathway feedback [254].

As a guanosine triphosphatase (GTPase), KRAS is a binary molecular switch that cycles between active guanosine triphosphate (GTP)-bound and inactive guanosine diphosphate (GDP)-bound states, a mechanism that is highly conserved among GDP/GTP binding proteins. Activation of RAS proteins is stimulated by the guanosine exchange factor (GEF) protein, which displaces GDP from the nucleotide binding site and results in spontaneous GTP binding. The active RAS proteins are inactivated upon GTP hydrolysis to GDP, a process that is catalyzed by the GTPase activating proteins (GAPs). In the GTP-bound active state, KRAS is able to bind and activate its effector proteins recruited to the plasma membrane, such as RAF-kinases (BRAF and RAF1), PI3K and RalGDS [255].

Over 90% of somatic mutations in *KRAS* occur at codons 12, 13 or 61 (i.e., 'hotspot mutations'), and other non-canonical codons are also mutated at low frequencies (Fig 4.2A). All the canonical mutations in *KRAS* prompt the loss of the intrinsic and/or the GAP-stimulated GTPase activity, leading to a constitutive activation of KRAS (Fig 4.2B; [256, 257]). In addition, mutations that decrease affinity for nucleotides were also observed in tumours (e.g., A146 mutations), allowing GDP to dissociate rapidly, thus resulting in abnormal accumulation of KRAS in the GTP-bound form even without stimulus from upstream signals or GEFs [258]. Therefore, all these somatic mutations result in constitutively active KRAS in the cell, hyperactivating the EGFR pathway and driving growth factor-independent proliferation

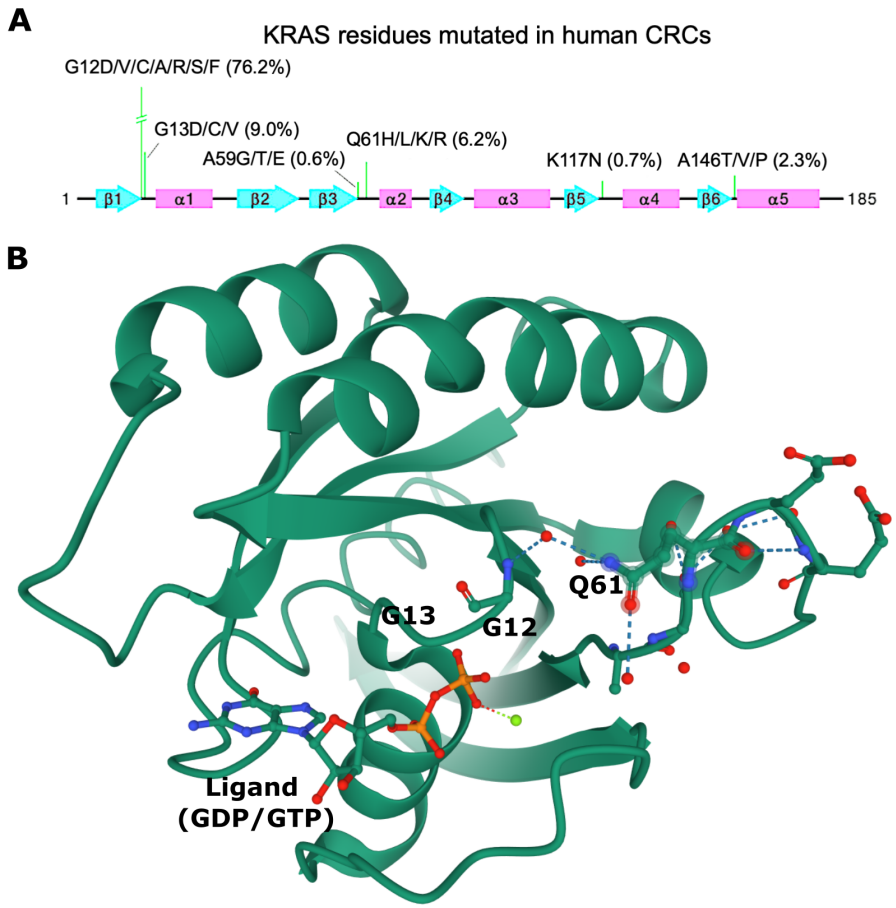

Figure 4.2. A structural overview of KRAS mutational spectrum in human CRCs.

(A) KRAS residues with more than 3 mutations detected among TCGA CRC genomes in the cBioPortal database are shown with frequencies of mutations in parentheses. All these residues are in the switch regions or in the G1-G5 sequence motifs that plays crucial roles in recognition and binding of guanine nucleotide. (B) Positions of KRAS residues affected by hotspot mutations (G12, G13, and Q61) are mapped on the tertiary structure of KRAS (PDB: 4Q21). Note that all three residues are located in the ligand-binding pocket of KRAS.

of cancer cells. In particular, a number of mouse models highlighted that mutated KRAS increases the susceptibility of the intestinal mucosa to chemical carcinogenesis [259] and leads to accelerated tumour formation in combination with loss of *APC* [248, 260].

4.1.4 A hypothesis of KRAS-mSNPs in colorectal cancer

The analyses in chapter 3 have demonstrated that inherited genetic variants can interact with p53 mutational status in tumours to influence risk, response to therapy and prognosis of common cancers. Recently, inherited genetic variants in *MC1R,*

IRF4 and *FLA2G6* that were known to predispose low-penetrant melanoma risk have shown subtype-specific association for *BRAF V600E* mutations (Table 1.4; [138, 140]). However, the interaction between inherited cancer risk variants and mutational status of the EGFR pathway in CRC is undefined. In this chapter, I aim to address this question, and to refine genotype-risk association in CRC subtypes with or without KRAS hotspot mutations.

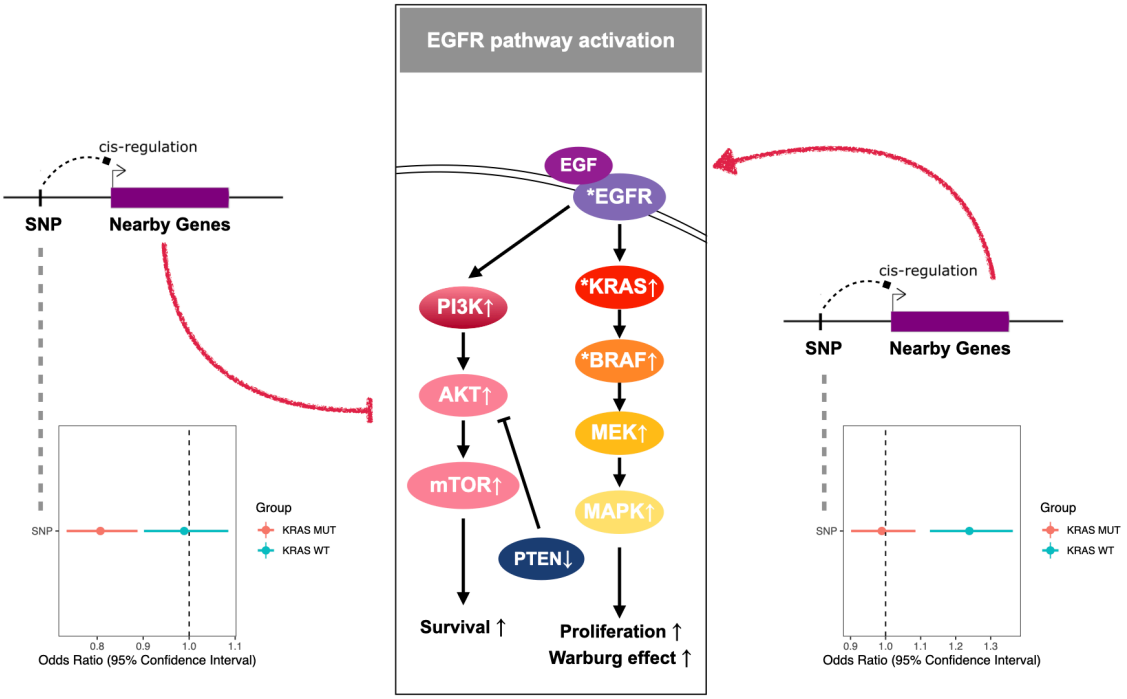

Figure 4.3. A schematic for the hypothesised interactions between inherited genetic variants and KRAS mutational status on CRC risk.

Left panel: some inherited genotypes could associate with KRAS MUT CRC but not with KRAS WT CRC if it mitigates EGFR pathway activity or bolsters pathway activities that antagonise the EGFR pathway. Right panel: some inherited genotypes could associate with risk for KRAS WT CRC specifically if it sensitises EGFR pathway activities like the mutated KRAS or enhances pathway activity that complements the function of the EGFR pathway in promoting tumorigenesis.

The interaction between inherited genetic variants and somatic mutations affecting the EGFR pathway could take many forms. Similar to the p53 poly(A) SNP (Fig 3.5), inherited genotypes could associate with risk for KRAS WT CRC specifically if it sensitises EGFR pathway activities like the mutated KRAS or enhances pathway activity that complements the function of the EGFR pathway in promoting tumorigenesis (Fig 4.3). Conversely, inherited genotypes could associate with KRAS MUT CRC but not with KRAS WT CRC if it mitigates EGFR pathway activity or bolsters pathway activities that antagonise the EGFR pathway (Fig 4.3).

In either case, such genetic variants would demonstrate an imbalance in risk association between CRC subtypes differing in KRAS mutational status, which I refer to as the KRAS-mSNPs onwards (Fig 4.3). In this chapter, I show that stratified association studies by KRAS mutational status in tumours represent a strategy to identify the hypothesised KRAS-mSNPs and to refine genotype-risk association in CRC by molecular subtype.

4.2 Results

4.2.1 Identification of comprehensive CRC cohorts for stratified association studies

To investigate the potential interaction between inherited genetic variants and KRAS mutational status in CRC, I proposed to perform stratified association studies by KRAS mutational status in tumours. As discussed in sections 1.1.3 and 3.2.1, this required large population cohorts of cases and controls, with both germline data and molecular information about the tumours. Therefore, I assembled datasets from four UK-based clinical trials on CRC, totalling 6,630 CRC patients (Fig 4.4; [164–167, 261]). Considering the similarity in clinicopathological features of trial participants and the compatibility of technology used for genotyping and molecular subtyping, I pooled the trial cohorts whenever possible to maximise statistical power (see Methods, section 2.1.2.1).

The identification of KRAS-mSNPs consisted of two phases (Fig 4.4). The phase I study included 4,368 cases of high-risk stage II or stage III CRC (American Joint Committee on Cancer, or AJCC TNM staging) participating in the VICTOR and QUASAR2 trials (referred to as the VQ2 study onwards), matched with 2,673 controls from the 1958 Birth Cohort in the Wellcome Trust Case Control Consortium 2 (WTCCC2). The phase II study included of 2,244 patients with metastatic CRC enrolled in the COIN and COIN-B trials (collectively referred to as the COIN study onwards), paired with 2,800 controls from the WTCCC2 Blood Service Group. Details of the clinical trials have been reported previously [164–167, 261] and are provided in the Methods (section 2.1.2.1).

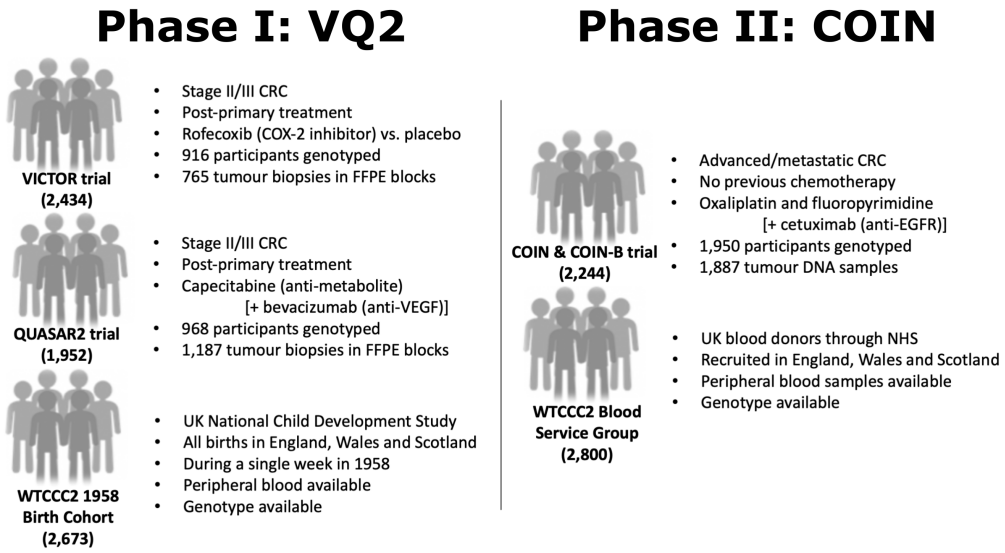

Figure 4.4. An overview of clinical trial cohorts assembled for stratified association studies by KRAS mutational status in CRC.

Considering the similarity in clinicopathological features of trial participants and the compatibility of technology used for genotyping and molecular subtyping, I pooled the trial cohorts whenever possible to maximise statistical power (see Methods, section 2.1.2.1).

The primary endpoint of the stratified association studies was to compare allelic differences of individual variant genotypes in tumours with KRAS wild type (WT) or mutant (MUT) against those in healthy controls analysed in parallel (Fig 4.5A). To achieve this goal, SNPs of cases and controls were typed (see Methods, section 2.1.2.1); wherever tumour blocks are available, tumour DNA were extracted and sequenced for identifying somatic mutations in common CRC driver genes, including *KRAS, BRAF* and *NRAS* (see Methods, section 2.1.2.3). According to the presence or absence of hotspot mutations at codons 12, 13 or 61, I assigned KRAS mutational status to tumours as either WT or MUT (Fig 4.5B). Cancer cases with both genotype and KRAS mutational information were included in the KRAS mutations-stratified association studies. In total, 1,703 KRAS WT and 992 KRAS MUT cases were included in both phases (KRAS mutation frequency: 36.8%), against 5,472 healthy controls (Table 4.1).

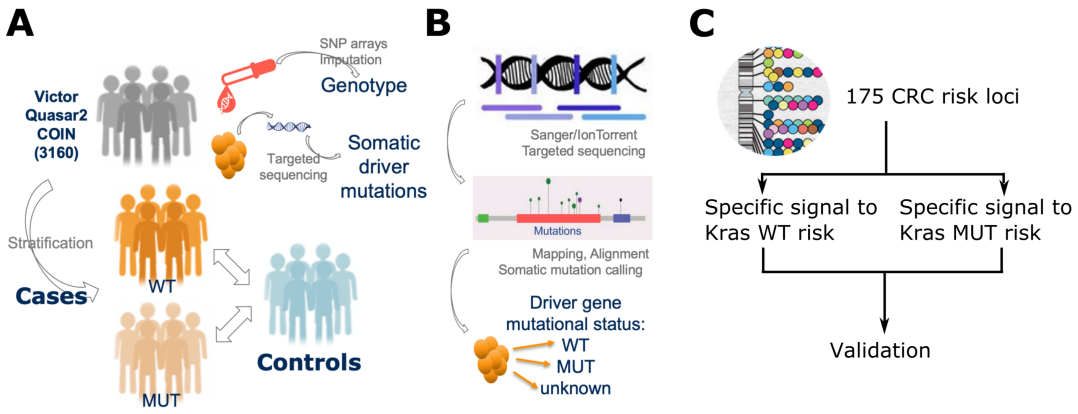

Figure 4.5. An overview of the study design for the stratified association studies by KRAS mutational status in CRC.

(A) Schematics describe the strategies used to ascertain genotypes of inherited genetic variants and somatic mutations of participants in the trial cohorts. Blood DNA was extracted and genotyped on SNP arrays for selected markers. Marker genotypes were then imputed to infer genotypes of variants across the genome (Methods, section 2.1.2.2). Whenever possible, tumour DNA was extracted, and somatic driver mutations were identified by targeted sequencing (Sanger or IonTorrent; see Methods, section 2.1.2.3). (B) On the basis of detected somatic mutations, the tumour samples were then classified into those carrying the WT or MUT gene product. (C) To investigate whether known risk variants interact with KRAS mutational status in tumours to affect CRC risk by molecular subtype, I examined a total of 175 known CRC susceptibility variants (see Methods, section 2.2.2) for association signals that are specific to cancers with either KRAS WT or MUT in phase I and II datasets. Stratum-specific signals were then aggregated via meta-analysis (see Methods, section 2.3.6). Stratum-specific signals that were validated in both phases were prioritised. A formal power calculation indicated that this study provided at least 98% power to detect an odds ratio of 1.20 per allele CRC risk for 75% of variants (MAF > 0.10), and 82% power for variants with a MAF of 0.05. However, the statistical power to detect comparable association signals for variants of low frequency (MAF < 0.05) was less than 80% (see Methods, section 2.3.6).

	Cases			Controls
Cohort	**KRAS WT**	**KRAS MUT**	**Total**	
VQ2	812	421	1,233	2,673
COIN	891	571	1,462	2,800
Total	1,703	992	2,695	5,472

Table 4.1. An overview of sample sizes in the sample cohorts.

4.2.2 KRAS-Stratified association studies in CRC

KRAS-stratified GWASs were conducted in phase I and II independently, and combined by meta-analysis in the respective strata (KRAS WT or MUT) as follows. First, I performed stringent quality control measures on the marker genotype data obtained from SNP arrays (see Methods, section 2.1.2.1). Second, I imputed the

genotypes of over 8 million SNPs using the UK10K and 1000 Genomes Project phase 3 data as the reference panels (see Methods, section 2.1.2.2). Third, I filtered out SNPs with a minor allele frequency (MAF) $< 1\%$ and imputation quality score < 0.8 (see Methods, section 2.1.2.2). Fourth, I assessed the difference in allele frequencies of individual variants between cases and controls in each stratum (KRAS WT or MUT), using logistic regression assuming an additive genetic model (see Methods, section 2.3.6). Finally, I aggregated the effect size estimates in phase I and II studies by an inverse-variance weighted fixed-effect meta-analysis (see Methods, section 2.3.6). A formal power calculation indicated that a study including 2695 cases and 5472 controls provided at least 98% power to detect an odds ratio of 1.20 per allele CRC risk for 75% of variants (MAF > 0.10), and 82% power for variants with a MAF of 0.05 (see Methods, section 2.3.6). However, the statistical power to detect comparable association signals for variants of low frequency (MAF < 0.05) was less than 80%.

Stratified association studies in the individual phases revealed a total of 9 genomic loci associated with KRAS WT CRC at genome-wide significance ($P <$ 5e-08, logistic regression), and 7 loci associated with KRAS MUT CRC (Fig 4.6A-D). All loci are novel, and almost all variants are of low frequency (MAF < 0.05). I found that one locus on chromosome 20 was validated in the meta-analysis, which was associated with risk for KRAS MUT tumours at genome-wide significance (Fig 4.6F; Supplementary Data Table 4.1). However, I observed that the signals at this locus primarily arose from Phase II, and that the variants at this locus were also associated with risk for KRAS WT tumours at suggestive significance (Supplementary Data Table 4.1). Given that this study has insufficient statistical power for identifying cancer risk variants of low frequency, it remains unclear whether these putative risk loci represent genuine candidates for the KRAS-mSNPs I hypothesised (see also section 4.1.4).

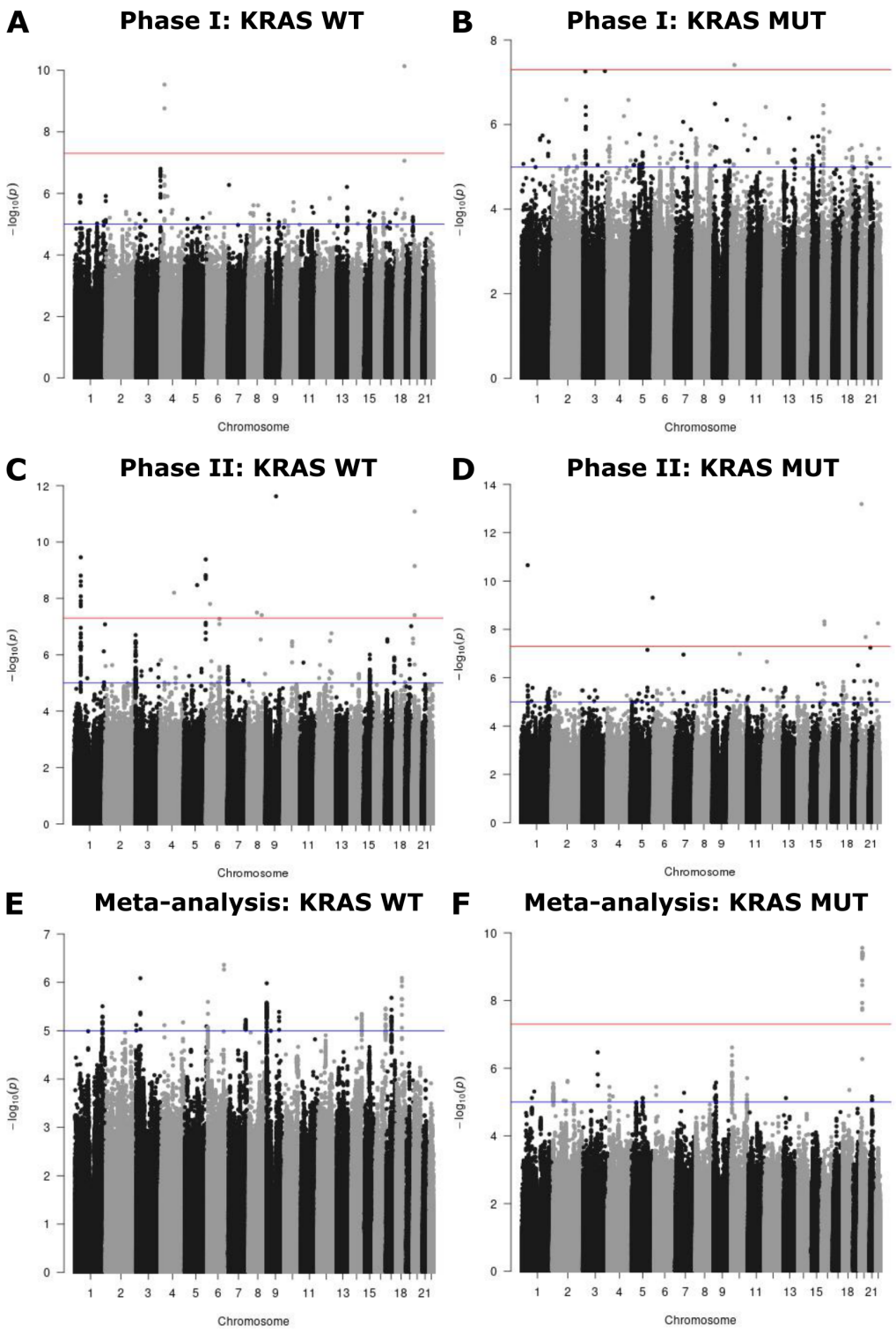

Figure 4.6. Manhattan plots of stratified association studies by KRAS mutational status in CRC.

(A, B) Phase I study for cases with KRAS WT (A) or MUT (B). (C, D) Phase II study for cases with KRAS WT (C) or MUT (D). (E, F) Meta-analysis of phase I and II studies for cases with KRAS WT (E) or MUT (F). The red lines mark the genome-wide significance level (P = 5e-08), and the blue lines mark the suggestive significance level (P = 1e-05).

4.2.3 KRAS-stratified association studies refine known genotype-risk association by CRC molecular subtype

Even though this study was not statistically powered for a genome-wide search for KRAS-mSNPs, it could be used to explore the interaction between inherited CRC risk variants and KRAS mutational status in tumours and potentially refine associations of known CRC risk loci by molecular subtype. To test this idea, I adopted a candidate approach to focus on CRC risk variants that were previously identified in large-cohort studies.

As of February 2020, 175 CRC risk variants have been reported (see Methods, section 2.2.2; Supplementary Data Table 4.2). Initially, I examined the allelic differences of these 175 variants between cases and controls in both strata (KRAS WT or MUT) in phase I and II independently, to identify association signals that are specific to either KRAS WT or MUT CRC risk. Specifically, using a P-value cut-off at 0.05, I tested whether certain inherited genotypes were associated with KRAS WT CRC risk, but not with KRAS MUT CRC risk; or they were associated with KRAS MUT CRC risk, but not with KRAS WT CRC risk (Fig 4.5C). Next, I checked whether the trend of associations was consistent in Phase I and II studies. That is, variant of interest were associated with KRAS WT CRC risk in both phases, but not associated with KRAS MUT CRC risk in either phase; or they were associated with KRAS MUT CRC risk in both phases but not with KRAS WT CRC risk in either (Fig 4.5C).

Out of the 175 CRC risk variants, one (rs10795668, 10p14, MAF = 0.32) fulfilled the selection criteria for the hypothesised KRAS-mSNPs. In particular, the minor allele A of rs10795668 was depleted in KRAS MUT CRC when compared to healthy controls, thus associated with lower risk for KRAS MUT CRC (meta-analysis: OR = 0.78, 95% confidence interval (CI) = [0.70, 0.86], P = 2.80e-06, FDR = 4.9e-04; Fig 4.7A). However, the same allele showed comparable frequencies between KRAS WT CRC and healthy controls, thus not associated with risk for KRAS WT CRC (OR = 0.98 [0.90, 1.07], P = 0.65, FDR = 1; Fig 4.7A). Of note, this variant has been reported by multiple CRC GWASs, both in European and East Asian populations [52, 241, 262–268].

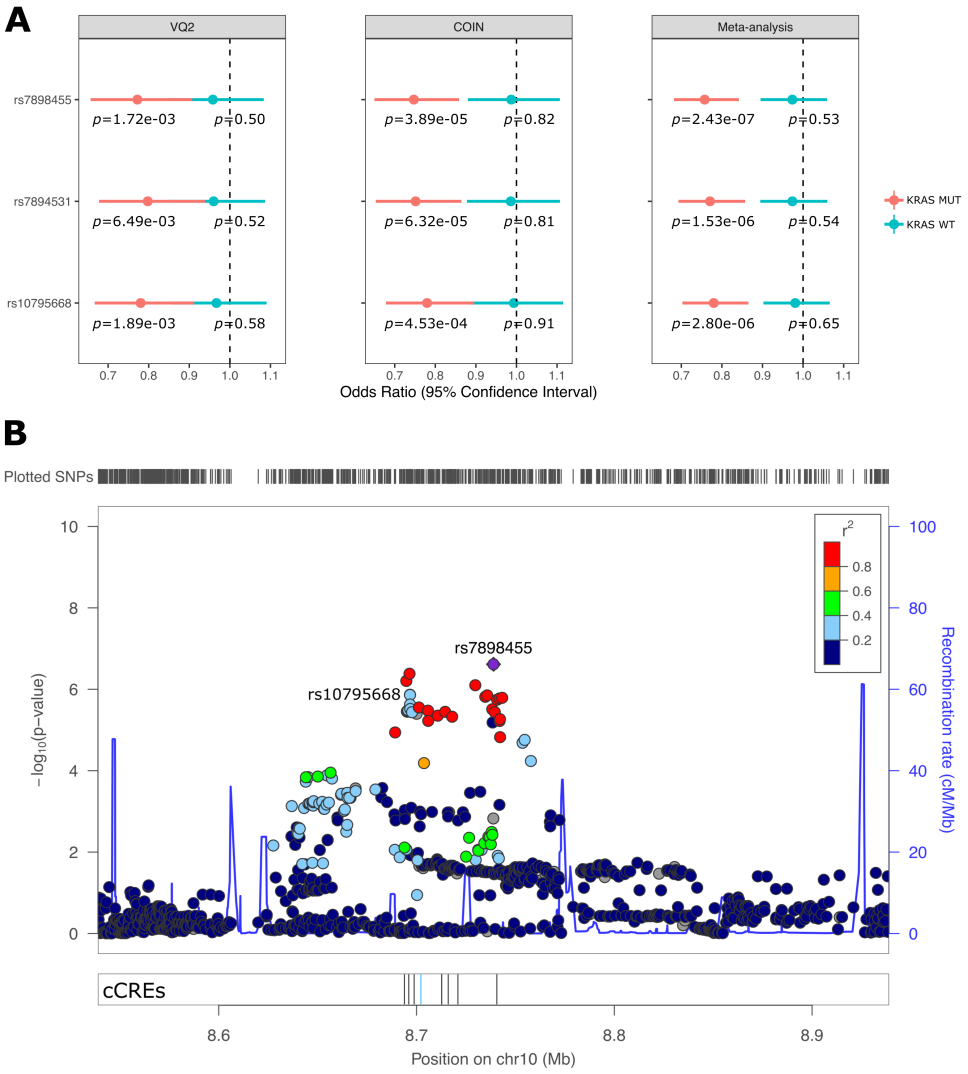

Figure 4.7. KRAS-mSNPs at the 10p14 locus.

(A) Forest plots of per-allele ORs for representative inherited genetic variants at the 10p14 locus in stratified association studies based on KRAS mutational status in tumours. Columns correspond to the respective sample cohorts or meta-analysis. (B) Fine mapping association results at 10p14 in the KRAS MUT association studies. P-values from the meta-analysis are shown against chromosomal positions of individual variants at the locus. The top variant rs7898455 (purple) is in high linkage disequilibrium (LD) with the known CRC risk variant rs10795668. The colour bar reflects the estimated correlation coefficient (r^2) between each variant and the top variant rs7898455 (purple), derived genotype data from non-Finnish European population in 1000 Genomes Project (phase 3 release). No genes reside in the locality, and the nearest genes are GATA3, TAF3, and ATP5C1 500 kb in the upstream. Candidate cis-regulatory elements (cCREs) known in this LD block has been marked for their locations in the region. Black cCREs were characterised for high H3K27ac contents within the elements and the cyan cCRE represents element high in CTCF binding (ENCODE data; Supplementary Data Table 4.4). Genomic positions shown are based on Genome Reference Consortium Human Build 37 (GRCh37).

Spanning over 50 kb on chromosome 10p14 (8.68-8.72 Mb; GRCh37), this locus has a structure of extensive linkage disequilibrium (LD; Fig 4.7B). Therefore, I also

examined the associations of other 24 variants at this locus that are in LD with rs10795668 ($r^2 > 0.8$; European population). In total, 21 of the 25 linked variants showed specificity in CRC risk association by KRAS mutational status; 15 of these 21 variants demonstrated consistent trend of specificity for KRAS MUT CRC risk (Fig 4.7B; Supplementary Data Table 4.3). Fine-mapping revealed the lead SNP at this locus is rs7898455 (KRAS MUT: OR = 0.76 [0.68-0.84], P = 2.43e-07, FDR = 0.001; KRAS WT: OR = 0.97 [0.90-1.06], P = 0.53, FDR = 1; Fig 4.7A).

Three variants (rs10795668, rs7894531, rs11255841; MAF range: [0.31-0.32]) within this LD block have been reported to associate with CRC susceptibility before [52, 263, 266]. The minor alleles of all three variants were found to be less frequent among CRC cases compared to controls, thus associated with lower CRC risk, with characteristic ORs estimated to be in the range of [0.84-0.88]. This suggests the presence of a common protective haplotype representative of the minor alleles for lower CRC risk at this locus. In particular, the variant rs10795668 represents the most robust signal at this locus so far [262, 269–278], with the minor allele consistently associated with lower CRC risk (ORs in the range of [0.84-0.94]). Intriguingly, the current study identified that these alleles were all associated with lower risk for KRAS MUT CRC, but not with risk for KRAS WT CRC at all (Fig 4.7A; Supplementary Data Table 4.3).

In summary, stratified association study by KRAS mutational status in CRC represent a powerful method to refine association patterns of known CRC risk variants by molecular subtype. I identified 15 linked variants on chromosome 10p14 that were specifically associated with risk for KRAS MUT CRC, but with risk for KRAS WT CRC. The minor alleles of these variants were associated with decreased CRC risk, consistent with conventional GWASs where tumours were unstratified by KRAS mutational status. Nevertheless, the OR estimates in this stratified association study (in the KRAS MUT stratum) were notably greater than previous unstratified GWASs (stratified OR = 0.78 versus unstratified OR range: [0.84, 0.94]). These results suggest that stratified association study can refine genotype-risk association that is specific to cancers with specific driver mutations.

4.2.4 Genotype by KRAS mutational status interaction on CRC risk is independent of BRAF mutations

Besides KRAS hotspot mutations, the EGFR pathway can also be activated by BRAF mutations in CRC (10%; Table 4.2). To test whether the observed interaction between variants at 10p14 and KRAS mutational status on CRC risk could be related to BRAF mutations in tumours, I performed another set of stratified association study by KRAS-BRAF mutational status in CRC, using the same sample cohorts (see Methods, section 2.1.2). In total, 1,299 KRAS&BRAF WT and 1,222 KRAS/BRAF MUT cases were included in both phases (mutation frequency: 48.5%), against 5,472 healthy controls (Table 4.2).

	Cases							Controls
	KRAS		BRAF		KRAS-BRAF			
Cohort	WT	MUT	WT	MUT	WT	MUT	Total	
VQ2	812	421	1,099	160	646	574	1,233	2,673
COIN	891	571	1,194	112	653	648	1,462	2,800
Total	1,703	992	2,293	272	1,299	1,222	2,695	5,472

Table 4.2. An overview of samples sizes in the KRAS/BRAF-stratified association studies.

As in section 4.2.3, I examined the interaction between the 175 CRC risk variants and KRAS-BRAF mutational status on CRC risk in the sample cohorts. I found that the locus on chromosome 10p14 represented the strongest signal of interaction with in the meta-analysis; again, the variants were specifically associated with risk for KRAS/BRAF MUT CRC (lead SNP: rs7898455; OR = 0.79 [0.72-0.87], P = 1.89e-06; Fig 4.8). However, stratifying tumours based on KRAS-BRAF mutational status (KRAS&BRAF WT vs. KRAS/BRAF MUT) did not improve the association of rs10795668 with KRAS MUT CRC risk (OR = 0.81 [0.73-0.89], P = 1.03e-05; Fig 4.8).

Considering that KRAS and BRAF mutations are mutually exclusive in CRC (as discussed in section 4.1.2) and that KRAS signals via both RAF1 and BRAF [254], this result suggests that the association of rs10795668 with risk for KRAS MUT CRC observed in section 4.2.3 is likely independent of BRAF mutational status in CRC.

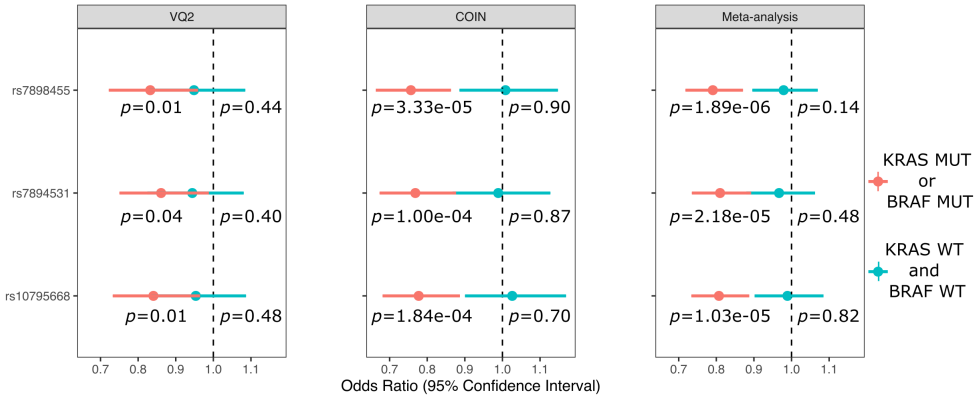

Figure 4.8. KRAS-mSNPs in KRAS-BRAF-stratified association studies.
Forest plots of per-allele ORs for representative inherited genetic variants at 10p14 in stratified association studies by KRAS/BRAF mutational status in CRC. Columns correspond to the respective sample cohorts or meta-analysis.

4.2.5 Further analyses of candidate variant rs10795668

4.2.5.1 rs10795668 genotype and KRAS mutational status in TCGA tumours

To further examine the relationship between rs10795668 and KRAS mutational status in tumours, I utilised an additional comprehensive dataset from the TCGA pan-cancer cohort. As in section 3.2.2, I selected 7,021 patients of European ancestry, who were diagnosed with 31 different tumour types. In this section, I aim to test whether the minor allele A of the variant rs10795668 is depleted in KRAS MUT tumours, as seen in section 4.2.3.

In this pan-cancer cohort, 7.6% of patients (530 out of 6,969) have KRAS hotspot mutations in their tumours. These KRAS MUT tumours consist of 24 different cancers: 30.0% are CRC (COAD and READ), 25.1% are lung adenocarcinoma (LUAD), 19.6% are pancreatic adenocarcinoma (PAAD), 9.4% are uterine corpus endometrial carcinoma (UCEC), 4.5% are stomach adenocarcinoma (STAD). When I compared variant genotype by KRAS mutational status in tumours (Fig 4.9A), I found that the variant rs10795668 was not significantly associated with KRAS mutational status in the pan-cancer cohort (OR = 0.99 [0.87-1.13], $P = 0.46$; logistic regression, adjusted for tumour types; Fig 4.9B); nor was it among the CRC tumours (OR = 1.2 [0.88-1.64], $P = 0.14$; logistic regression; Fig 4.9B). Nevertheless, I observed a trend that tumour types with increasing KRAS mutation frequency had lower frequency

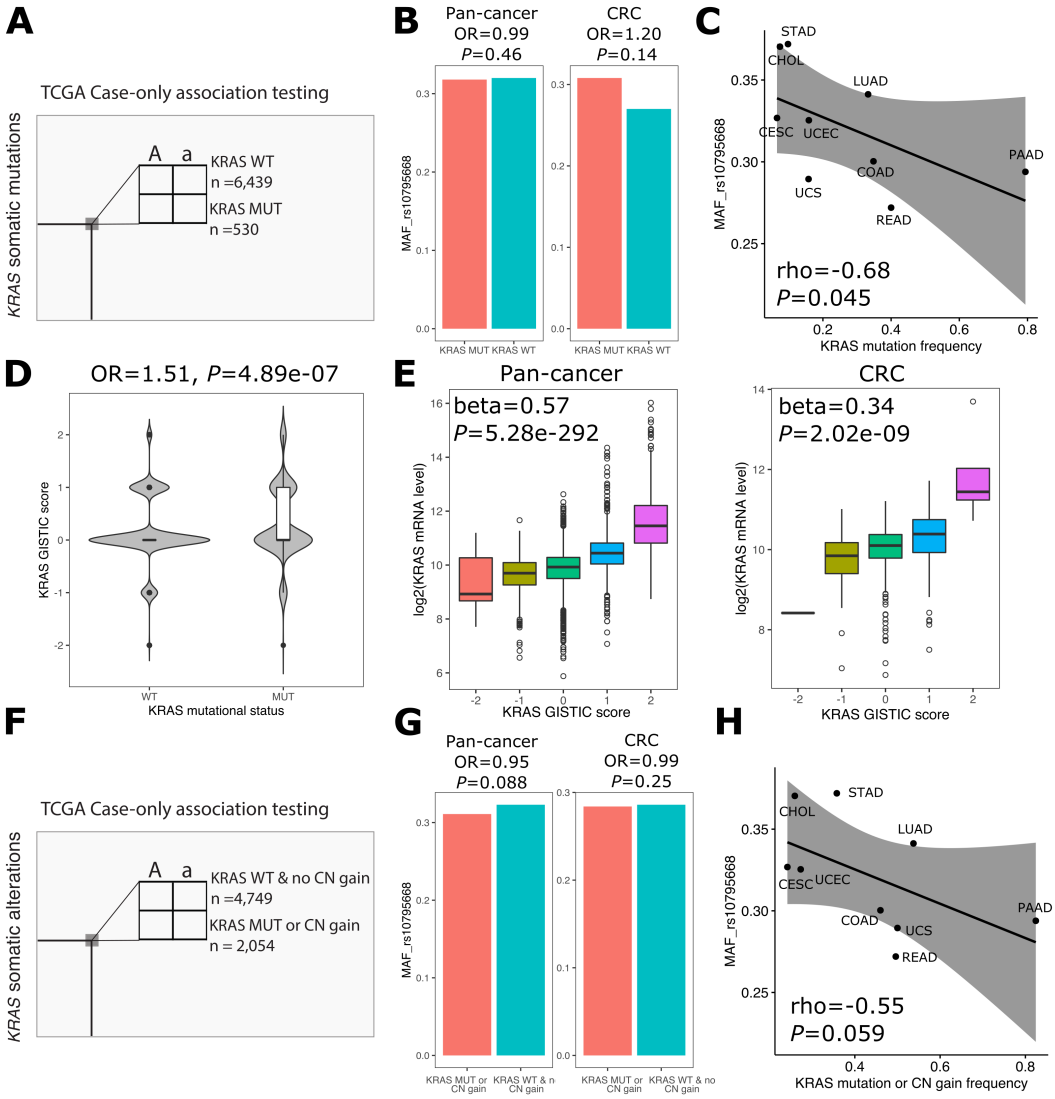

Figure 4.9. The relationship between rs10795668 genotype and KRAS somatic alterations in TCGA tumours.

(A) A schematic of the association testing between the genotype of an inherited genetic variant and KRAS mutational status among TCGA cases. (B) Bar graphs of the allelic difference of the variants rs10795668 between KRAS WT and MUT tumours in the TCGA pan-cancer or CRC cohort. (C) Scatter plot of the minor allele frequency of rs10795668 and KRAS mutation frequency in individual tumour types in the TCGA cohort (with a KRAS mutation frequency > 5%). The relationship between genotype and KRAS mutational status in individual tumour types was evaluated by Spearman correlation test. (D) A violin plot of KRAS GISTIC score against KRAS mutational status in the TCGA pan-cancer cohort. (E) A box plot showing the relationship between KRAS GISTIC score and KRAS expression levels in the TCGA pan-cancer or CRC cohort. (F) A schematic of the association testing between the genotype of an inherited genetic variant and KRAS alteration status in TCGA cases. (G) Bar graphs of the allelic difference of the variant rs10795668 between tumours with (i) KRAS WT no CN gain and (ii) KRAS MUT or CN gain in the TCGA pan-cancer or CRC cohort. (Continued on the following page.)

(Previous page.) (H) Scatter plot of the minor allele frequency of rs10795668 and KRAS alteration frequency in individual tumour types in the TCGA cohort (with a KRAS mutation frequency > 5%). The relationship between genotype and KRAS mutational status in individual tumour types was evaluated by Spearman correlation test. CESC: Cervical squamous cell carcinoma and endocervical adenocarcinoma; CHOL: Cholangiocarcinoma; COAD: Colon adenocarcinoma; LUAD: Lung adenocarcinoma; PAAD: Pancreatic adenocarcinoma; READ: Rectum adenocarcinoma; STAD: Stomach adenocarcinoma; UCS: Uterine Carcinosarcoma; UCEC: Uterine Corpus Endometrial Carcinoma.

of the minor allele A (rho = -0.68, $P = 0.045$; Spearman correlation test; Fig 4.9C).

Somatic mutations and CN gains as well as mutant ASI in *KRAS* frequently co-occur in tumours, which were correlated with increased mutant allele transcription and gene activity [129, 164, 279]. In the TCGA pan-cancer cohort, 25% of the tumours have CN gain in the *KRAS* region; 30.2% have either hotspot mutations or CN gain in *KRAS*; 2.6% have both hotspot mutations and CN gain in *KRAS*. In particular, 33.2% of KRAS MUT tumours show CN gain in the *KRAS* region (32% are CRC tumours), while 24.3% of KRAS WT tumours show CN gain, suggesting that KRAS MUT is correlated with CN gain in *KRAS* (OR = 1.55, $P = 1.30e\text{-}05$; Fisher's exact test). Moreover, I found that KRAS MUT was strongly correlated with higher KRAS GISTIC score (OR = 1.51, $P = 4.89e\text{-}07$; logistic regression test, adjusted for tumour types; Fig 4.9D), and that CN gain *KRAS* was correlated with higher KRAS expression levels (beta = 0.57, $P = 5.28e\text{-}292$; linear regression; Fig 4.9E). As a result, CN gain in *KRAS* could also sensitive the EGFR pathway for activation, akin to KRAS hotspot mutations that render the EGFR pathway constitutively active.

When I stratified the TCGA patients into two groups based of KRAS somatic alterations (MUT or CN gain vs. WT & no CN gain) and performed an association test on variant genotype against KRAS somatic alterations in tumours (Fig 4.9F), I found that despite not statistically significant, the frequency of the minor allele A was slightly lower in pan-cancer tumours with KRAS MUT or CN gain than those with KRAS WT & no CN gain (OR = 0.95 [0.87-1.02], $P = 0.088$; logistic regression, adjusted for tumour types; Fig 4.9G). However, I did not observe a similar finding in CRC tumours (OR = 0.99 [0.73-1.34], $P = 0.25$; logistic regression; Fig 4.9G), possibly due to limited sample size (n = 275). Of note, I didn't observe a significant

correlation between rs10795668 MAF and KRAS alteration frequency across tumour types either (rho = -0.55, $P = 0.059$; Fig 4.9H).

Together, the TCGA dataset did not provide strong support for the hypothesis that rs10795668 significantly tracks with KRAS hotspot mutations or CN gain in tumours. Additional analysis would be necessary to further examine this relationship.

4.2.5.2 rs10795668 genotype and nearby gene expressions in TCGA tumours

As mentioned in section 4.2.3, the LD block harbouring KRAS-mSNPs situates in an intergenic region on chromosome 10p14 (Fig 4.7B). Despite numerous CRC GWAS reporting associations of variants in this LD block with CRC risk, interpreting the functional consequences of these variants remain a challenge due to the missing link between variant genotypes and cancer driver gene or pathway activity in this region.

None of the 15 KRAS-mSNPs are known *cis*-eQTLs (see Methods, section 2.3.3). Since eQTLs identification mainly relies on gene transcript levels measured in normal tissues, this lack of association does not rule out the role of KRAS-mSNPs on regulating gene expression in cancerous tissues. Especially, I identified eight candidate *cis*-regulatory elements (cCREs) that are located within this LD block (see Methods, section 2.2.4; Fig 4.7B). These cCREs are cell type agnostic (i.e., they are identified in both normal and cancerous tissues; [185]). Seven of these elements are marked with high H3K27ac content, an epigenetic modification associated with higher activation of transcription (i.e., active enhancer marker); while one flags high CTCF binding activity, which facilitates chromatin looping and long-range chromatin interactions [185]. The presence of these regulatory elements within this LD block suggests that this region could be involved in regulating the expression of nearby or distal genes.

In total, there are 8 genes near this locus (define by a maximal distance of 2Mb; Fig 4.11). I found that only the expression of *ATP5C1*, about 850 kilobases upstream of rs10795668, was correlated with rs10795668 genotype. *ATP5C1* encodes the gamma subunit of ATP synthase, which plays a role in cellular metabolism in

the mitochondria. One study indicated that the rs10795668-A allele was correlated with 1.21-fold increased expression of *ATP5C1* in a sample of 40 microsatellite stable and CpG island methylation phenotype-negative CRCs (for which KRAS mutational status was not determined), but not in paired adjacent normal tissues [280].

Cancer cells are commonly found to have switched their mode of metabolism from respiration in the mitochondria to glycolysis in the cytosol; a phenomenon termed as the 'Warburg effect' [281, 282]. One mechanism to trigger this switch is a decrease in the expression of mitochondrial oxidative phosphorylation chain genes such as *ATP5B*, which encodes the beta subunit of ATP synthase [283, 284]. Given the central role of ATP synthase in cellular respiration, Loo et al. postulated that increased levels of ATP synthase in the presence of the rs10795668-A allele would strengthen respiration in the mitochondria and mitigate cancer risk by curbing the Warburg effect [280]. This fits well with previous GWAS findings that the rs10795668-A allele was associated with lower CRC risk [52, 241, 262–268].

Cancers driven by mutated KRAS are characterised by Warburg effect to promote tumour growth (Fig 4.1), and they typically show elevated nutrient uptake, glycolysis, glutaminolysis, synthesis of fatty acids and nucleotides [285]. Therefore, enhanced mitochondrial respiration function could potentially antagonise RAS-MAPK pathway-directed metabolic reprogramming in tumours.

To further investigate the relationship between rs10795668 genotype and *ATP5C1* expression levels in tumours, I examined variant genotype and *ATP5C1* mRNA levels in TCGA (see Methods, section 2.3.4). I observed that the rs10795668-A allele tends to associate with lower *ATP5C1* mRNA levels among pan-cancer tumours that were not distinguished by KRAS mutational status, though not statistically significant (beta = -0.022, $P = 0.08$, linear regression; Fig 4.10A). Among KRAS WT tumours, the A allele was associated with 1.03-fold lower *ATP5C1* expression per allele (beta = -0.036, $P = 7.1e\text{-}03$, linear regression; Fig 4.10B); whereas among KRAS MUT tumours, the A allele was associated with 1.08-fold higher *ATP5C1* expression per allele (beta = 0.11, $P = 0.013$, linear regression; Fig 4.10B). A formal testing indicated that rs10795668 genotype interacts with KRAS mutational status on *ATP5C1* gene expression in tumours ($P_{interaction} = 1.6e\text{-}03$). Adjustment for tu-

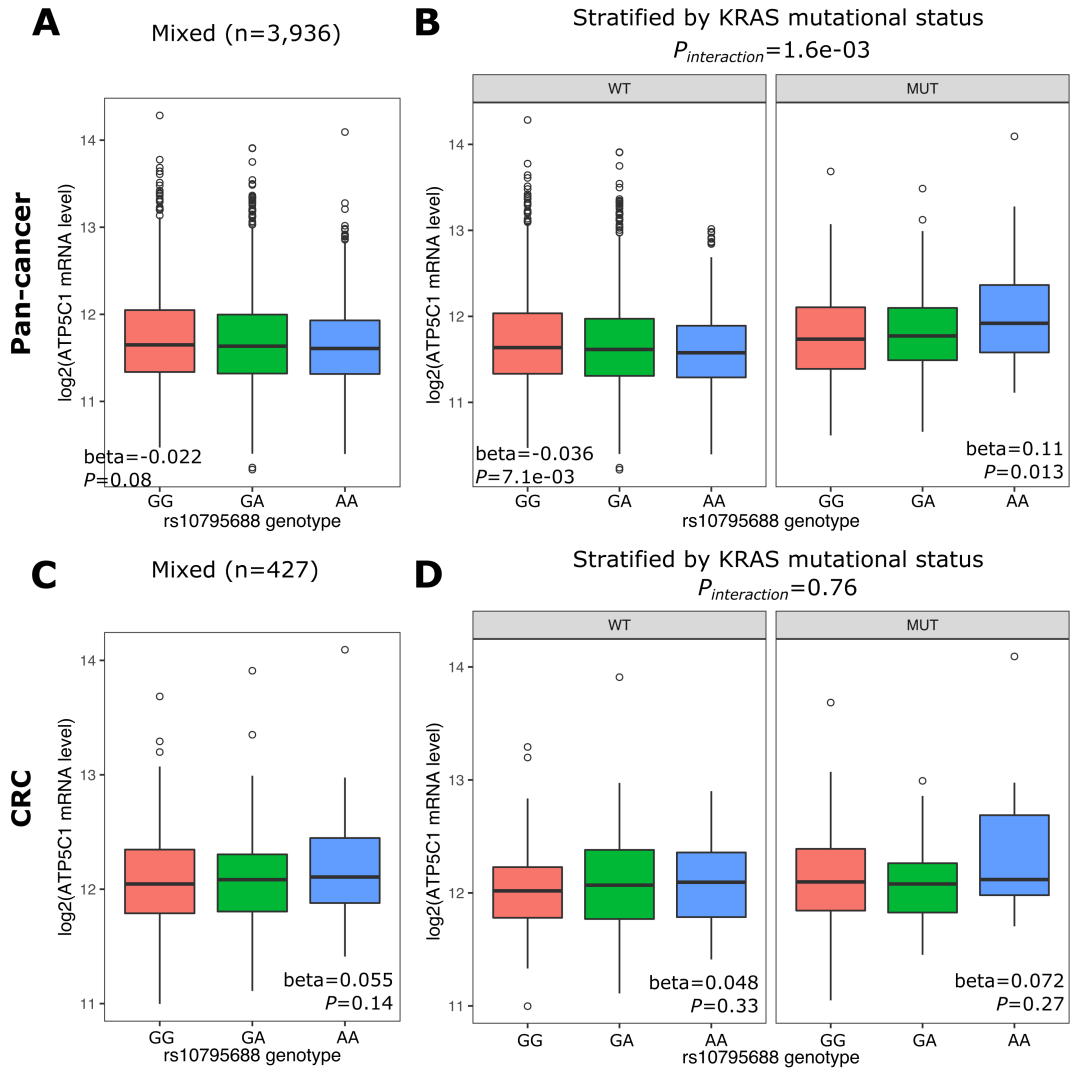

Figure 4.10. rs10795668 genotype by KRAS mutational status interaction on ATP5C1 gene expression in TCGA tumours.

(A) A box plot of the relationship between rs10795668 genotype and ATP5C1 mRNA levels in TCGA pan-cancer tumours. (B) Box plots show the relationship between rs10795668 genotype and ATP5C1 mRNA levels in TCGA pan-cancer tumours stratified by KRAS mutational status. (C) A box plot of the relationship between rs10795668 genotype and ATP5C1 mRNA levels in TCGA CRC tumours. (D) Box plots show the relationship between rs10795668 genotype and ATP5C1 mRNA levels in TCGA CRC tumours stratified by KRAS mutational status. Betas and nominal P-values were obtained from linear regression. Interaction p-values were obtained by evaluating the multiplicative term of genotype and KRAS mutational status in a linear model.

mour types did not affect the statistical significance (adjusted $P_{interaction}$ = 2.4e-03), indicating that this genotype by KRAS mutational status on gene expression was independent of tumour types. These findings suggest that the functional role of rs10795668 on *ATP5C1* gene expression in tumours depends on KRAS mutational status. Meanwhile, I did not find similar interaction among TCGA CRC tumours

(Fig 4.10C, D), possibly due to limited sample sizes (n = 427 [Mixed], 184 [KRAS WT], and 118 [KRAS MUT]).

In summary, the findings in this section provide a potential mechanistic explanation for the genotype-risk association observed in section 4.2.3 that the minor allele A of rs10795668 was exclusively associated with lower risk for KRAS MUT CRC (Fig 4.7A). Among KRAS MUT CRC, the minor allele A was associated with increased *ATP5C1* expression, which enhances respiration in the mitochondria, thus counteracts Warburg effect and reduce CRC risk (Fig 4.11). One possibility of the dependency of rs10795668 genotype on KRAS mutational status to impact nearby gene expression in tumours is that gene regulation at this locus might be contingent on hyperactive RAS-MAPK pathway activities. This could be a direction for future experimental investigation.

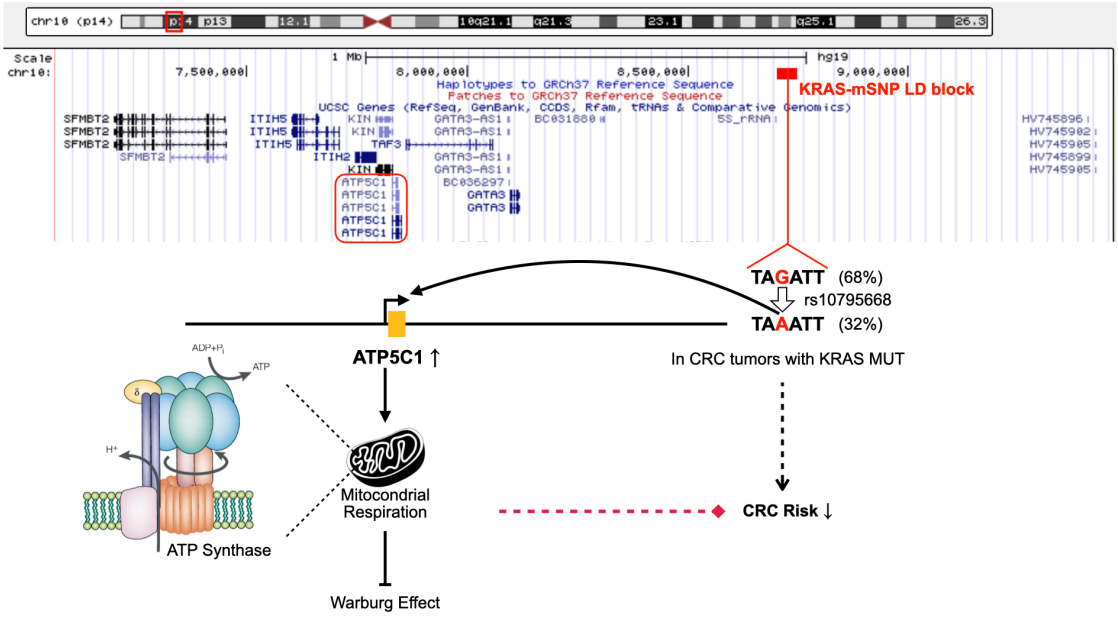

Figure 4.11. A hypothesis for the identified interaction between rs10795668 genotype and KRAS mutational status on nearby gene expression and CRC risk.

In KRAS MUT CRC, the minor allele A allele selectively upregulates the expression of ATP5C1. Higher abundance of ATP5C1 enhances ATP synthase activity in the mitochondria, thus strengthens mitochondrial respiration and curbs Warburg effect. This could account for the associated lower risk for KRAS MUT CRC.

4.2.5.3 rs10795668 genotype and CRC prognosis

Studies showed that expression of mitochondrial oxidative phosphorylation chain genes has prognostic value in CRC [286]. Considering the results in section 4.2.5.2, studying KRAS-mSNPs on CRC prognosis could also be informative. Therefore, I investigated the relationship between rs10795668 genotype and patient survival outcomes in the VICTOR and QUASAR2 clinical trial cohorts. In view of the differing treatment regimens in the two trials, survival analyses were conducted separately for each cohort (see Methods, section 2.3.7). Examination of variant genotype and patient survival outcomes in the COIN/COIN-B trials was not feasible due to a lack of clinical outcome data.

An important consideration is the relationship between KRAS mutational status and patient survival outcomes in CRC. Several studies showed that KRAS hotspot mutations were associated with increased risk of cancer recurrence and worse prognosis in stage II/III CRC [287–289]. Specifically, in the QUASAR2 study, KRAS mutations were associated with shorter recurrence-free survival (RFS) (HR = 1.99, 95% CI = [1.37-2.91], P = 3.44e-04, log-rank test; [174]), while in the VICTOR study they significantly predicted recurrence and lung metastases [290]. I validated these results in the TCGA pan-cancer cohort: patients with KRAS MUT tumours had shorter progression-free interval (PFI) while adjusting for gender, age at diagnosis, and tumour types in a multivariate Cox regression model (HR = 1.17 [1.01-1.36], P = 0.03, log-rank test; Fig 4.12A). There was a similar trend among TCGA CRC cases (HR = 1.36 [1.93-2.01], P = 0.12, log-rank test; Fig 4.12B).

In an unstratified analysis in the VICTOR trial cohort, the rs10795668-A allele showed no significant association with recurrence-free survival (RFS; HR = 1.11 [0.83, 1.48], P = 0.49, log-rank test; Fig 4.13A) or OS (HR = 0.95 [0.65, 1.39], P = 0.80, log-rank test; Fig 4.13C). However, following stratification by KRAS mutation status, I found that the rs10795668-A allele was marginally associated with worse RFS in cases with KRAS WT CRC (HR = 1.78 [1.01, 3.13], P = 0.045, log-rank test; Fig 4.13B), but not in cases with KRAS MUT CRC (HR = 0.81 [0.37, 1.77], P = 0.59; Fig 4.13B). While interesting, the interaction between rs10795668 genotype and KRAS mutational status in CRC tumors was not statistically significant ($P_{interaction}$

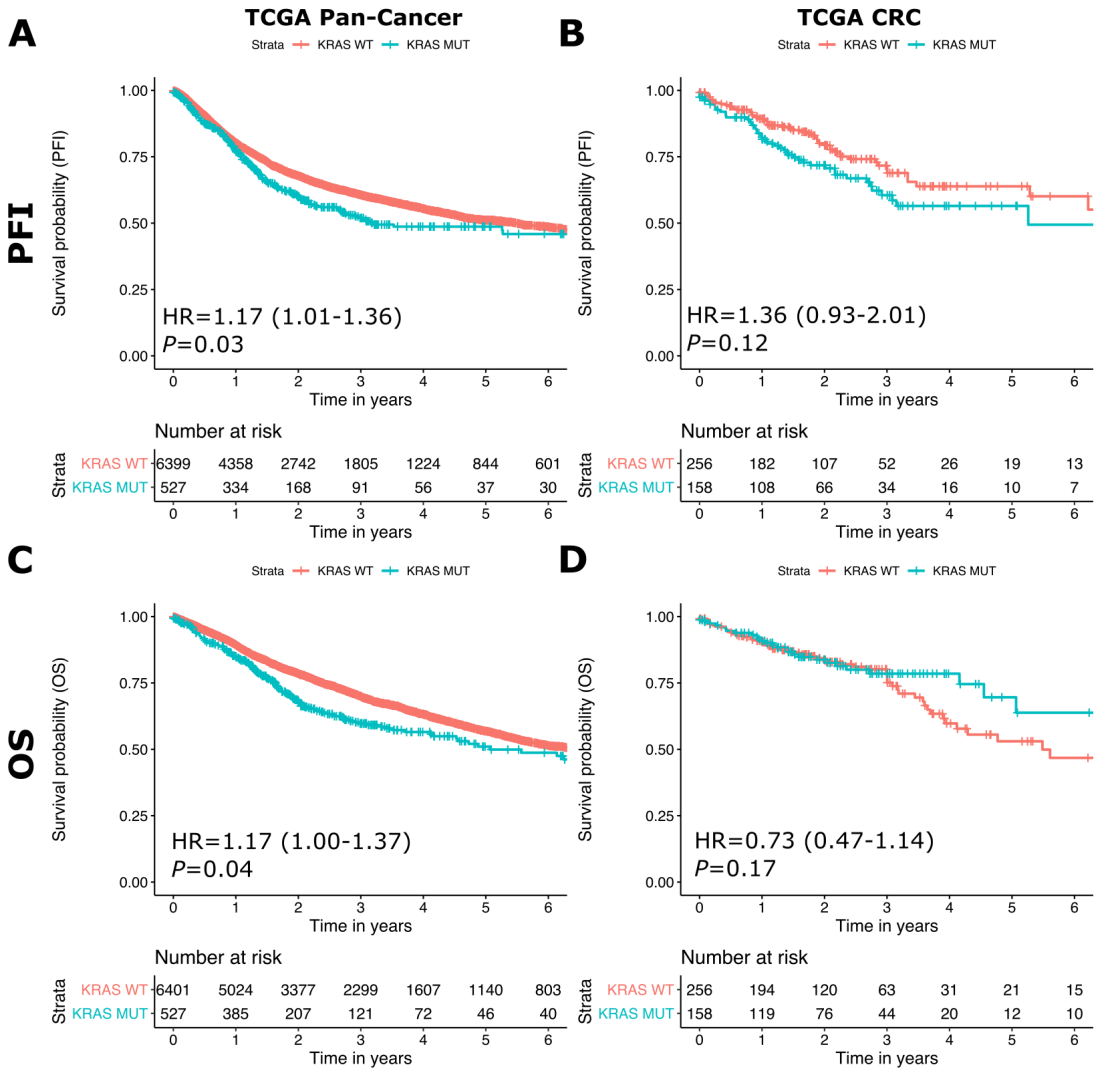

Figure 4.12. KRAS mutational status and clinical outcomes in TCGA.

(A, C) Kaplan Meier curves of KRAS mutational status against progression-free interval (PFI) (A) or overall survival (OS) (C) among TCGA pan-cancer cases. (B, D) Kaplan Meier curves of KRAS mutational status and progression-free interval (PFI) (B) or overall survival (OS) (D) among TCGA CRC cases. The hazard ratio (HR) and the 95% confidence interval were estimated from multivariate Cox regression, adjusted for age and sex. The nominal P-value was calculated by log-rank test.

= 0.11). Meanwhile, I did not detect any correlative relationship between rs10795668 genotype and OS by KRAS mutational status (Fig 4.13D).

In the QUASAR2 trial cohort, I also identified a similar trend of association between rs10795668 genotype and RFS by KRAS mutational status, although it was not statistically significant (KRAS WT: HR = 1.35 [0.90, 2.02], $P = 0.15$; KRAS MUT: HR = 1.06 [0.66, 1.71], $P = 0.81$; log-rank test; Fig 4.14B). However, this is not validated in the TCGA cohort, either among pan-cancer (Fig 7.2) or

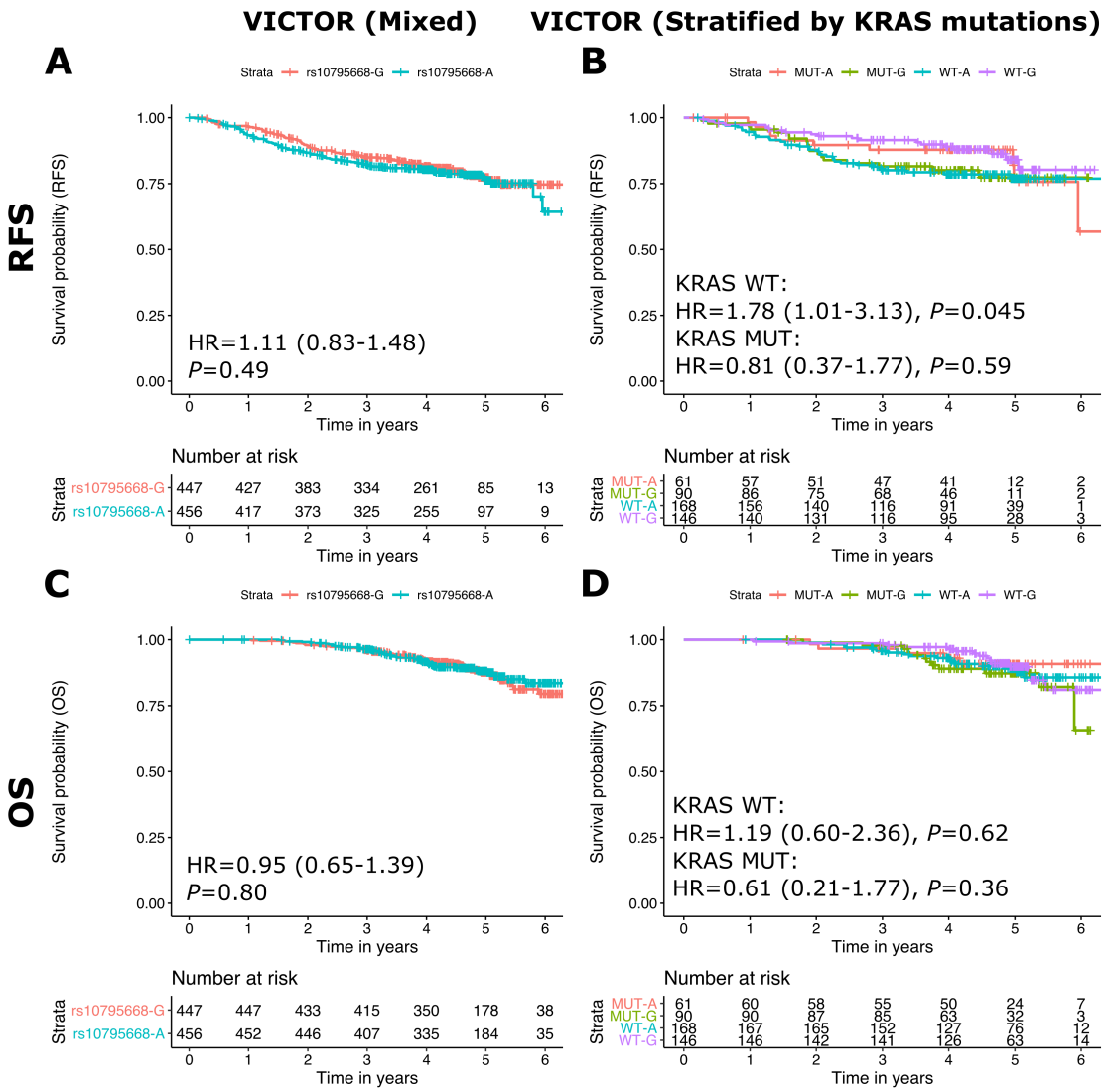

Figure 4.13. rs10795668 genotype and CRC prognosis in the VICTOR trial.

(A, C) Kaplan-Meier curves for rs10795668 genotype (assuming a dominant model for the minor allele A) against recurrence-free survival (RFS) (A) and overall survival (OS) (C), where tumours were not stratified on KRAS mutational status. Number of patients in each group and hazard ratios (HR) comparing patient survival outcome with either G or A allele are indicated. P-values were calculated by log-rank test in the multivariate COX proportional hazard model, adjusted for age at diagnosis, gender, tumour location, AJCC TNM stage, chemo-radiation status, and treatment. (B, D) Kaplan-Meier curves for rs10795668 genotype (assuming a dominant model for the minor allele A) against RFS (B) and OS (D), where tumours were stratified on KRAS mutational status. The correlative relationship between genotype and patient survival outcomes was evaluated in cases with KRAS WT or KRAS MUT separately. The hazard ratio (HR) and the 95% confidence interval were estimated from multivariate Cox regression, adjusted for age, sex, tumour location (rectum vs. colon), AJCC TNM stage (II vs. III), chemo-radiation history, and treatment (Vioxx vs. placebo). The nominal P-value was calculated by log-rank test.

among CRC tumours (Fig 7.3).

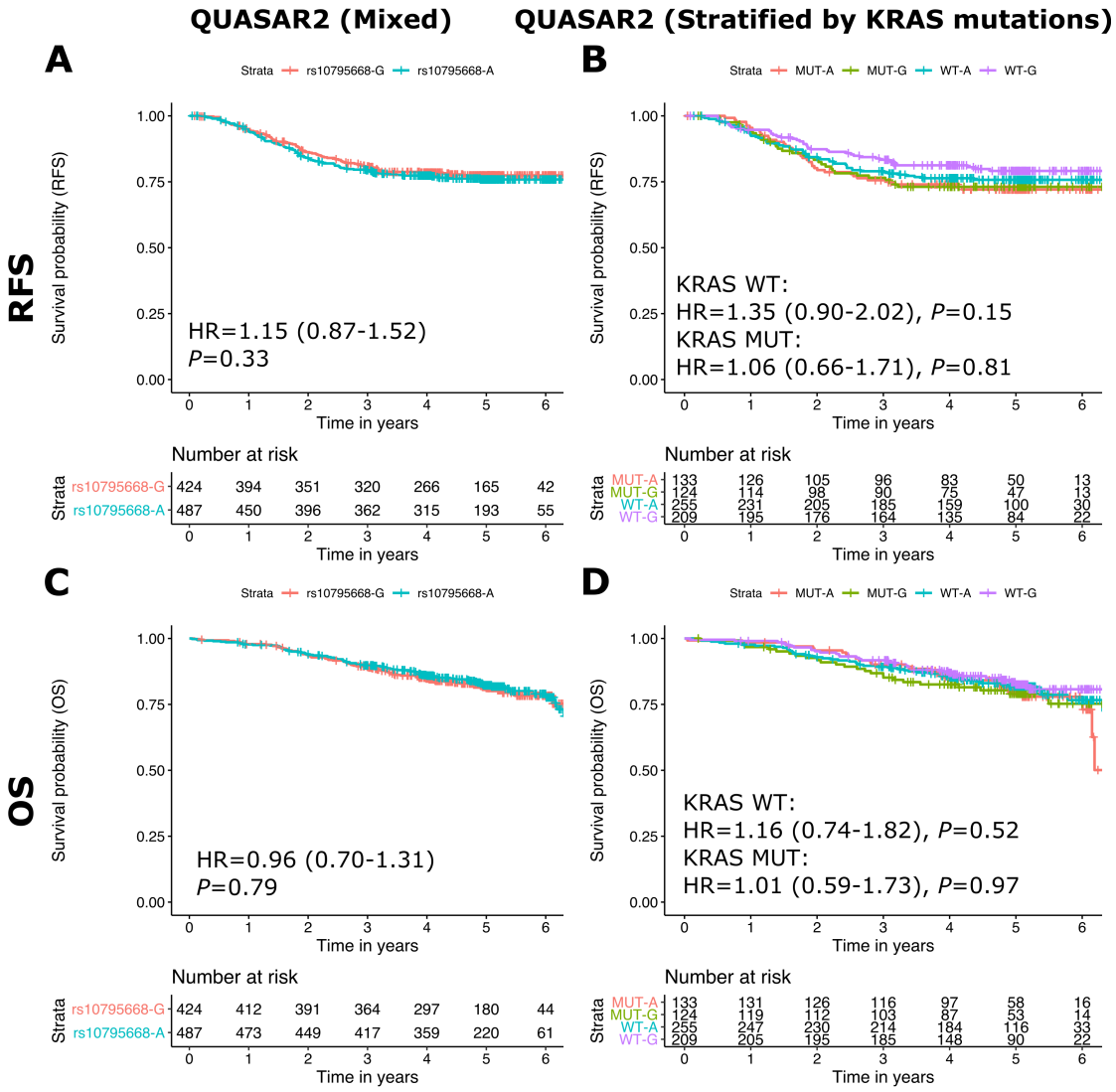

Figure 4.14. rs10795668 genotype and CRC prognosis in the QUASAR2 trial.

(A, C) Kaplan-Meier curves for rs10795668 genotype (assuming a dominant model for the minor allele A) against recurrence-free survival (RFS) (A) and overall survival (OS) (C), where tumours were not stratified on KRAS mutational status. Number of patients in each group and hazard ratios (HR) comparing patient survival outcome with either G or A allele are indicated. P-values were calculated by log-rank test in the multivariate COX proportional hazard model, adjusted for age at diagnosis, gender, tumour location, pathological tumour stage, nodal stage, and treatment. (B, D) Kaplan-Meier curves for rs10795668 genotype (assuming a dominant model for the minor allele A) against RFS (B) and OS (D), where tumours were stratified on KRAS mutational status. The correlative relationship between genotype and patient survival outcomes was evaluated in cases with KRAS WT or KRAS MUT separately. The hazard ratio (HR) and the 95% confidence interval were estimated from multivariate Cox regression, adjusted for age, sex (male vs. female), tumour location (rectum vs. colon), primary tumour stage (pT4 vs. pT1-3), nodal stage (N1/2 vs. N0), treatment (bevacizumab and capecitabine vs. capecitabine). The nominal P-value was calculated by log-rank test.

4.3 Discussion

As introduced in section 1.2.4, CRCs in distinct molecular subtypes behave very differently, both in terms of clinicopathological features and response to therapies or disease progression. Identifying inherited genetic variants that predispose risk to specific subtype of CRC or CRC with specific driver mutations would not only deepen our understanding of genetic susceptibility to CRC, but could also aid in personalised cancer risk management. In this chapter, I present evidence that a known CRC risk variant, rs10795668 at 10p14, interacts with KRAS mutational status in tumours to associate with cancer risk, nearby gene expression and potentially with prognosis in CRC.

Standard CRC GWASs have identified 175 inherited genetic variants that were associated with differential risk for developing CRC in the general population. However, the interaction between these risk variants and somatic driver mutations in CRC is not defined. Using comprehensive data from four CRC clinical trials paired with two healthy control populations (Phase I and II studies; Fig 4.4), I showed that 15 variants in a LD block at a known CRC risk locus 10p14 were associated with differential risk for developing CRC with KRAS hotspot mutations (KRAS MUT), but not for KRAS WT CRC (Fig 4.7). Multiple CRC GWASs reported that the rs10795668-A allele at this locus was associated with decreased CRC risk (OR range: [0.84-0.94]). By stratifying tumours based on their KRAS mutational status, I found that the rs10795668-A allele was exclusively associated with a decreased risk for KRAS MUT CRC, with a greater effect size (OR = 0.78); whereas it was not associated with risk for KRAS WT CRC. This indicates that stratified association study by KRAS mutational status in tumours represents a powerful platform to refined genotype-risk association in CRC by molecular subtype (see also Fig 4.3).

In contrast, a stratified association study by KRAS-BRAF mutational status in tumours did not improve the genotype-risk association by KRAS mutational status in tumours alone (Fig 4.7A vs. Fig 4.8). This suggests that genotype by KRAS mutational status interaction on CRC risk is independent of BRAF mutations in tumours, and that the risk association of rs10795668 and other variants at 10p14 should be interpreted in the context of mutated KRAS-driven hyperactivation of

the RAS-MAPK pathway in tumours.

Among the TCGA pan-cancer tumours, I observed a trend that tumour types with increasing KRAS mutation frequency tend to have lower frequency of the rs10795668-A allele (Fig 4.9C). This is in line with the findings in the stratified association study by KRAS mutational status in tumours (Fig 4.7). Nevertheless, the data did not strongly support the notion that the variant rs10795668 strictly tracks with KRAS hotspot mutations or CN gain in tumours (Fig 4.9). Therefore, investigation in additional CRC cohort would be beneficial to confirm this relationship.

This KRAS MUT-specific genotype-risk association could potentially be explained by the functional role of the variant rs10795668 on nearby gene expression, which is dependent on mutated KRAS-driven RAS-MAPK pathway activity (Fig 4.10). In KRAS MUT tumours, the rs10795668-A allele was associated with higher expression of *ATP5C1*, which encodes the gamma subunit of ATP synthase, thus bolstering mitochondrial respiration and counteracting Warburg effect. Tumours driven by mutated KRAS frequently reprogram their metabolism towards the Warburg effect. Therefore, the rs10795668-A allele plausibly protects against CRC risk by promoting *ATP5C1* expression to antagonise the activity of the RAS-MAPK pathway in KRAS MUT CRC (Fig 4.11). This can be readily tested experimentally. For instance, a pair of isogenic cancer cell model carrying either the major allele G or the minor allele A would differ in *ATP5C1* expression levels under the influence of mutated KRAS; the functional outcomes of high or low *ATP5C1* expression in the cell can be compared by the modality of cellular respiration and their potential for cellular survival and proliferation. It might also of interest to test whether mutated BRAF would have an impact on *ATP5C1* expression.

Evaluation of patient survival outcomes in the VICTOR clinical trial cohort suggested that the rs10795668-A allele was associated with poorer RFS among patients with KRAS WT CRC compared to carriers of the other allele, but the same allele showed a trend of improved clinical outcomes among patients with KRAS MUT CRC (Fig 4.13). Despite that the QUASAR2 trial data and the TCGA data did not strongly support this finding, the observed differential outcome in VIC-

TOR data indicates that the prognostic value of the variant rs10795668 might need to be interpreted in the context of whether the RAS-MAPK pathway activity has been disrupted by somatic driver mutations. Investigation in additional CRC cohort would help examine this germline by somatic interaction on CRC prognosis further. On the other hand, this interaction could help reconcile conflicting findings on the role of rs10795668 on CRC prognosis [291–294].

The findings presented in this chapter expanded the scope of germline by somatic interaction on cancer risk first studied in chapter 3, from the p53 tumour suppressor pathway to the RAS-MAPK oncogenic pathway. Together, these findings support the notion that inherited genetic variants and mutational status of cancer driver genes interact to associated with cancer risk and prognosis. In this chapter, I provided evidence that inherited risk variants can specifically associate with cancer carrying KRAS hotspot mutations (Fig 4.7A), suggesting a dependency of associated cancer risk on mutated KRAS and hyperactive RAS-MAPK pathway activity. This represents a different mode of germline by somatic interaction on cancer risk, as opposed to the inverted risk association pattern in chapter 3 (Fig 3.5A), suggesting that the form of germline by somatic interaction on cancer risk can be diverse and likely dependent on the functional role of the cancer driver genes of interest as well as the cancer hallmark pathway they are involved in.

As discussed above, this study provides a way to refine genotype-risk association by molecular subtype, thus informing the development of cancer subtype specific PRSs for personalised disease prediction. In addition, the findings imply that a specification on cancer subtypes might be beneficial in CRC screening and cancer prevention programs.

Chapter 5

The role of inherited genetic variants on prognosis and adaptive immune response in colorectal cancer

5.1 Introduction

5.1.1 Prognostic factors in colorectal cancer

As introduced in section 1.2.4, colorectal cancer (CRC) is a leading cause of cancer morbidity and mortality worldwide. Despite substantial improvement in CRC survival over the last five decades, there are still around 16,600 CRC deaths in the UK every year, accounting for 10% of all cancer deaths (Cancer Research UK, 2018). Besides ongoing efforts in improving treatment regimens for CRC patients, emphasis has also been on identifying prognostic and predictive factors that could aid in personalised risk stratification and prognosis prediction. Currently, TNM stage at diagnosis provides the most prognostic value, detailing loco-regional and distant spread [295]. Alongside age and gender effects, infiltration of immune cells in tumours is also known to influence CRC prognosis [111–114]. However, current schemes of risk stratification and prognosis prediction in CRC have limited efficacy.

Especially, many patients with early-stage (i.e., non-metastatic) CRC are either over-treated or under-treated [296, 297]. Therefore, additional prognostic biomarkers for distinguishing high-risk individuals from those with low risks would benefit clinical management of early-stage CRC.

Inherited genetic factors are known to play a critical role in CRC risk and pathophysiology (see also sections 1.1.3 and 4.1.1; [4, 43, 44]). Moreover, familial concordance for CRC prognosis has been reported, suggesting that inherited components can also have an impact on CRC disease progression [298]. Several studies of modest sample sizes suggested that common inherited genetic variants may predict clinical outcomes among patients with early-stage CRC [292, 299–301]. However, the prognostic value of these variants was not confirmed in larger cohorts, notably the SOCCS cohort [302–309]. As a result, the role of inherited genetics on prognosis in early-stage CRC remains controversial. I reason that this controversy may have risen from under-powered studies that reported false positive findings.

To systematically evaluate the role of inherited genetic variants on prognosis in early-stage CRC, in this chapter, I present a large-scale study on germline determinants of patient survival outcomes, utilising 3,858 cases from the QUASAR2 and SCOT clinical trials. This represents one of the most powered studies for early-stage CRC, comparable to those based on the SOCCS (n = 3,886; [305]) and NCCTG N0147 Alliance (n = 4,319; [309]) trial cohorts.

5.1.2 Adaptive immune response in colorectal cancer

The complex ecosystem of a tumour contains a mixture of diverse cell populations, including neoplastic cells, extracellular matrix, as well as auxiliary non-neoplastic cells, such as resident mesenchymal cells, endothelial cells, and infiltrating immune cells. Interaction and crosstalk between cancer cells and auxiliary cells in the tumour microenvironment (TME) shape the course of tumour development [58, 310]. Importantly, innate and adaptive immune responses play an important role in tumour development by selecting immune evasive clones, inducing immunosuppression, and facilitating metastasis [109, 110]. In particular, adaptive immune response in the TME has been suggested to affect tumour heterogeneity (i.e., cancer cell composi-

tion), to influence disease progression, and to modulate response to immunotherapy [311–313].

As discussed in section 1.2.4 and above, the strength of adaptive immune response in the TME is a key prognostic factor in CRC: a high level of infiltrating lymphocytes in tumours is associated with favourable outcomes [311, 314]. In particular, immunological profile or 'immunoscore' (i.e., a summary of the types, density, and location of immune cells) of tumour samples predicted patient survival in multiple studies [111–114].

Among tumour infiltrating immune cells, CD8+ T cells play a critical role in the anti-tumour response: cytotoxic T cells recognise tumour-specific antigens (neoantigens; discussed further below, see section 5.1.3) expressed on cancer cells and target them for cytolysis [315]. A number of studies have evaluated the prognostic role of the presence or the abundance of CD8+ T cells in tumours, often distinguishing between the tumour centre and invasive margin (Table 5.1). Overall, increasing tumour infiltration of CD8+ T cells is associated with improved clinical outcomes in CRC [316].

As introduced in section 1.3.4, adaptive immune response in the TME can be affected by a number of factors that are tumour-intrinsic or tumour-extrinsic [110]. Many of the tumour-intrinsic factors have been studied extensively. These include clonal heterogeneity, total mutation load, neoantigen load, copy number variations, gene- or pathway-level somatic mutations in the tumour [332, 333]. In comparison, the role of inherited genetic variants in anti-tumour adaptive immune response is less studied. Evidence is emerging that inherited genetic variants play a role in the interaction between the tumour and the host immune system [151–155, 158].

In this chapter, I aim to identify inherited genetic variants that predict tumour infiltrating CD8+ T cells in CRC, and investigate the relationship between these variants, adaptive immune response and prognosis in early-stage CRC.

Year	Samples Size	TNM Stage	Counting Site	Statistical Regime§	Outcomes	Ref.
1998	131	I-IV	Tumour Centre	Categorical	OS (HR=0.61 [0.41, 0.89])	[317]
2001	245	II/III	Tumour Centre	Cut-point	DFS (HR=0.35 [0.16, 0.78]) OS (HR=0.33 [0.15, 0.73])	[318]
2004	272	I-IV	Tumour Centre	Cut-point	CSS (HR=0.71 [0.48, 1])	[319]
2004	152	III	Tumour Centre	Cut-point	OS (HR=0.43 [0.22, 0.85])	[320]
2010	215	I-IV	Tumour Centre	Categorical	DFS (HR=0.68 [0.16, 2.88])	[321]
2010	768	I-IV	Tumour Centre	Categorical	OS (n.s.) CSS (n.s.)	[322]
2012	130	I-III	General	Cut-point	CSS (HR=0.51 [0.26, 0.98])	[323]
2012	216	II/III	Tumour Centre	Cut-point	OS (HR=0.68 [0.43, 1.08])	[324]
2014	426	I-IV	Tumour Stroma	Categorical	CSS (HR=0.71 [0.55, 0.92])	[325]
2014	365	I-III	Tumour Centre	Categorical	CSS (HR=0.58 [0.44, 0.77])	[326]
2015	157	I-III	Tumour Centre	Categorical	DFS (HR=0.23 [0.07, 0.67])	[327]
2015	190	II-IV	Tumour Centre	Cut-point	OS (HR=1.00 [0.99, 1.01])	[328]
2016	300	II-IV	Tumour Centre	Cut-point	DFS (HR=0.59 [0.39, 0.90]) OS (HR=0.67 [0.42, 1.07])	[329]
2017	557	I-IV	Tumour Centre	Categorical	OS (HR=0.53 [0.29, 0.95])	[330]
2018	573	II	Invasive Margin	Cut-point	RFS (HR=0.72 [0.53, 0.97) OS (HR=0.63 [0.46, 0.86])	[331]
2019	1,804	II/III	Tumour Centre	Continuous	TTR (HR=0.92 [0.87,0.97]) OS (HR=0.93 [0.87, 0.99])	[194]

Table 5.1. A review on the prognostic value of tumour infiltrating CD8+ T cells in CRC.

§ Cut-point regime: Mean cell count was analysed dividing the distributions of CD8+ T cells into two groups of equal size, using the median value as cut-point. CSS: cancer-specific survival; DFS: disease-free survival; OS: overall survival; RFS: recurrence-free survival; TTR: time to recurrence.

5.1.3 Human leukocyte antigen variations and adaptive immune response

T cells can recognise fragmented peptides (antigens) displayed on the cell surface. This is a critical process in immune response to infected or transformed cells [109].

In tumours, somatic mutations can give rise to variant amino acid sequences (i.e., neoantigens) that are presented on the surface of cancer cells and elicit adaptive immune response against them [334]. Neoantigen load in tumours partly determines the magnitude of anti-tumour adaptive immune response [335], and has been correlated with clinical efficacy of immunotherapy in CRC [336, 337]. Also, inherited genetic variants that modify cellular antigen presentation play an essential role in tumour development (see also section 1.3.4; [151, 152]).

Antigens are presented by the major histocompatibility complex (MHC; also called human leukocyte antigen, or HLA, in humans) molecules. While previous GWASs frequently examined the association between autosomal inherited genetic variants and cancer-related traits, variants in the HLA region were usually excluded due to extreme sequence variation and widespread linkage disequilibrium (LD) within the region [338]. As a result, the role of inherited genetic variants in the HLA region in cancer is largely unclear, especially in cancer prognosis. Therefore, in this chapter, I also aim to explore the prognostic value of HLA variants in early-stage CRC.

The classical MHC region on chromosome 6p21.3 spans approximately 4 Mb and comprises over 160 protein-coding genes [339]. MHC has evolved over many millions of years to become a master coordinator of specificity in both adaptive and innate immune responses [340]. In humans, HLA genes encode cell surface antigen-presenting proteins (Fig. 5.1A; [339]). The HLA region is known for its link to susceptibility or resistance to a number of infections, autoimmune and other pathologies [142–150].

HLA class I molecules (HLA-A, -B, and -C) are expressed by all nucleated cells; they present peptides of cytosolic and nuclear origin at the cell surface to CD8+ cytotoxic T cells [341, 342]. While HLA class I molecules are ubiquitously expressed, the class II molecules (HLA-DR, HLA-DQ and HLA-DP) are primarily expressed by professional antigen presenting cells (e.g., dendritic cells, macrophages, and B cells); they present foreign peptides to CD4+ helper T cells [341, 342]. Typically, the peptides presented by HLA class I and II molecules differ in their origin (i.e., endogenous for class I molecules and exogenous for class II molecules) and are derived

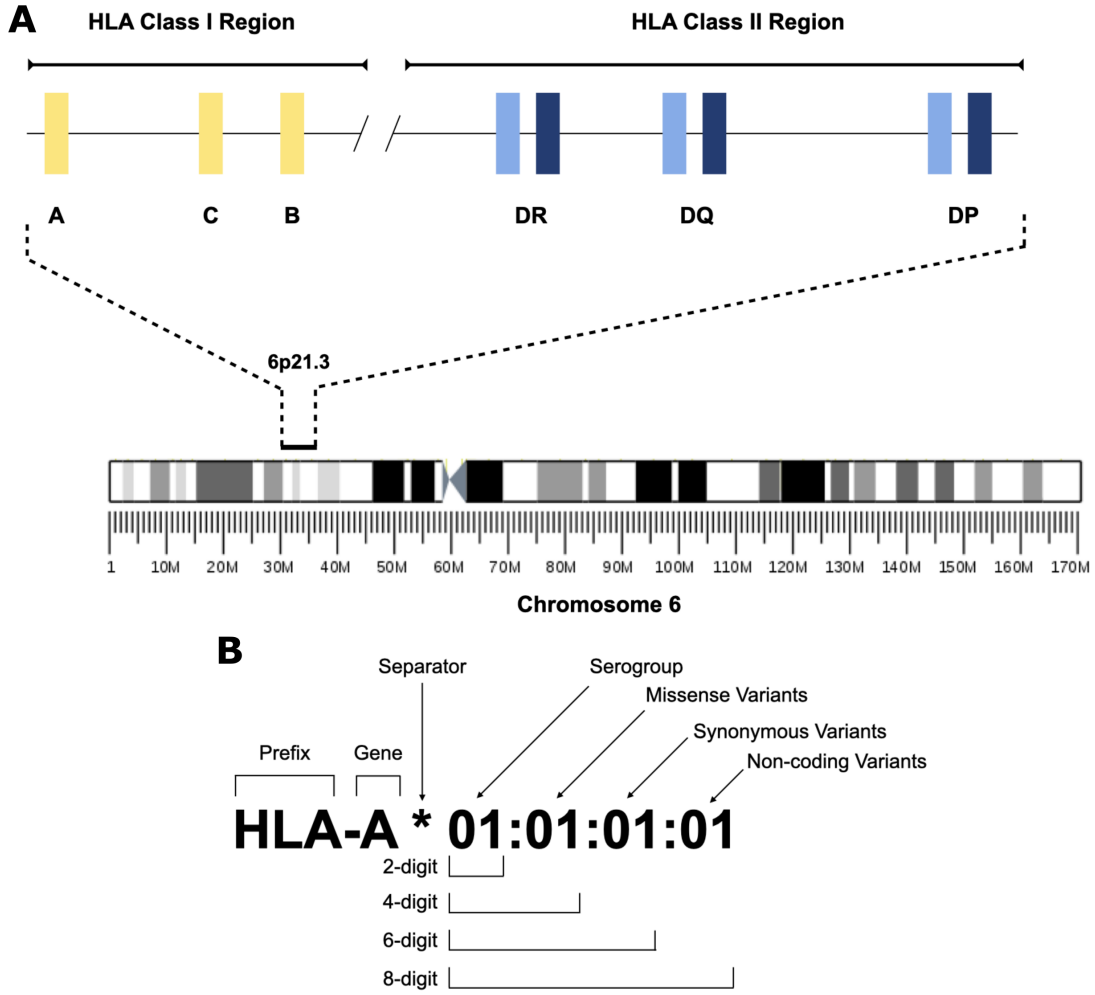

Figure 5.1. Illustration of HLA gene locus and nomenclature.

(A) The location of the HLA class I and class II genes on chromosome 6. (B) Example of an HLA type (allele) using the standard nomenclature. Typically, the first two digits refer to the serological antigen; the third and fourth digits denote HLA types that differ by one or more missense variants; the fifth and sixth digits specify HLA types that differ by one or more synonymous variants; the seventh and eighth digits distinguish HLA types that differ by non-coding variants.

from different cellular processes. Nevertheless, cross-presentation exists between the two pathways of antigen presentation: exogenous antigens are presented by HLA class I molecules [343] and cytosolic proteins can be presented by HLA class II molecules when they are degraded through autophagy or other pathways [344]. Therefore, both class I and II molecules could play a part in presenting neoantigens on the surface of cancer cells.

As mentioned above, genes in the HLA region are highly polymorphic. These polymorphisms result in variation in the anchor residues to which peptides dock,

thus creating distinct peptide-binding grooves that present unique antigens and enabling presentation of diverse peptide sequences [345]. Up to date, over 8000 unique alleles have been documented for the classical HLA genes, and more than 2600 haplotypes have been reported for the *HLA-B* gene alone [346]. Within the HLA coding regions, some codons contain multiple polymorphisms (sometimes with multiple alleles), creating a range of possible amino acids at the corresponding residues in the HLA molecule. Reflecting the original antibody-based serotyping used to study HLA molecular isoforms, diverse HLA alleles are usually classified into 'types' (Fig. 5.1B). Currently, the four-digit typing (e.g., HLA-A*01:01), referring to a unique amino acid sequence of the gene product (thus ignoring synonymous changes), is the standard method to represent HLA genotypes [347].

The interaction between tumour neoantigens and the host immune response is dynamic [332, 348]. This process actively selects the cancer cell populations for immune evasive phenotypes (i.e., immunoediting), as commonly seen in cases with mutations and loss of heterozygosity in HLA genes [349–351]. Moreover, HLA genotype-specific neoantigen presentation restricts immunoediting in cancer [151, 152]; HLA class I genotypes influence cancer response to immunotherapy and disease progression in advanced melanomas [352]. Therefore, an assessment of the relationship between inherited HLA variants and clinical outcomes in early-stage CRC could potentially be of additional benefit to risk stratification and prognosis prediction for patients with early-stage CRC.

5.1.4 Inherited genetic variants and immune quantitative traits in cancer

A number of inherited genetic variants have been associated with immune traits such as immune cell levels in healthy and diseased tissues [156, 157, 353], as well as susceptibility to immune-mediated diseases [354, 355]. In cancer, evidence is emerging that the extent of anti-tumour immune response is influenced not only by somatic driver mutations in *KRAS, BRAF* and *TP53* [332, 333], but also by common inherited genetic variants [153–155, 333]. In turn, inherited genetic variants that modify immune traits potentially have an impact on clinical outcomes [333, 356].

So far, four studies have performed systematic search for common inherited genetic variants that are associated with immune gene expression or immune composition of the TME based on the TCGA pan-cancer cohort data. Thorsson and colleagues found that a *cis*-eQTL (rs822337) 1 Kb upstream of *CD274* was weakly correlated with PD-L1 expression in tumours [333]. Lim and colleagues identified 103 'gene signature QTLs' that were associated with predicted immune cell abundance in the TME [153], but no variant was associated with predicted tumour infiltrating CD8+ T cells abundance. Shahamatdar and colleagues revealed two inherited genetic variants were associated with tumour infiltrating T cell phenotypes: rs3366 in the 3'-UTR of *SIK1* was associated with follicular helper T cell abundance in the tumour bulk; rs4819959, a *cis*-eQTL of *IL17RA*, was associated with T helper 17 cell signature, predicted by the expression of three genes, including *IL17RA* [154]. Sayaman and colleagues determined that common inherited genetic variants explain 20% of the variation in CD8+ T cell subset enrichment in tumours, and they identified half a dozen variants associated with the phenotype [155]. Note that all four studies relied heavily on *in silico* prediction of immune cell abundance in tumours based on gene expression profiles, which is not without bias. In addition, the TCGA pan-cancer cohort is an admixed population cohort, and it lacks high-quality patient follow-up data.

With regards to data quality, clinical trial sample cohorts represent the ideal for prognostic studies. However, it is not a standard practice to obtain germline data of trial participants; most CRC clinical trials do not have information about patients' inherited genomes. Therefore, it will be valuable to have a CRC clinical trial cohort ascertained of both genotypic and immune phenotypic data to facilitate the identification of immune quantitative trait loci (iQTLs) in tumours. In this chapter, I leveraged the QUASAR2 clinical trial cohort to map iQTLs in tumours, based on tumour infiltrating immune cell levels determined by *in situ* immunohistochemistry (IHC) staining.

5.1.5 Establishing a causal relationship between adaptive immune response and CRC prognosis

As mentioned in section 5.1.2, the abundance of tumour infiltrating immune cells can predict disease progression, response to therapy, and clinical outcomes. However, potential confounding factors make it challenging to draw a causal link from correlation identified in observational studies (Fig 5.2A), especially when the confounding factors are unknown or inadequately measured [357]. While multiple studies have shown an association between the levels of tumour infiltrating CD8+ T cells in tumours and CRC prognosis (Table 5.1), their causal relationship remains to be delineated. In this chapter, I will perform causal inference to address this question [358, 359].

Among the established schemes of causal discovery such as Granger causality, Bayesian networks, and structural causal models (reviewed in [361]), Mendelian randomisation (MR) is arguably the most widely used in epidemiological studies [362]. MR is a study design in which genetic variants are employed as instrumental variables for estimating the unconfounded effect of an exposure (e.g., infiltrating immune cell abundance) on a disease outcome (e.g., cancer prognosis) (Fig 5.2B). Although common genetic variants typically have small impact on complex disease phenotypes, combining multiple variants as a genetic instrument increases the statistical power to detect associations between exposure and outcome [363, 364]. Importantly, because MR leverages the random assortment of alleles at meiosis, the estimated relationships are less vulnerable to confounding than observational studies. When individual-level data is available, a two-stage Mendelian randomisation can be used to interrogate the causal relationship of interest, within which the Cox regression framework can be retained to accommodate for time-to-event data (Fig5.2C; [196]).

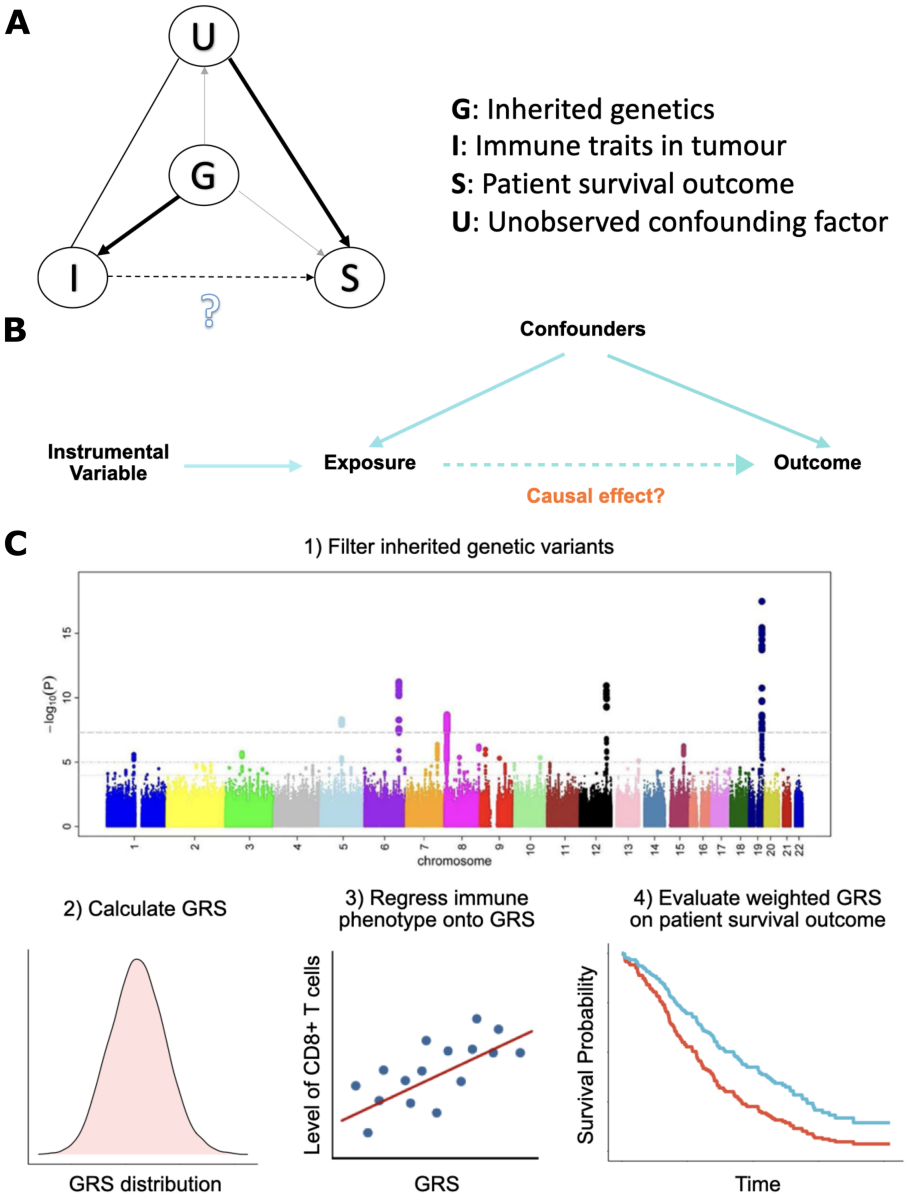

Figure 5.2. Using Mendelian randomisation (MR) to infer the potential causal relationship between adaptive immune response in tumours and CRC prognosis.

(A) Possible causal relationships between inherited genetic variants, immune quantitative traits in tumours, confounding factor, and patient survival outcome. (B) A schematic of the conceptional framework of MR, adapted from [360]. The instrumental variable is based on inherited genetic variants associated with immune quantitative trait of interest in tumours, and the outcome refers to clinical outcomes among patients with early-stage CRC. The effect of the instrumental variable should be independent from the confounding factors and should affect outcome only through exposure. In the presence of a causal relationship, the association between instrumental variable and CRC prognosis should be proportionate to its association with the immune quantitative trait, given the relationship between the exposure and outcome. (C) Workflow for the two-stage MR employed in this study.

5.2 Results

5.2.1 Survival analyses in the QUASAR2 and SCOT clinical trial cohorts

To systematically evaluate the prognostic value of common inherited genetic variants in early-stage CRC, I proposed to perform genome-wide survival analyses in the QUASAR2 and SCOT clinical trial cohorts independently, followed by a meta-analysis. A total of 3,858 CRC cases were included in this study and their basic characteristics are summarised in Table 5.2. Note that all patients have early-stage CRC (AJCC TNM stage: II/III), who underwent adjuvant chemotherapy (see Methods, section 2.1.3).

Following established quality control measures for each dataset, the genotypes of over 8 million SNPs in each study were imputed, using 1000 Genomes [2] and UK10K [172] data as reference (see Methods, section 2.1.3.1). After filtering SNPs with a minor allele frequency (MAF) < 0.01 and imputation quality score < 0.8, a total of 8,307,314 genetic variants were included in this study. Power calculation indicated a power of at least 0.89 to detect a hazard ratio (HR) of 1.5 for variants with MAF > 0.05, and a power of 0.79 for variants with MAF > 0.025. Statistical power estimation with various parameter settings is presented in Fig 5.3.

The primary endpoint of this study was time to CRC recurrence (TTR; defined as the time from randomisation to colorectal cancer relapse, with censoring at last contact or death in case of no recurrence), and the secondary endpoint was overall survival (OS; defined as all-cause death, with censoring at last contact). First, I examined the effect of clinicopathological factors that are known to influence patient survival outcomes in the two cohorts using a multivariate Cox proportional hazard regression model (Fig 5.4), testing the association of age, sex, tumour location (colon or rectum), pathological tumour (pT) and nodal (N) stage, treatment and randomisation arm (for SCOT trial only) with clinical outcomes (see Methods, section 2.3.8). I observed that pT stage and N stage were strongly associated with both TTR and OS (Fig 5.4), consistent with previous reports [174, 306]. Note that mutational status in cancer driver genes such as *KRAS* and *BRAF* were not

	QUASAR2	SCOT
Number of informative patients	929	2,929
Number of events		
CRC recurrence	205	538
All-cause death	165	186
Age		
Mean (SD)	63 (9.7)	64 (9.0)
< 20	0	0
20-49	81	206
50-59	200	581
60-69	378	1,364
70-79	247	741
80-89	23	1,134
Sex		
Male	534	1,1795
Female	395	1,134
AJCC TNM Stage		
I	0	0
II	346	586
III	583	2,343
IV	0	0

Table 5.2. Basic characteristics of QUASAR2 and SCOT clinical trial cohorts.

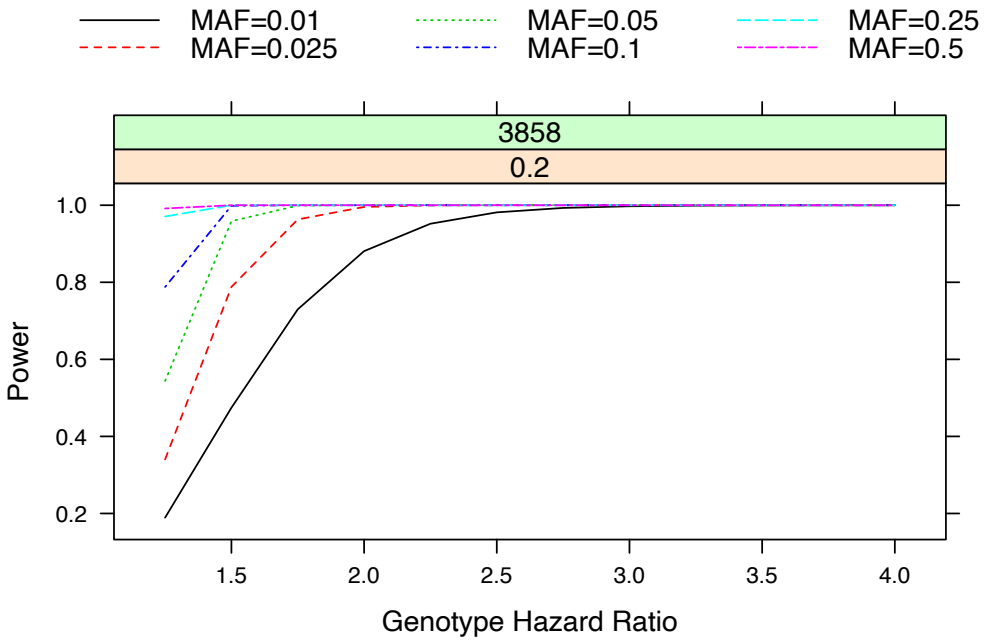

Figure 5.3. Calculated statistical power of the current study for genetic variants with varying minor allele frequency (MAF) and genotype hazard ratios. The green box indicates the total number of samples included in this study; the yellow box indicates the event rate.

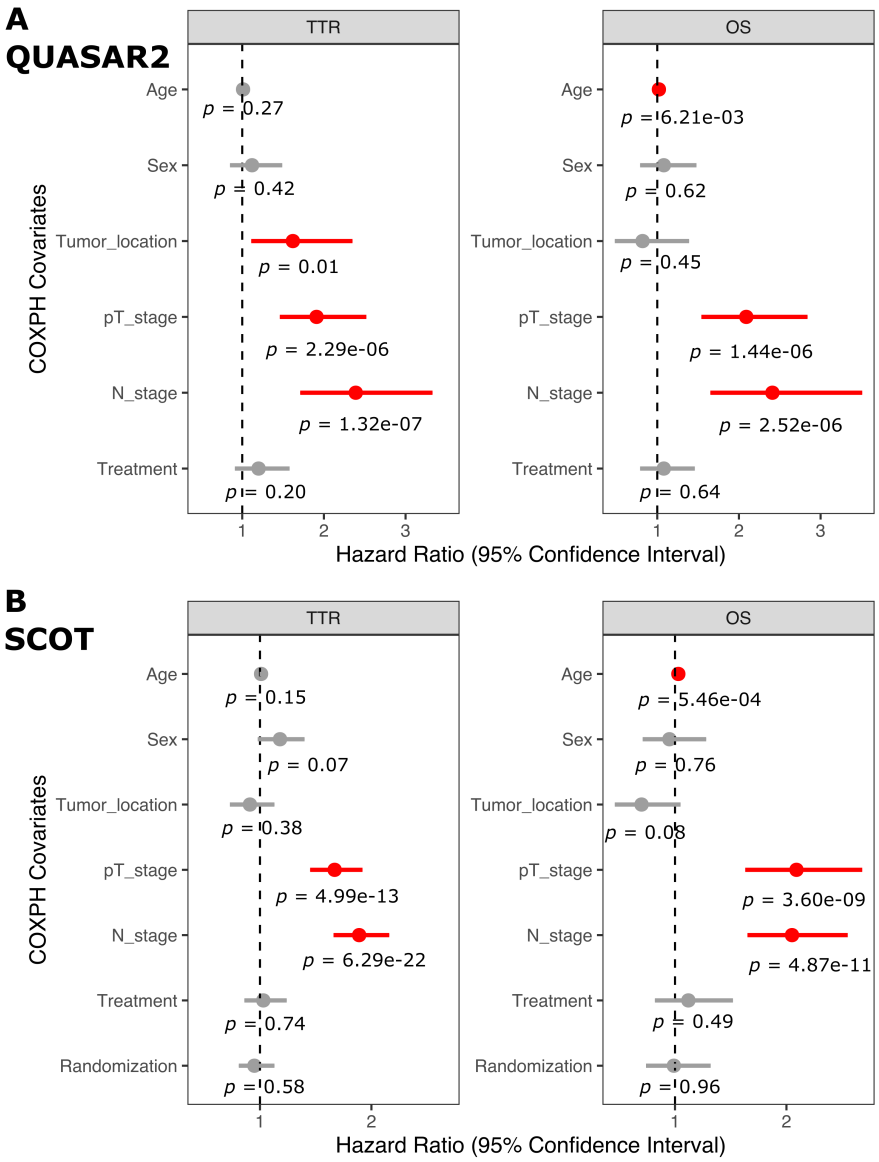

Figure 5.4. Associations of known clinicopathological factors on patient survival outcomes in the sample cohorts.

(A) QUASAR2: age as a continuous variable; sex as a categorical variable (male vs. female); tumour location as a categorical variable (rectum vs. colon); primary tumour stage as a categorical variable (pT4 vs. pT1-3); nodal stage as a categorical stage (N1/2 vs. N0); treatment as a categorical variable (bevacizumab and capecitabine vs. capecitabine). (B) SCOT: age as a continuous variable; sex as a categorical variable (male vs. female); tumour location as a categorical variable (rectum vs. colon); primary tumour stage as a categorical variable (pT1-2 vs. pT3 vs. pT4); nodal stage as a categorical stage (N0 vs. N1 vs. N2); treatment as a categorical variable (FOLFOX vs. CAPOX); randomisation arm as a categorical variable (24 vs. 12 weeks). Hazard ratio (HR) estimates of each variable on clinical outcomes (TTR or OS) were calculated using a univariate Cox proportional hazard regression model, and the corresponding statistical significance (nominal P-values) were evaluated using log-rank tests.

included here due to a lack of molecular data in the SCOT clinical trial cohort. Subsequently, prespecifying the listed factors as covariates, I assessed the relationship

between individual variant genotype and patient survival outcomes using a multivariate Cox regression model, assuming an additive or dominant genetic effect (see Methods, section 2.3.8). Next, hazard ratio (HR) estimates for each genetic variant in the two cohorts were combined through an inverse-variance weighed fixed-effect meta-analysis (see Methods, section 2.3.8).

5.2.2 Evaluating the prognostic value of CRC risk loci

Previous GWASs have identified a number of variants associated with CRC risk. These genetic variants may also influence survival outcomes of patients with early-stage CRC. In this section, I evaluate their potential prognostic value in the QUASAR2 and SCOT clinical trial cohorts.

To date, 175 loci have been associated with CRC risk (see Methods, section 2.2.2; see also section 4.2.3). I examined the relationship between variant genotype and patient survival outcomes in a multivariate Cox regression model adjusted for confounders, as mentioned in section 5.2.1. I found that 16 loci were associated with TTR or OS with a nominal P-value below 0.05 in the meta-analysis. The most significant results include an intronic variant in *GNA12* (rs1182197, 7p22.2, MAF = 0.33), the minor allele of which was associated with longer TTR (HR = 0.86 [0.77-0.95], P = 4.02e-03); and an intronic variant in *TMEM91* (rs9797885, 19q13.2, MAF = 0.28), the minor allele of which was associated with improved OS (HR = 0.77 [0.65-0.92], P = 3.98e-03). However, none of the associations were statistically significant after adjusting for multiple hypothesis testing (FDR > 0.05; Supplementary Data Table 5.1).

I also checked 35 genetic variants that were previously associated with CRC clinical outcomes in cohorts of mixed TNM stages and of varying samples sizes (range: [505-3,494]). One variant (rs209489) has been associated with both disease-free survival and OS at genome-wide significance in a cohort of 3,494 patients with metastatic CRC [304]; while the rest were associated with patient survival outcomes at suggestive significance. However, these associations were not validated in the QUASAR2 and SCOT cohorts (Supplementary Data Table 5.2).

Previously, a list of 50 variants that were correlated with immune-related gene

expressions were shown to associate with prognosis in cutaneous melanoma [356]. I thus evaluated the association of these 50 immune gene eQTLs with survival outcomes in the QUASAR2 and SCOT clinical trial cohorts. Two variants were marginally associated with TTR in the meta-analysis with a nominal P-value below 0.05: rs8101605 (eQTL for *LILRB1*; HR = 1.15 [1.01-1.32], $P = 0.041$) and rs2517681 (eQTL for *HLA-G*; HR = 0.89 [0.80-0.99], $P = 0.030$). However, these associations were not significant after adjustment (FDR = 1; Supplementary Data Table 5.3).

The findings presented above are consistent with those reported of the SOCCS clinical trial, where a total of 5,675 CRC cases of all TNM stages (I-IV) were included [307, 308]: neither known CRC risk variants nor prognostic variants reported for multiple cancer types were associated with CRC patient survival outcomes. This suggests that the prognostic value of known CRC risk loci in early-stage CRC can be ruled out, and that previously reported CRC prognostic variants are either stage-specific or likely false-positive findings due to small sample sizes.

5.2.3 A genome-wide search for prognostic variants in early-stage CRC

In this section, I present the findings from genome-wide survival analyses in the QUASAR2 and SCOT clinical trial cohorts (see section 5.2.1). Within individual cohorts, I identified a total of four loci associated with TTR at genome-wide significance (Fig 5.5A, B). However, none of the loci overlapped between the two cohorts. Similarly, six loci were significantly associated with OS, but only within individual cohorts (Fig 5.5C, D). Meta-analysis revealed a number of loci that were significantly associated with both TTR and OS in early-stage CRC (Fig 5.6). In the following subsections, I am going to present the top candidate prognostic variants identified in this study. A full list of candidate loci reaching suggestive significance in the meta-analysis are provided in the Supplementary Data Table 5.4 and 5.5.

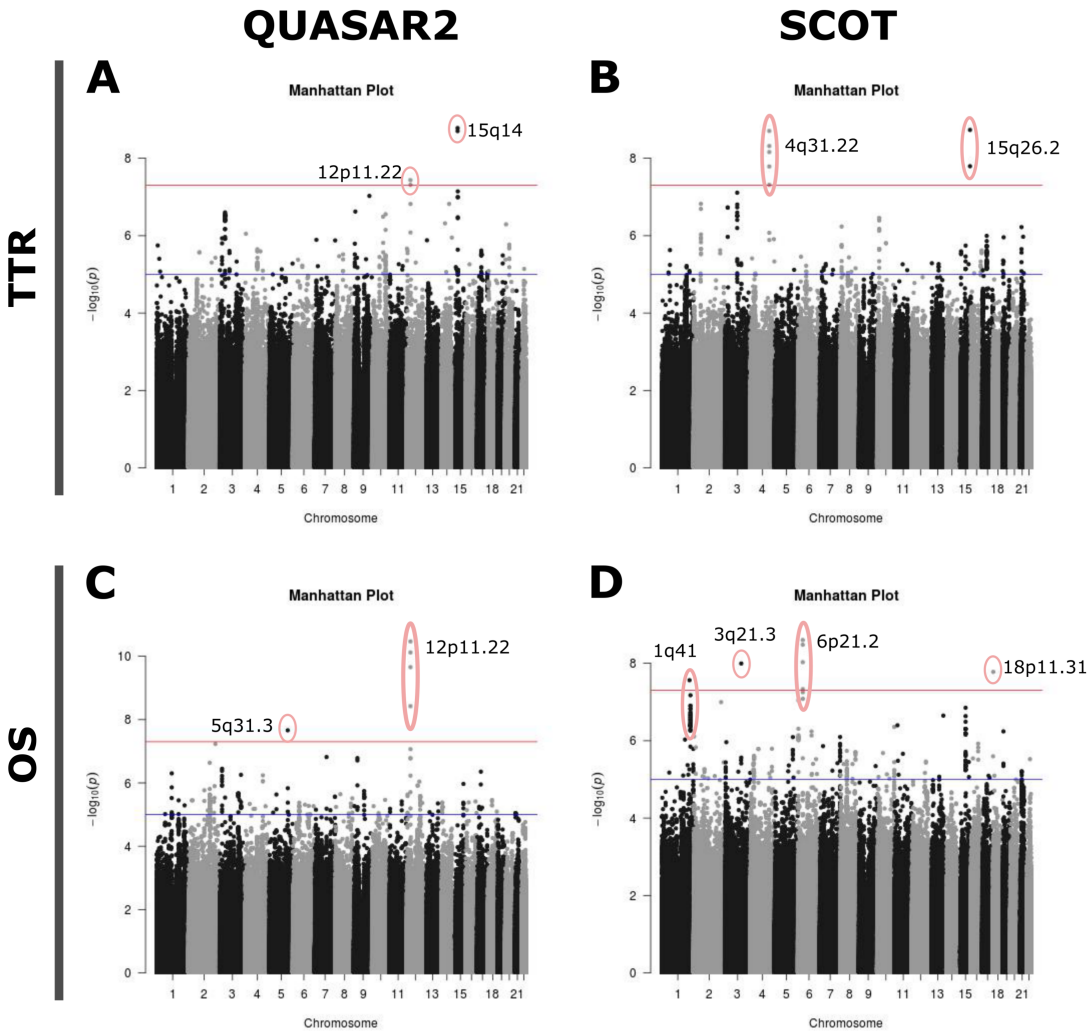

Figure 5.5. Manhattan plots of survival GWASs in respective colorectal cancer clinical trial cohorts: (A, C) QUASAR2 and (B, D) SCOT.

Inherited genetic variants were tested for their effects on patient survival outcomes independently, while accounting for known risk factors in multivariate Cox proportional hazard regression models. Variants of genome-wide significance ($P < 5.0e\text{-}08$) are highlighted and their cytobands annotated. Clusters of variants in high linkage disequilibrium (LD; $R^2 > 0.8$) are noted in (C, D).

5.2.3.1 A locus at 15q26.2 is associated with TTR in early-stage CRC

Meta-analysis revealed a locus at chromosome 15q26.2 (rs72767781 and rs72767784; $R^2 = 0.65$, $D' = 1$) that was consistently associated with TTR in both cohorts (Fig 5.6A). In particular, the minor allele of rs72767781 associated with shorter TTR in patients with early-stage CRC (QUASAR2: HR = 1.71, 95% CI: [1.01-2.89], $P = 0.046$; SCOT: HR = 2.26 [1.71-3.01], $P = 1.60e\text{-}08$; meta-analysis: HR = 2.12 [1.66-2.73], $P = 3.21e\text{-}09$) (Fig 5.7). This variant is of low frequency in the European population, with a MAF of 0.03 in both QUASAR2 and SCOT clinical trials. As

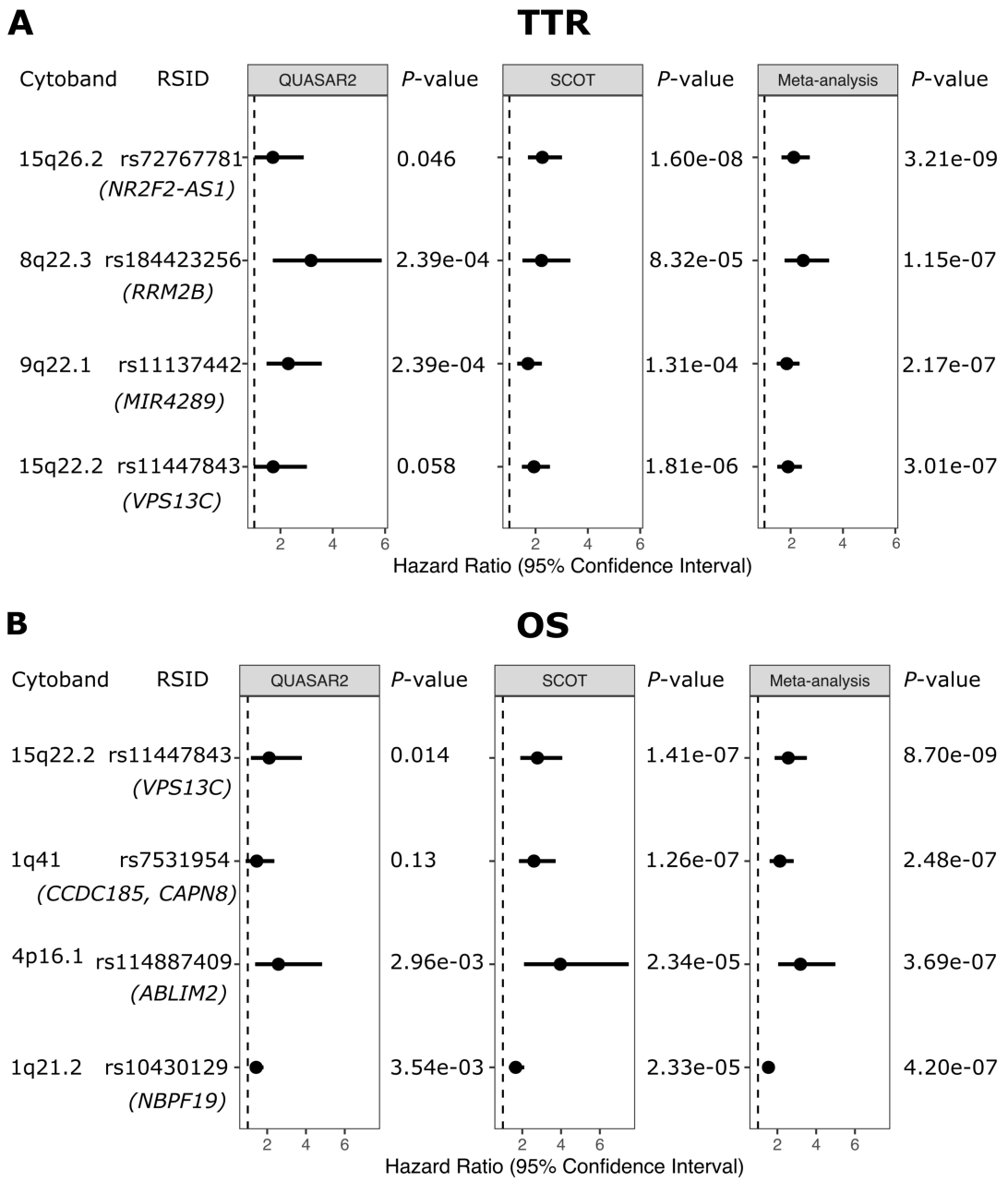

Figure 5.6. Meta-analysis reveals candidate prognostic variants for early-stage CRC.

Forest plots showing the association of top candidate prognostic variants with TTR (A) or OS (B) in early-stage CRC.

discussed in section 5.2.1, this study had sufficient statistical power for discovery.

Interestingly, an exploratory subgroup analysis revealed that the minor allele of variant rs72767781 was associated with substantially shorter TTR among patients with high-risk stage II CRC (QUASAR2: HR = 3.15 [1.22-8.11], $P = 0.018$; SCOT: HR = 3.13 [1.32-7.38], $P = 9.34\text{e-}03$; Fig 7.4A, B), compared to those with stage II or

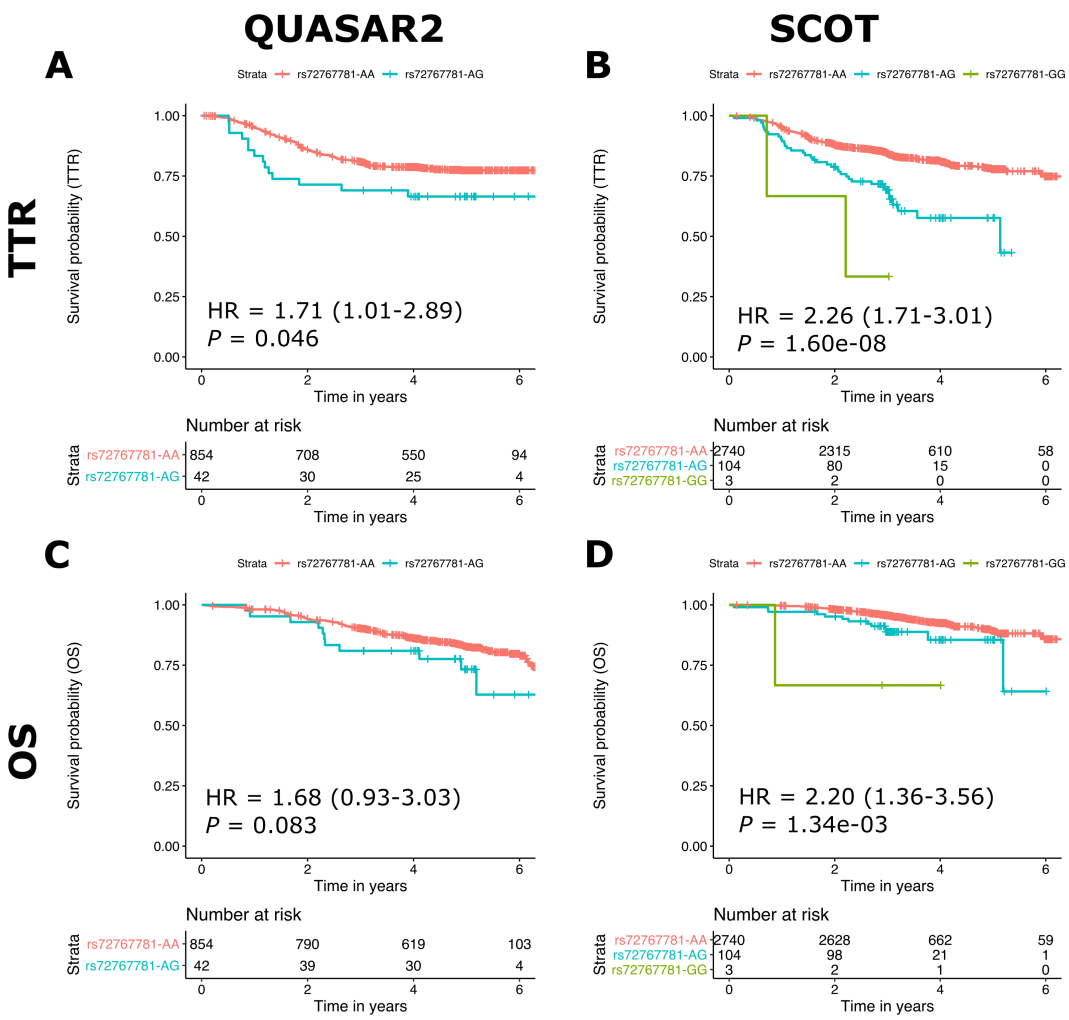

Figure 5.7. rs72767781 (15q26.2, MAF = 0.03) is associated with patient survival outcomes in QUASAR2 (A, C) and SCOT (B, D).

(A, B) For TTR. (C, D) For OS. The hazard ratio (HR) and the 95% CI were estimated from multivariate Cox regression, adjusted for known confounders; the nominal P-value was calculated by log-rank test (see Methods, section 2.3.8).

III CRC (Fig 5.7A, B). This result suggests that the association of rs72767781 with clinical outcomes in CRC could be specific to stage II disease. Further validation is warranted to confirm this finding.

The two variants rs72767781 and rs72767784 are of low frequency in the European population, with a MAF of 0.03 in the QUASAR2 and SCOT cohorts. Both variants situate in an intergenic region, near an RNA gene *NR2F2-AS1*. This locus does not harbour any known cCRE [185]. As a result, the biological interpretation of this association is unclear.

5.2.3.2 A locus at 15q22.2 is associated with OS in early-stage CRC

Meta-analysis revealed a locus at chromosome 15q22.2 (lead variant: rs11447843, MAF = 0.027) that was associated with OS at genome-wide significance (HR = 2.56 [1.86-3.52], P = 8.70e-09; Fig 5.6B). The minor allele of rs11447843 was associated with poorer OS in both cohorts (QUASAR2: HR = 2.10 [1.16-3.79], P = 0.014, Fig 5.8C; SCOT: HR = 2.78 [1.90-4.05], P = 1.41e-07, Fig 5.8D). Interestingly, the same allele was associated with shorter TTR at suggestive significance in the meta-analysis (HR = 1.91 [1.49-2.43], P = 3.01e-07; Fig 5.6A, Fig 5.8A, B).

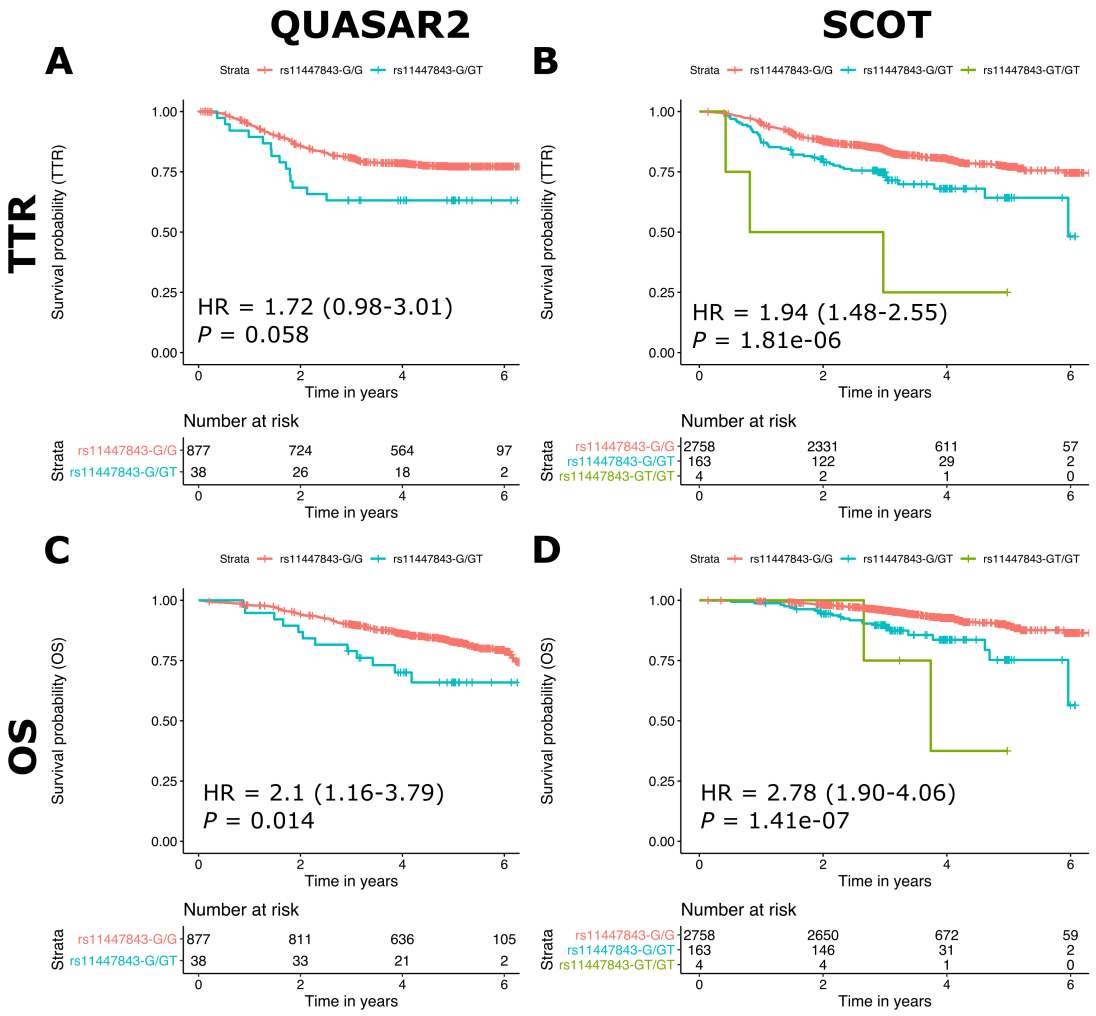

Figure 5.8. rs11447843 (15q22.2; MAF = 0.027) is associated with patient survival outcomes in QUASAR2 (A, C) and SCOT (B, D).

(A, B) For TTR. (C, D) For OS. The hazard ratio (HR) and the 95% CI were estimated from multivariate Cox regression, adjusted for known confounders; the nominal P-value was calculated by log-rank test (see Methods, section 2.3.8).

The variant rs11447843 situates in an intergenic region near *VPS13C*, which

encodes Vacuolar Protein Sorting 13 Homolog C. The biological significance of this variant remains to be elucidated. Considering the dual associational signal with both TTR and OS, it might represent a genuine prognostic variant early-stage CRC. Further validation of this association would be beneficial.

5.2.3.3 rs184423256 and clinical outcomes in early-stage CRC

Besides the genome-wide significant variants at 15q26.2, I identified a number of loci that were associated with TTR at suggestive significance in the meta-analysis (Fig 5.6A; Supplementary Data Table 5.4). One example is the variant rs184423256 at 8q22.3 (HR = 2.48 [1.77-3.46], P = 1.15e-07; Fig 5.6A). The minor allele of rs184423256 was associated with shorter TTR in both cohorts (QUASAR2: HR = 3.17 [1.71-5.87], P = 2.39e-04, Fig 5.9A; SCOT: HR = 2.23 [1.50-3.33], P = 8.32e-05, Fig 5.9B). I also observed that the same allele was associated with poorer OS at suggestive significance in the meta-analysis (QUASAR2: HR = 2.94 [1.57-5.49], P = 7.29e-04, Fig 5.9C; SCOT: HR = 2.99 [1.64-5.44], P = 3.32e-04, Fig 5.9D; meta-analysis: HR = 2.97 [1.93-4.59], P = 8.28e-07). Of note, these findings were not supported by the TCGA data (Fig 7.5A, C; Fig 7.6A, C).

The variant rs184423256 is relatively rare in the European population, with a MAF of 0.014 in both QUASAR2 and SCOT cohorts. It is located in the intronic region of *RRM2B*, which encodes the small subunit of a p53-inducible ribonucleotide reductase (p53R2), an essential enzyme in DNA synthesis. Currently, there is very limited evidence suggesting the biological basis of the association between rs184423256 and clinical outcomes in early-stage CRC. Up to date, no cCRE has been identified in this region [185]. This variant is not known to impact RRM2B expression levels either in normal tissues or in tumours (Fig 7.7). Therefore, further investigation will be essential to interpret the biological implications of this association.

5.2.3.4 rs114887409 and OS in early-stage CRC

Another example of candidate prognostic variant for early-stage CRC is the rs114887409 at 4p16.1. It showed association with OS in the meta-analysis at suggestive significance (HR = 3.19 [2.04-4.99], P = 3.69e-07; Fig 5.6B). The minor allele was

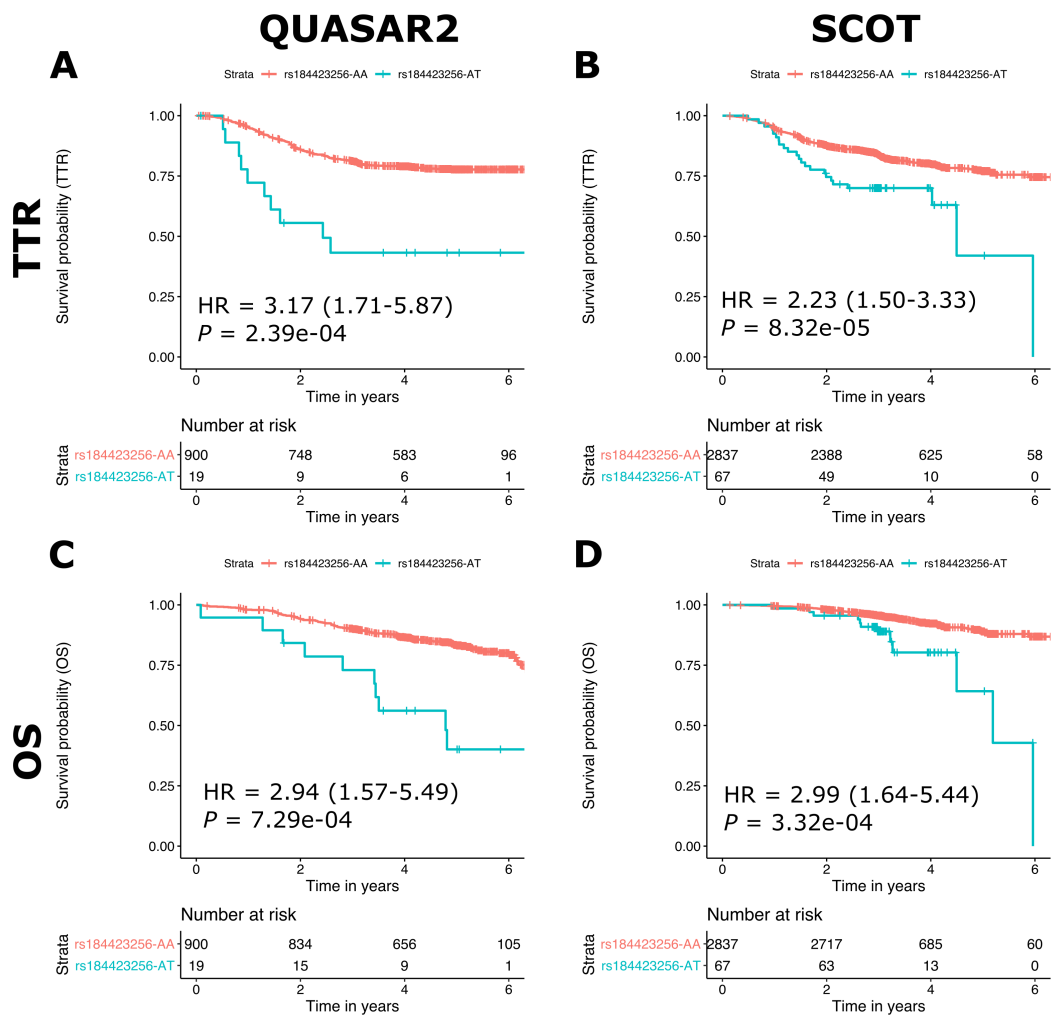

Figure 5.9. Suggestive variant rs184423256 (8q22.3, MAF = 0.014) and patient survival outcomes in QUASAR2 (A, C) and SCOT (B, D).

(A, B) For TTR. (C, D) For OS. The hazard ratio (HR) and the 95% CI were estimated from multivariate Cox regression, adjusted for known confounders; the nominal P-value was calculated by log-rank test (see Methods, section 2.3.8).

consistently associated with poorer OS in both clinical trial cohorts (QUASAR2: HR = 2.58 [1.38-4.83], P = 2.96e-03, Fig 5.10C; SCOT: HR = 3.96 [2.09-7.49], P = 2.34e-05, Fig 5.10D). In contrast, this variant was not significantly associated with TTR (meta-analysis: HR = 1.73 [1.16-2.59], P = 7.22e-03; Fig 5.10A, B).

Similar to the other top candidate loci shown above, the variant rs114887409 is of low frequency in the European population, with a MAF of 0.018 in the QUASAR2 and SCOT cohorts. It locates in the intronic region of *ABLIM2*, which encodes actin binding LIM protein 2. As the name suggests, this protein has a strong affinity for binding to F-actin and is known to involve in Netrin-1 signalling and axon guidance

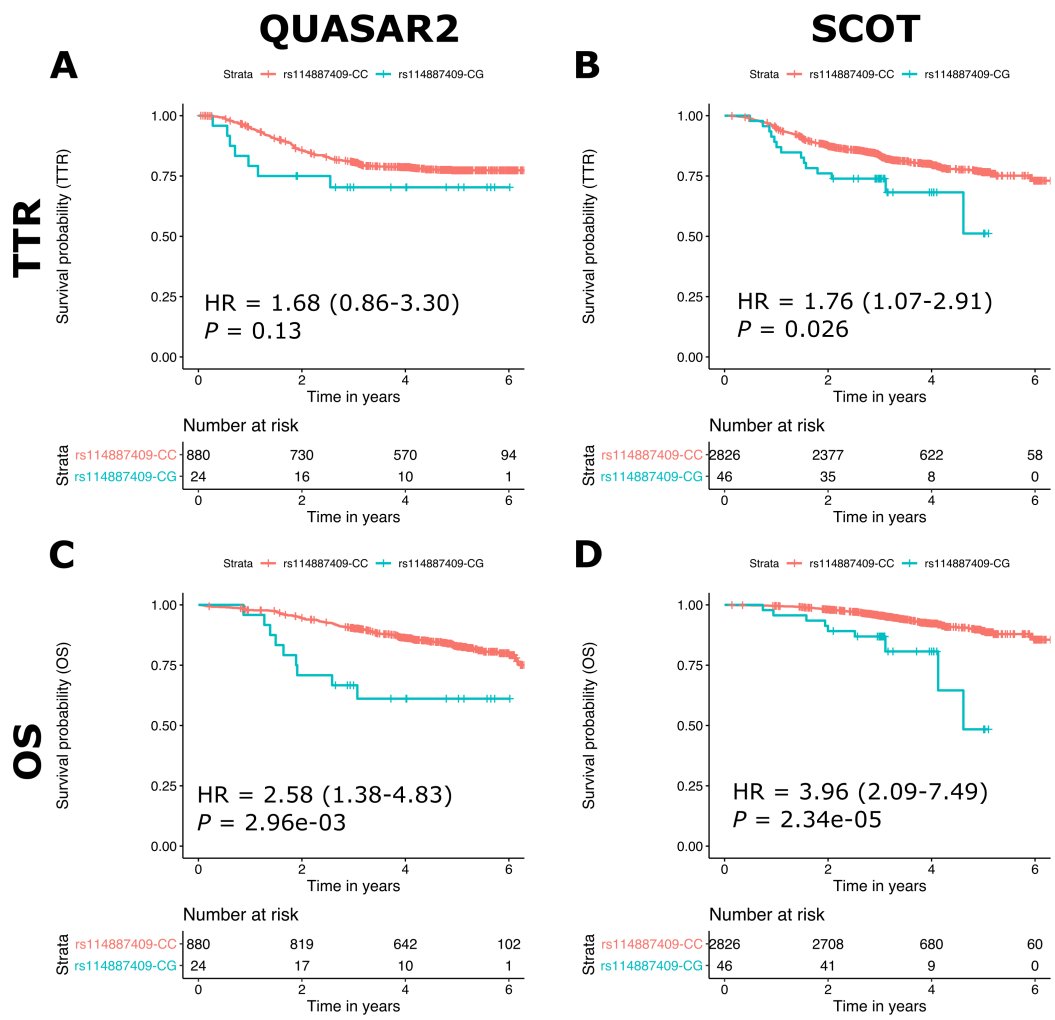

Figure 5.10. Suggestive variant rs114887409 (4p16.1, MAF = 0.018) and patient survival outcomes in QUASAR2 (A, C) and SCOT (B, D).

(A, B) For TTR. (C, D) For OS. The hazard ratio (HR) and the 95% CI were estimated from multivariate Cox regression, adjusted for known confounders; the nominal P-value was calculated by log-rank test (see Methods, section 2.3.8).

[365]. The biological basis of this association remains unknown.

5.2.3.5 Prognostic associations of candidate variants vary by KRAS mutational status in tumours

As mentioned in section 4.2.5.2, KRAS mutational status is known to influence clinical outcomes in CRC [174, 287–289]. However, it was not included as a covariate in the survival analyses reported above due to lack of molecular data in the SCOT clinical trial data (see also section 5.2.1). Considering the potential interactions between the CRC risk variants and KRAS mutational status in tumours on prog-

nosis (see section 4.2.5.2), I also examined the statistical relationship between the candidate prognostic variants presented above and KRAS mutational status in the QUASAR2 cohort, for which KRAS hotspot mutations in the tumours were determined. In this section, I provide evidence that inherited genetic variants interact with KRAS mutational status in tumours to associate with CRC prognosis.

To test the interactions between the top candidate prognostic variants and KRAS mutational status in tumours on clinical outcomes in the QUASAR2 cohort, I evaluated the statistical significance of variant by KRAS interaction in the original multivariate Cox regression model (see Methods, section 2.3.8). Of the four variants tested, I found that two variants (rs184423256 and rs114887409) and KRAS hotspot mutations interact to associate with TTR and OS (Table 5.3).

RSID	TTR	OS
rs72767781	0.41	0.72
rs11447843	0.53	0.37
rs184423256	0.044	0.45
rs114887409	0.051	0.088

Table 5.3. A summary of statistical significance of variant by KRAS interaction on CRC patient survival outcomes in the QUASAR2 cohort.

*The interaction was evaluated by calculating the P-value ($P_{interaction}$) of the multiplicative term 'SNP*KRAS' in the multivariate Cox regression model using a log-rank test (see Methods, section 2.3.8)*

I observed that carriers of the minor alleles of rs184423256 and rs114887409 showed a much shorter TTR if they have a KRAS MUT tumour, as compared to any other genotype-mutation groups (Fig 5.11, Fig 5.13). For instance, the rs184423256-AT genotype was associated with a much shorter TTR among patient with KRAS MUT tumours (HR = 12.22 [3.47-43.06], P = 9.81e-05; Fig 5.11D), than those with KRAS WT tumours (HR = 2.72 [1.17-6.29], $P = 0.02$; Fig 5.11C). Alternatively, this interaction also manifests that KRAS MUT was associated with shorter TTR among patients with the rs184423256-AT genotype than KRAS WT (HR = 12.24 [1.33-122], $P = 0.027$; Fig 5.11F), whereas KRAS mutational status was not associated with TTR among patients with the rs184423256-AA genotype (HR = 1.14 [0.83-1.58], $P = 0.42$; Fig 5.11E). Inter-group comparison revealed that rs184423256 genotype significantly interacts with KRAS mutational status to associate with TTR in the QUASAR2 cohort ($P_{interaction} = 0.044$; Fig 5.11B). A similar phenomenon was seen

for the variant rs114887409 (Fig 5.13). Moreover, carriers of the minor alleles of these two variants had much poorer OS if they have a KRAS MUT tumour (Fig 5.12; Fig 5.14). I noticed that the findings of rs184423256-KRAS interaction on CRC prognosis were not replicated in TCGA (Fig 7.5B, D; Fig 7.6B, D).

These results presented in this section suggest that in addition to existing CRC risk stratification schemes, genotype information of candidate prognostic variants could potentially benefit clinical management by identifying a subset of early-stage CRC cases that tend to have a much worse prognosis. Nevertheless, validation in additional CRC cohorts would be essential.

5.2.4 HLA genotypes and CRC prognosis

As introduced in section 5.1.3, HLA variations play an important role in adaptive immune response, potentially influencing tumour development. However, the survival analysis presented in the previous sections did not include inherited genetic variants in the HLA region, because the extreme sequence variation and widespread LD in the region requires specialised reference panel and algorithms for genotype imputation. To investigate whether inherited genetic variants in the HLA genes predict prognosis in early-stage CRC, in this section, I present results from the survival analysis targeted at the HLA region in the QUASAR2 and SCOT clinical trial cohorts.

I imputed genotypes in the HLA region using a reference panel densely typed for HLA variations [177] and classified them into the gold-standard four-digit types (see Methods, section 2.1.3.3). In total, 908 patients in the QUASAR2 trial and 2755 patients in the SCOT trial were successfully genotyped and imputed for HLA variants.

5.2.4.1 HLA homozygosity and clinical outcomes in early-stage CRC

Previously, homozygosity in HLA class I genes (*HLA-A, -B* and *-C*) were associated with poorer survival in advanced melanoma patients undergoing immunotherapy [352]. However, whether HLA homozygosity has similar prognostic values in early-stage CRC is unclear. In this section, I examined the prognostic value of HLA

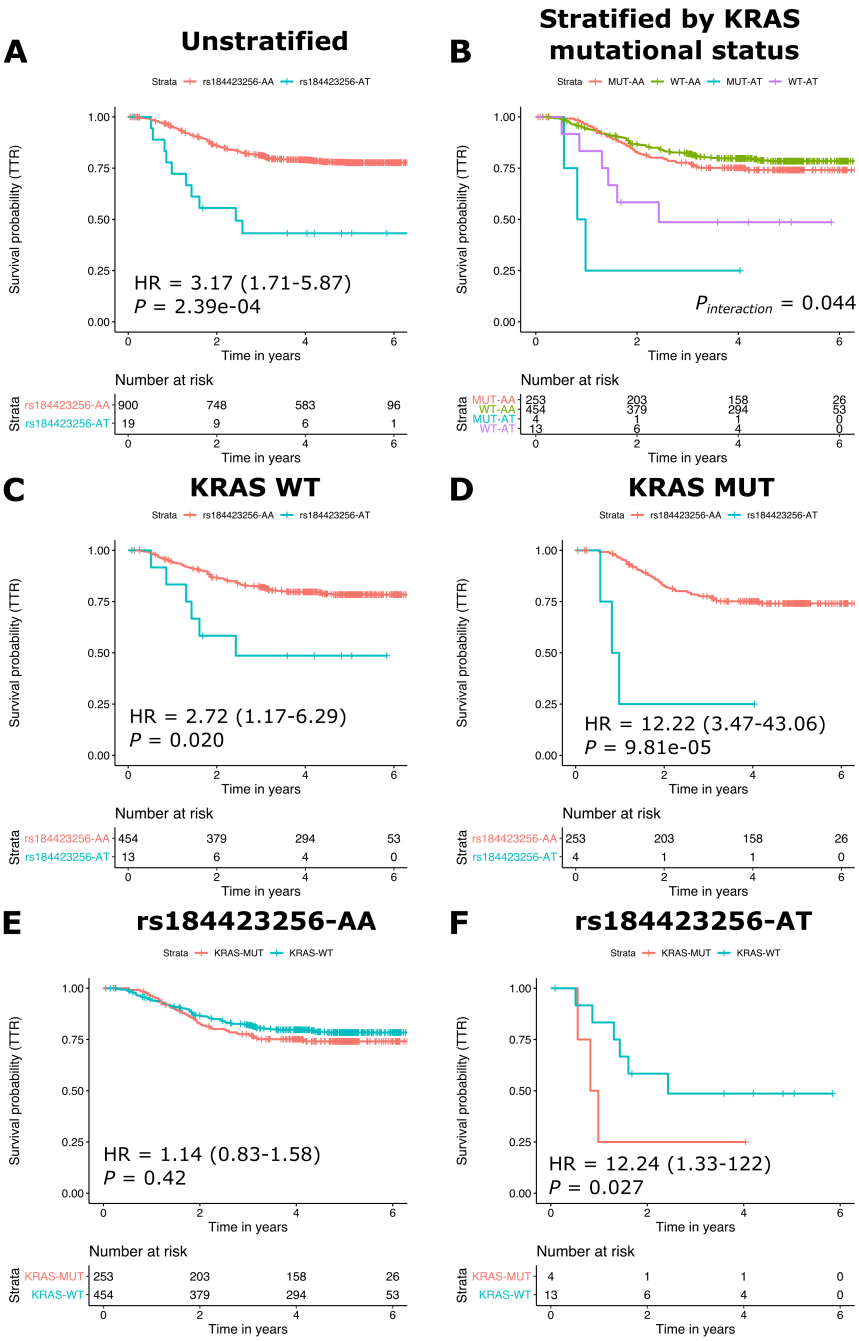

Figure 5.11. rs184423256 (8q22.3, MAF = 0.014) interacts with KRAS mutational status in tumours to associate with TTR in QUASAR2.

(A, B) Association of variant genotype with TTR when patients were unstratified (A; as in Fig 5.9A) or when patients were stratified by KRAS mutational status (B). (C, D) Association of variant genotype with TTR among patients with KRAS WT CRC (C) or KRAS MUT CRC (D). (E, F) Association of KRAS mutational status with TTR among patients with rs184423256-AA genotype (E) or rs184423256-AT genotype (F). The hazard ratio (HR) and the 95% CI were estimated from multivariate Cox regression, adjusted for known confounders; the nominal P-value was calculated by log-rank test (see Methods, section 2.3.8).

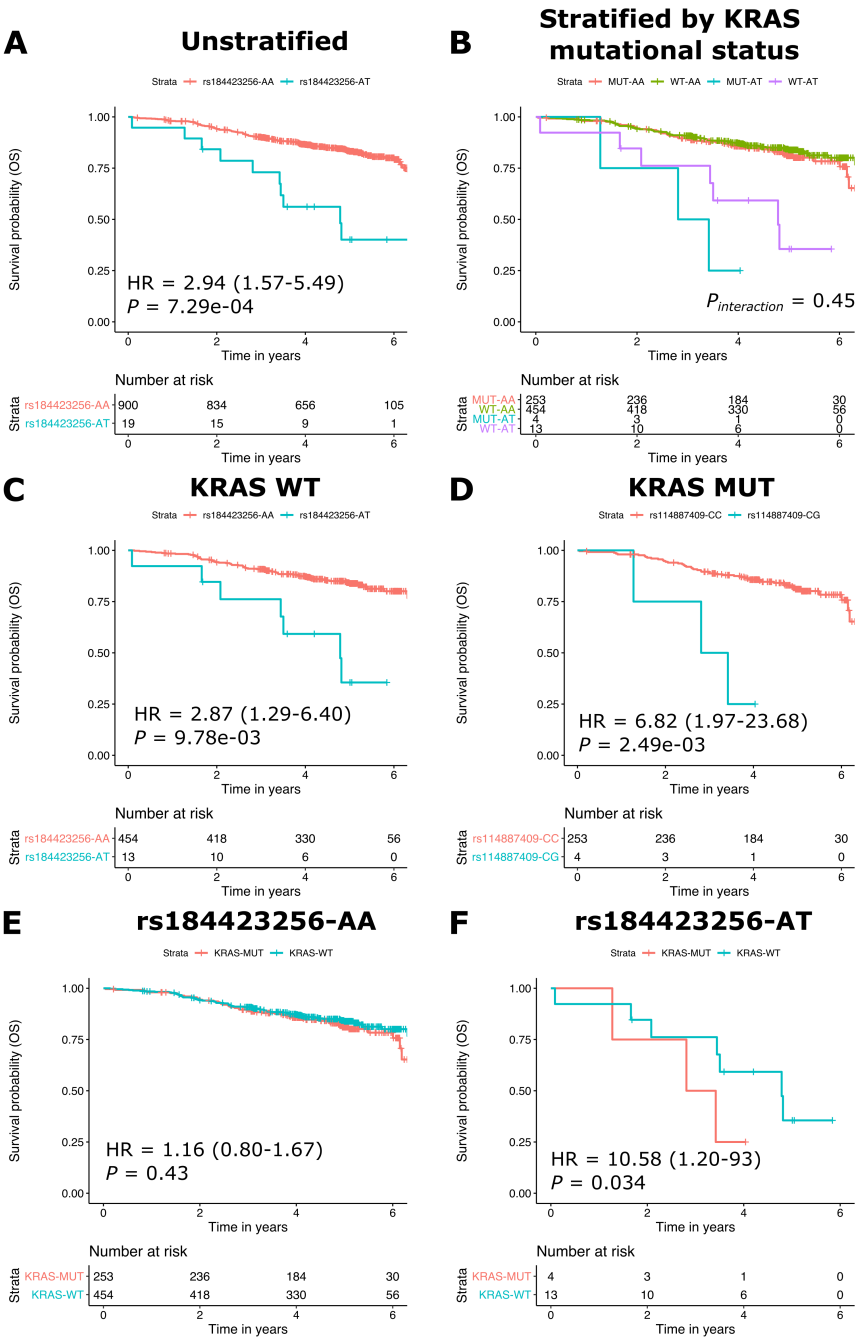

Figure 5.12. rs184423256 (8q22.3, MAF = 0.014) interacts with KRAS mutational status in tumours to associate with OS in QUASAR2.

(A, B) Association of variant genotype with OS when patients were unstratified (A; as in Fig 5.9C) or when patients were stratified by KRAS mutational status (B). (C, D) Association of variant genotype with OS among patients with KRAS WT CRC (C) or KRAS MUT CRC (D). (E, F) Association of KRAS mutational status with OS among patients with rs184423256-AA genotype (E) or rs184423256-AT genotype (F). The hazard ratio (HR) and the 95% CI were estimated from multivariate Cox regression, adjusted for known confounders; the nominal P-value was calculated by log-rank test (see Methods, section 2.3.8).

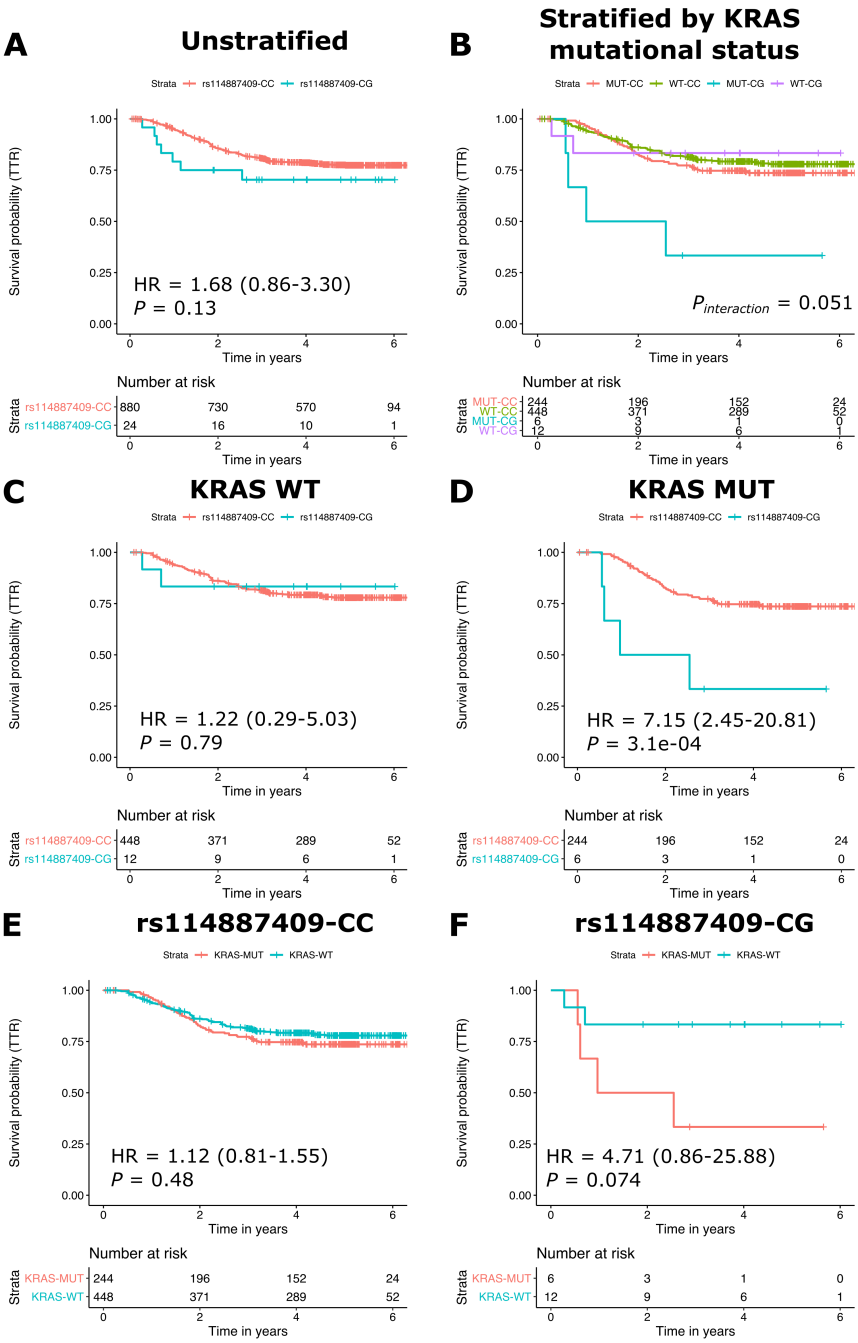

Figure 5.13. rs114887409 (4p16.1, MAF = 0.018) interacts with KRAS mutational status in tumours to associate with TTR in QUASAR2.

(A, B) Association of variant genotype with TTR when patients were unstratified (A; as in Fig 5.10A) or when patients were stratified by KRAS mutational status (B). (C, D) Association of variant genotype with TTR among patients with KRAS WT CRC (C) or KRAS MUT CRC (D). (E, F) Association of KRAS mutational status with TTR among patients with rs114887409-CC genotype (E) or rs114887409-CG genotype (F). The hazard ratio (HR) and the 95% CI were estimated from multivariate Cox regression, adjusted for known confounders; the nominal P-value was calculated by log-rank test (see Methods, section 2.3.8).

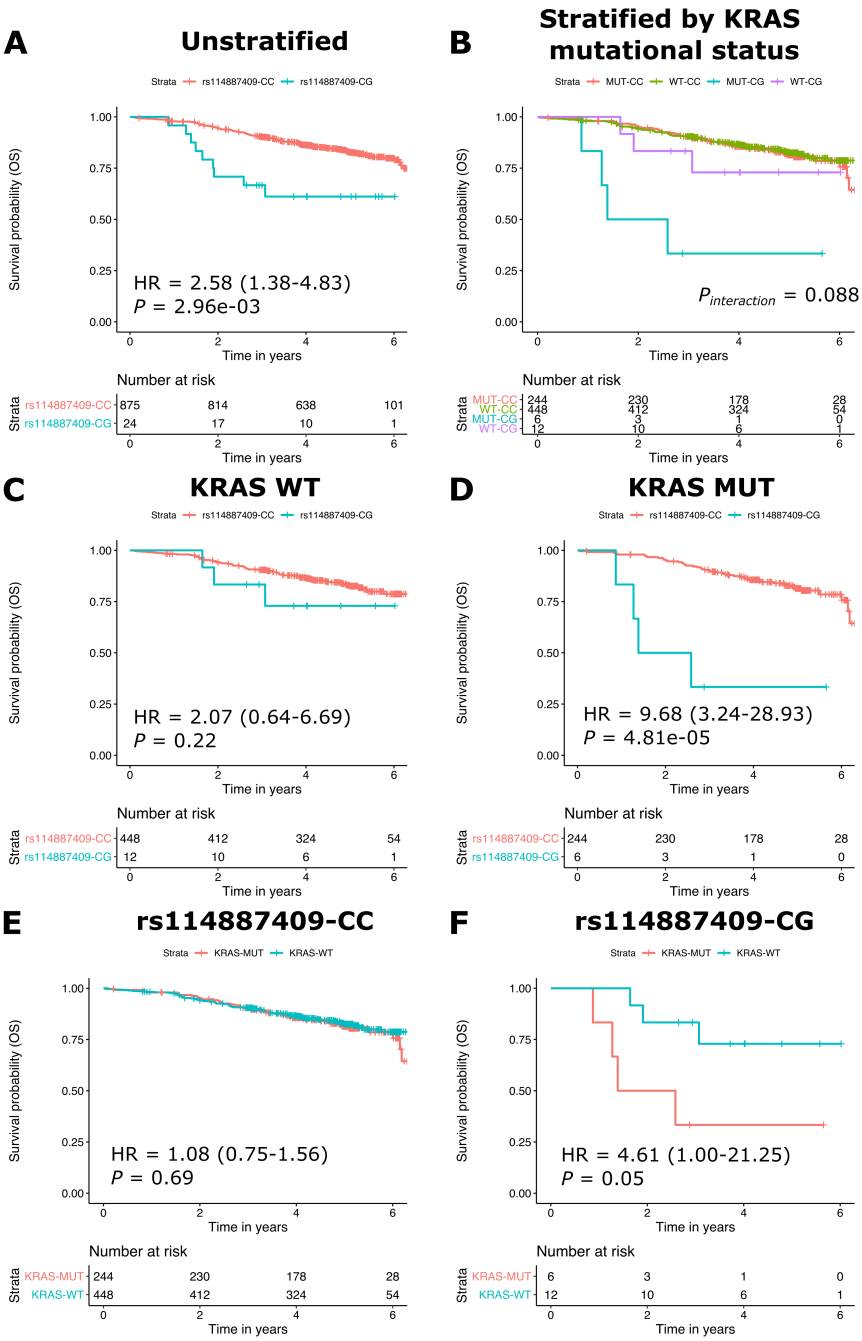

Figure 5.14. rs114887409 (4p16.1, MAF = 0.018) interacts with KRAS mutational status in tumours to associate with OS in QUASAR2.

(A, B) Association of variant genotype with OS when patients were unstratified (A; as in Fig 5.10C) or when patients were stratified by KRAS mutational status (B). (C, D) Association of variant genotype with OS among patients with KRAS WT CRC (C) or KRAS MUT CRC (D). (E, F) Association of KRAS mutational status with OS among patients with rs114887409-CC genotype (E) or rs114887409-CG genotype (F). The hazard ratio (HR) and the 95% CI were estimated from multivariate Cox regression, adjusted for known confounders; the nominal P-value was calculated by log-rank test (see Methods, section 2.3.8).

homozygosity in the QUASAR2 and SCOT cohorts.

Potentially, homozygous HLA alleles would restrict the diversity of antigens to be presented on the cell surface and alter the repertoire of antigen presentation, thus might influence immunoediting during tumour development. I compared the clinical outcomes of patients with varying number of homozygous loci in the three HLA class I genes (*HLA-A, -B* and *-C*). Assuming an additive effect, I calculated the hazard ratio of having homozygous genotype in HLA-I genes on TTR and OS in both cohorts, using a multivariate Cox regression model as in the preceding sections (see Methods, section 2.3.8).

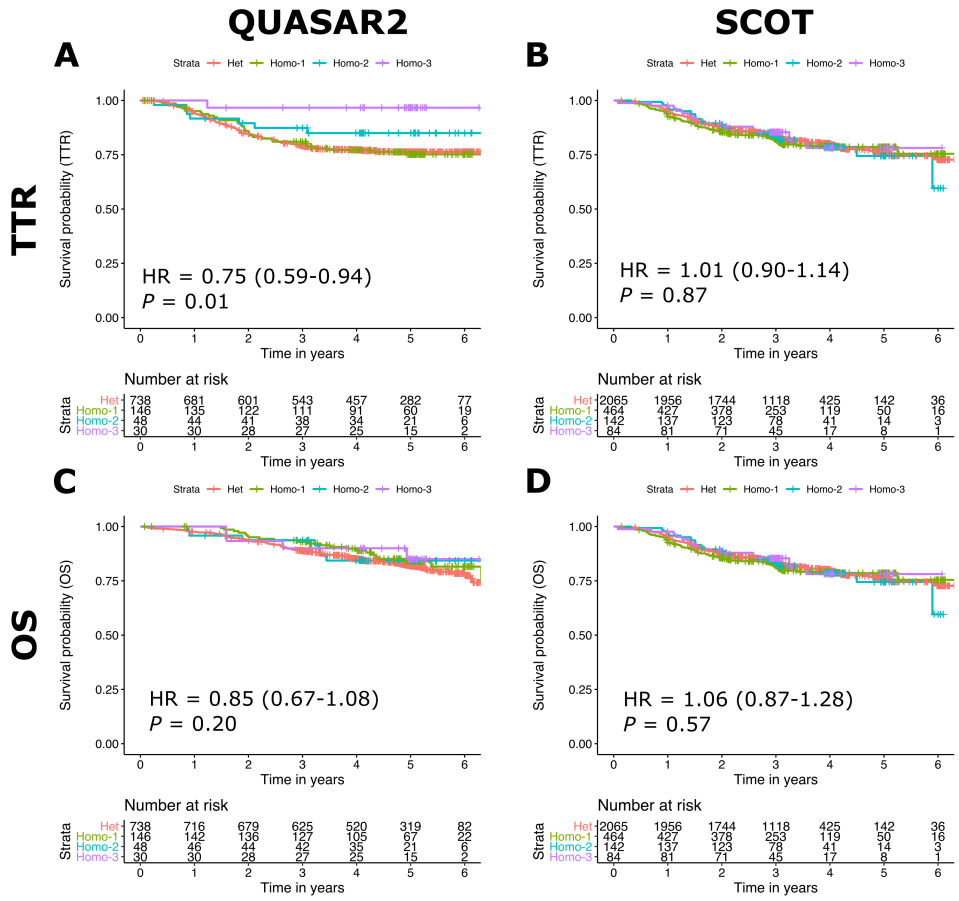

Figure 5.15. Homozygosity in HLA class I genes (HLA-A, -B and -C) and patient survival outcomes in QUASAR2 (A, C) and SCOT (B, D).

(A, B) For TTR. (C, D) For OS. The hazard ratio (HR) and the 95% CI were estimated from multivariate Cox regression, adjusted for known confounders; the nominal P-value was calculated by log-rank test (see Methods, section 2.3.8).

Overall, I found no significant difference in clinical outcomes between patient groups with different HLA class I homozygosity (Fig 5.15; Supplementary Data Table 5.6). I identified a trend that HLA class I homozygosity was associated with

longer TTR in QUASAR2 cohort (HR = 0.75 [0.59-0.94], P = 0.01; Fig 5.15A). However, this was not validated in the SCOT cohort (HR = 1.01 [0.90-1.14], P = 0.87; Fig 5.15B). A closer examination of the QUASAR2 dataset revealed that the putative relationship was dominated by the association of homozygosity in *HLA-B* gene with TTR (HR = 0.39 [0.20-0.77], P = 6.08e-03; Fig 7.8A), which is the most diverse amongst all HLA genes. After removing individuals with homozygous HLA-B genotype, homozygosity in *HLA-A* and *-C* was not associated with TTR (HR = 0.97 [0.66-1.4], P = 0.88; Fig 7.8C). Given that the association of homozygosity in *HLA-B* gene with TTR was not replicated in the SCOT cohort (HR = 1.05 [0.79-1.41], P = 0.72; Fig 7.8B), which has a larger sample size, the observed relationship between *HLA-B* and TTR in the QUASAR2 cohort is likely a spurious association. Alternatively, the observed association could be due to individual *HLA-B* genotypes. Future development could consider an in-depth subgroup analysis.

Next, I examined the clinical outcomes of patients with varying number of homozygous loci in the five HLA class II genes (*HLA-DPA1, -DPB1, -DQA1, -DQB1* and *-DRB1*). Homozygosity in HLA class II genes showed no reproducible association with clinical outcomes in the QUASAR2 and SCOT cohorts (Fig 5.16; Supplementary Data Table 5.7). I noticed that HLA class II homozygosity was associated with poorer OS in the SCOT cohort (HR = 1.15 [1.03-1.28], P = 0.01; Fig 5.16D). However, patients with homozygous alleles in all five HLA class II genes (annotated as 'Homo-5') had similar TTR as those with heterozygous alleles in all five genes (Fig 5.16D), obscuring the interpretation of this association. Given that the association was not replicated in QUASAR2 (HR = 0.95 [0.83-1.08], P = 0.42; Fig 5.16C), again, this is possibly a spurious finding.

In summary, the study of associations between homozygosity in HLA genes and clinical outcomes in the QUASAR2 and SCOT clinical trial cohorts showed that HLA homozygosity has no convincing association with clinical outcomes in early-stage CRC undergoing adjuvant chemotherapy. Study of advanced CRC cohorts or those undergoing immunotherapies would be of great benefit to further investigate the prognostic role of HLA homozygosity in CRC (see discussion, section 5.3).

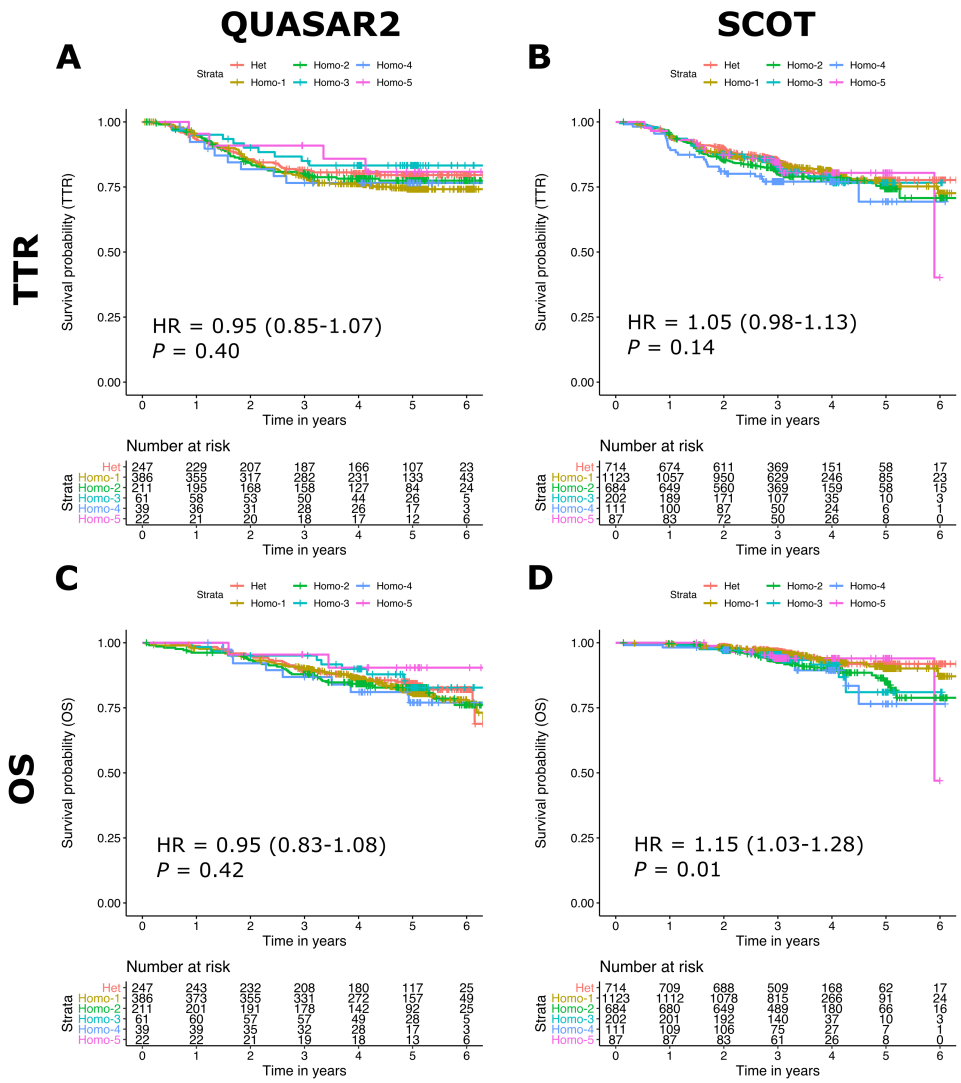

Figure 5.16. Homozygosity in MHC class II genes (HLA-DPA1, -DPB1, -DQA1, -DQB1 and -DRB1) on patient survival outcomes in QUASAR2 (A, C) and SCOT (B, D).

(A, B) For TTR. (C, D) For OS. The hazard ratio (HR) and the 95% CI were estimated from multivariate Cox regression, adjusted for known confounders; the nominal P-value was calculated by log-rank test (see Methods, section 2.3.8).

5.2.4.2 HLA class I supertypes

Even if HLA genotypes are grouped by four-digit typing (see section 5.1.3), HLA alleles are still diverse, most of which are rare (MAF < 0.01). This poses difficulties for survival analysis examining the association of individual HLA alleles with patient survival outcomes, primarily due to low statistical power. Studies suggest that HLA four-digit types can be further grouped based on similar peptide-anchor-binding specificities, which are critical for presenting antigens on the cell surface [143, 179, 180]. Particularly, four-digit types of related HLA-A and -B protein isoforms can be

grouped into 12 supertypes, covering the majority of *HLA-A* and *-B* alleles in human populations [179, 180]. Recently, a study demonstrated that HLA B44 supertype was associated with improved OS among advanced melanoma patients undergoing immunotherapy [352]. However, no such study has been conducted in CRC. In this section, I explore whether the 12 HLA supertypes are correlated with clinical outcomes in the QUASAR2 and SCOT clinical trial cohorts.

First, I assigned *HLA-A* and *-B* supertypes to patients in the two cohorts, based on the four-digit HLA types determined from genotype imputation (see Methods, section 2.1.3.3). The distribution of the supertypes in the two cohorts coincided with each other (Fig 5.17). The frequencies were also comparable to those reported in European populations [352].

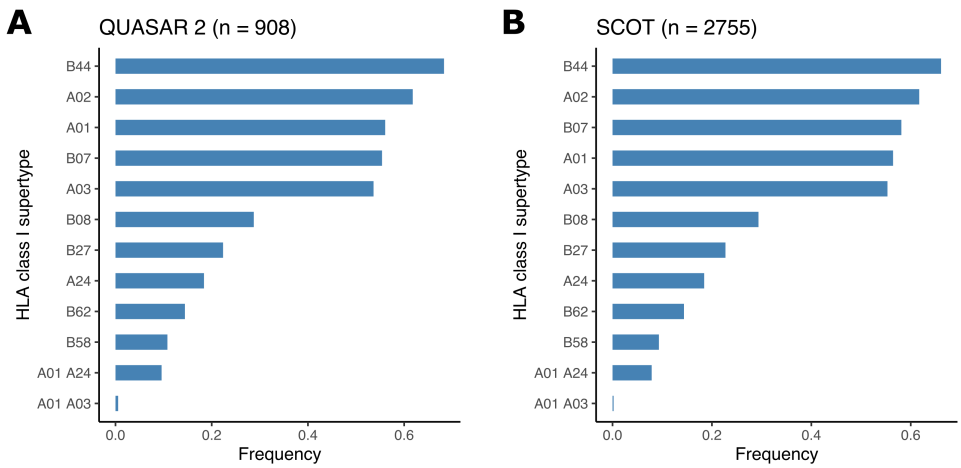

Figure 5.17. Frequency of HLA-A and -B supertypes and in QUASAR2 (A) and SCOT (B).

Next, I investigated whether individual *HLA-A* and *-B* supertypes have predictive value in CRC prognosis. Interestingly, two supertypes were associated with both TTR and OS in the QUASAR2 cohort (Fig 5.18). Specifically, the A03 supertype was associated with longer TTR (HR = 0.68 [0.54-0.87], P = 1.84e-03; Fig 7.9A) and better OS (HR = 0.74 [0.56-0.97], P = 0.03; Fig 7.9C); whereas the B62 supertype was associated with shorter TTR (HR = 1.43 [1.04-1.97], P = 0.03; Fig 7.10A) and poorer OS (HR = 1.50 [1.07-2.11], P = 0.02; Fig 7.10C). However, these associations were not replicated in the SCOT cohort (Fig 5.18; Fig 7.9B, D; Fig 7.10B, D).

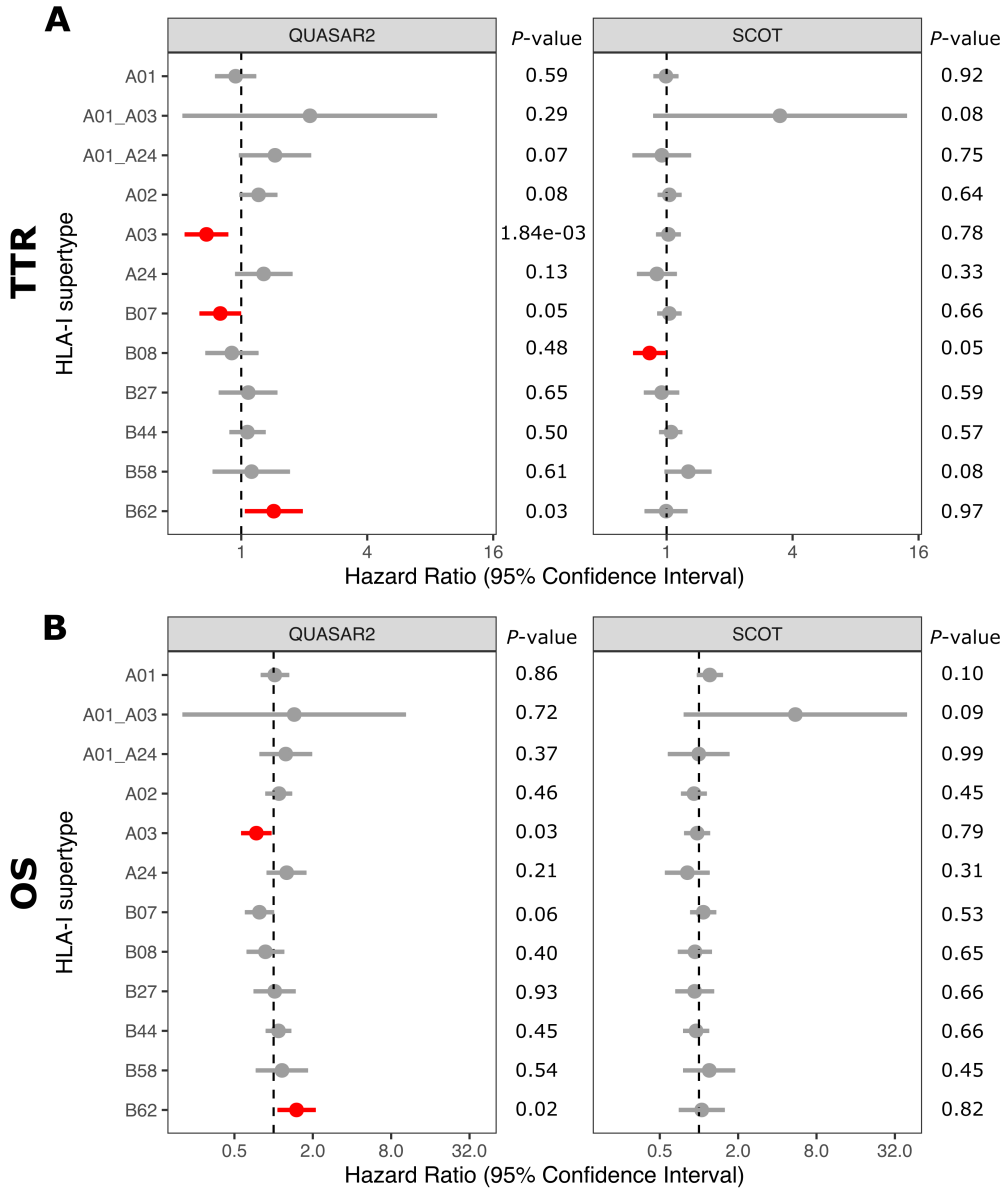

Figure 5.18. The association of HLA-A and -B supertypes with patient survival outcomes in QUASAR2 and SCOT.

(A) Forest plots showing the association of individual supertype with TTR. (B) Forest plots showing the association of individual supertype with OS. The hazard ratio (HR) and the 95% CI were estimated from multivariate Cox regression, adjusted for known confounders; the nominal P-value was calculated by log-rank test (see Methods, section 2.3.8).

This lack of concordance in results between the two cohorts (Fig 5.18; Supplementary Data Table 5.8) indicates that *HLA-A* and *-B* supertypes are not associated with clinical outcomes in early-stage CRC. It is possible that HLA-I-dependent immunoediting in CRC is stage-specific or plays a less prominent role in disease progression among patients undergoing adjuvant chemotherapy (see also section 5.2.1),

compared to patients with advanced melanoma treated with immune checkpoint blockade therapy (see discussion, section 5.3).

5.2.5 Mapping QTLs for infiltrating CD8+ T cells in CRC

Identifying genetic determinants of specific patterns of tumour infiltrating immune cells is crucial to understanding the role of inherited genetic variants in cancer. It can not only aid in characterising the variability of immune response to tumours between individuals [312], but also help uncovering genetic drivers of immune cell signatures in TME that are prognostic for disease progression and predictive of response to treatment [333]. As introduced in section 5.1.4, four studies have attempted mapping immune quantitative trait loci (iQTLs) in tumours [153–155, 333]. However, they used the TCGA pan-cancer cohort data and relied solely on predicted immune cell abundance or fraction in tumours based on gene expression profiles. Previously, tumour samples in the QUASAR2 cohort were characterised for infiltrating immune cell composition via *in situ* immunohistochemistry staining (see Methods, section 2.3.9; [194]). In this section, leveraging this high-quality data, I aim to map iQTLs in CRC, with a focus on tumour infiltrating CD8+ T cell levels.

Among the 929 QUASAR2 trial participants that have been genotyped, 760 have valid entries for tumour infiltrating CD8+ T cell density (Fig 5.19A). I performed a genome-wide scan for common inherited genetic variants associated with this immune quantitative trait (see Methods, section 2.3.9; Fig 5.19B) with no prespecified covariates. One locus at 14q23.1 (lead SNP: rs78959096, MAF = 0.013) was associated with tumour infiltrating CD8+ T cell density at genome-wide significance (Fig 5.19C); 21 loci reached suggestive significance (Fig 5.19D). A full list of these genetic variants is provided in Supplementary Data Table 5.9. Note that these variants do not overlap with the iQTLs previously mapped to inferred CD8+ T cell abundance in TCGA pan-cancer tumours [155].

The minor allele of the top variant, rs78959096 (14q23.1; nearby gene: *DACT1*), was associated with a 2.6-fold decrease in infiltrating CD8+ T cell density in CRC (beta = -1.38, P = 5.60e-09; Fig 5.19C). This variant is located in an intergenic region, 1.5 kb away from a known cCRE (EH38E1717801; high CTCF binding,

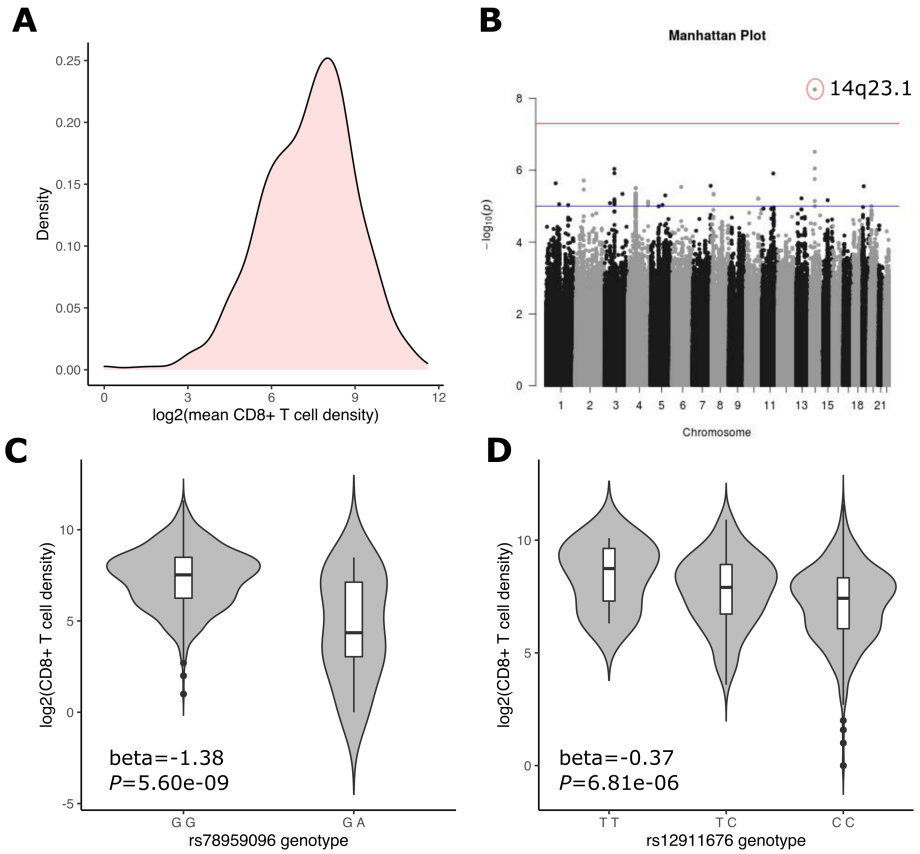

Figure 5.19. Mapping iQTLs for infiltrating CD8+ T cell density in CRC.
(A) Distribution of tumour infiltrating CD8+ T cell density (log2 transformed) in QUASAR2. (B) Manhattan plot of the iQTL mapping, testing the association of individual genetic variants with the immune trait. (C, D) Violin plots of the top variants associated with the infiltrating CD8+ T cell density in CRC.

indicative of chromatin looping activity; identified in both normal and cancerous tissues). Both the variant and the cCRE are downstream of *DACT1*, which encodes a protein involved in Wnt and GPCR signalling pathways. Whether this variant regulates nearby gene expression would be an important topic for future studies. In addition, I observed that one of the suggestive variants, rs12911676 (Fig 5.19D), was also associated with the relative percentage of Tregs in rectal cancer in the TCGA cohort (beta = 5.93, P = 2.12e-05; Fig 7.11; [366]).

These findings suggest that inherited genetic variants are associated with infiltrating CD8+ T cell density in CRC. Further validation and studies of variant functions would be highly valuable.

5.2.6 Causal inference on tumour infiltrating CD8+ T cell density and CRC prognosis

The data presented thus far demonstrate that inherited genetic variants are associated with CRC prognosis as well as infiltrating CD8+ T cell density in CRC. As introduced in 5.1.2, adaptive immune response plays a crucial part in influencing clinical outcomes in CRC. Nevertheless, the causal relationship between these two variables is yet to be delineated. In this section, I aimed to investigate the causal relationship between tumour infiltrating CD8+ T cell density and clinical outcomes in early-stage CRC.

5.2.6.1 Mendelian randomisation suggests a marginal causal association of tumour infiltrating CD8+ T cells with OS in early-stage CRC

As introduced in section 5.1.5, Mendelian randomisation (MR) leverages information from inherited genetic variants to estimate the unconfounded effect of an exposure on an outcome, and is the method of choice for inferring causal relationship between the two. In this section, I apply two-stage MR (Fig 5.2C) to evaluate the causal association of tumour infiltrating CD8+ T cell density with CRC prognosis in the QUASAR2 cohort.

First, I constructed a genetic instrument for tumour infiltrating CD8+ T cell density using the 22 iQTLs identified in section 5.2.5. Specifically, I calculated a weighted genomic risk score (GRS) for each individual using their genotype information at the 22 iQLTs (see Methods, section 2.3.10). As expected, the weighted GRS was significantly associated with the log2-transformed CD8+ T cell density in a linear regression model (P = 7.87e-61), explaining 30% of the variation in the immune trait (Fig 5.20B). The F-statistic of the linear regression was 323.5 for the weighted GRS, indicating that the constructed genetic instrument was a strong one for MR analysis (F > 10; [364]). None of the 22 variants utilised for the genetic instrument was associated with any demographic, pathological or known prognostic factors including age, gender, tumour location, stage and treatment (see Methods, section 2.3.10). Importantly, none of the 22 variants were or in LD ($R^2 > 0.2$) with any known prognostic variants in CRC (Supplementary Data Table 5.9).

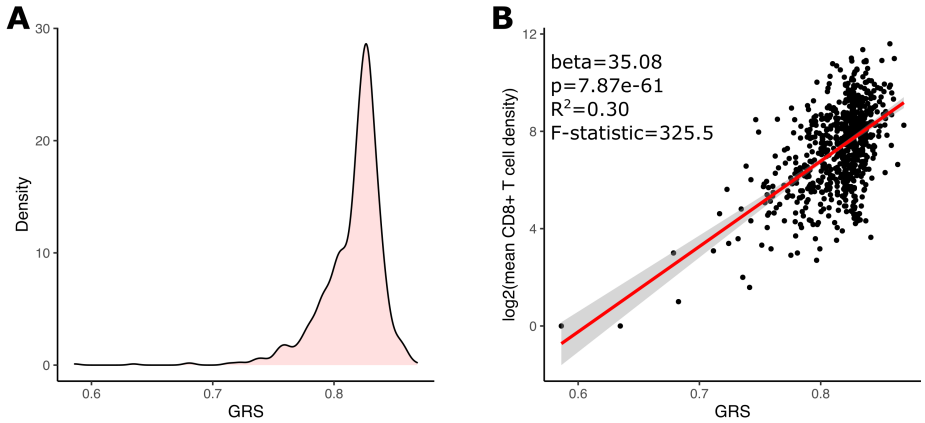

Figure 5.20. Constructing a genetic instrument variable for infiltrating CD8+ T cell density in CRC.

(A) Distribution of the weighted GRS that was used to construct the genetic instrument for infiltrating CD8+ T cell density in CRC. (B) Scatter plot of the log2-transformed CD8+ T cell density in CRC against the weighted GRS, fitted with a linear model.

Next, I evaluated the association between the genetic instrument for tumour infiltrating CD8+ T cell density and patient survival outcomes to estimate the causal effect of tumour infiltrating CD8+ T cell density on CRC prognosis (see Methods, section 2.3.10). I observed no significant association between the genetic instrument and TTR in the QUASAR2 cohort. Results from univariate and multivariate Cox regression adjusted for known prognostic factors (age, gender, tumour location, pathological and nodal stage, and treatment) failed to confirm a causal association between the exposure (tumour infiltrating CD8+ T cell density in CRC) and time to CRC recurrence in QUASAR2 (HR = 0.93 [0.80-1.07], $P = 0.29$; Table 5.4). On the other hand, I detected a marginal causal association between the exposure and OS. The causal estimate of tumour infiltrating CD8+ T cell density on all-cause mortality was 0.84 in a multivariate Cox regression model (95% CI: [0.72-0.97], $P = 0.018$; Table 5.4). This indicates that increasing tumour infiltrating CD8+ T cell density decreases the risk of all-cause death in early-stage CRC.

During the construction of genetic instrument for the immune trait, I noticed there were four outlier individuals with extreme small GRS (< 0.7; Fig 5.20B), which might bias the first-stage regression. However, the causal estimate of tumour infiltrating CD8+ T cell density on clinical outcomes in the dataset was not sensitive to these outlier samples, because the test statistics remained largely unchanged when I removed the four individuals from the two-stage MR (Table 7.1).

	CRC recurrence		All-cause death	
Cox model	Causal Estimate[b]	*P*-value	Causal Estimate[b]	*P*-value
Univariate	HR=0.90 (0.78-1.04)	0.17	HR=0.83 (0.71-0.96)	0.014
Multivariate[a]	HR=0.93 (0.80-1.07)	0.29	HR=0.84 (0.72-0.97)	0.018

Table 5.4. Main results of two-stage Mendelian randomisation analysis using individual-level data in the QUASAR2 clinical trial cohort.

[a] *Multivariate Cox regression model adjusted by age, sex, tumour location (rectum vs. colon), pT stage (p4 vs. p1-3), nodal stage (N1/2 vs. N0), and treatment (see Methods, section 2.3.8).* [b] *The causal estimate of exposure on survival outcome refers to change in hazard ratio per unit increase in log2-transformed tumour-infiltrating CD8+ T cell density.*

Taken together, these results suggest that the implemented two-stage MR does not support a causal association between tumour infiltrating CD8+ T cell density and TTR, whereas it provides marginal evidence for a causal association between tumour infiltrating CD8+ T cell density and OS in the QUASAR2 cohort. Investigation in additional CRC cohorts is required to validate these findings.

Although MR proves robust, it does come with strict assumptions and severe constraints. Amongst others, the assumption of linearity and the parametric natural of inference might have limited the scope of inquiry. Therefore, nonlinear relationships between tumour-infiltrating CD8+ T cell density and clinical outcomes in CRC remain plausible.

5.2.6.2 Mutual information inference

Despite the strong association demonstrated by Glaire et al. [194], I did not observe any causal link between tumour infiltrating CD8+ T cell density and TTR in the two-stage MR study in section 5.2.6.1. Considering potential nonlinear relationship between tumour infiltrating CD8+ T cell density and clinical outcomes in CRC, I was motivated to employ a different method to infer the causal relationship between the immune trait and patient survival outcomes in the QUASAR2 cohort. Mutual information (MI) has emerged as a non-parametric framework to examine nonlinear structure in data [197]. Related to entropy (i.e., uncertainty in observations of a random variable), it operates by evaluating dependencies of information contents between variables. In the context of causal inference, it has been suggested as a general measure of the presence of nonlinear relationship [367, 368]. In this section,

I aim to assess potential nonlinear relationships between tumour infiltrating CD8+ T cell density and patient survival outcomes in CRC using MI and related measures.

After optimising the binning scheme for estimating MI in the dataset, I calculated the MI between log2-transformed tumour infiltrating CD8+ T cell density and patient survival outcomes to be 1.45 (for CRC recurrence) and 1.44 (for all-cause death) (unit: bit; see Methods, section 2.3.11). These estimates were invariant to right-censoring (Fig 5.21A-B), as seen in sensitivity analyses where three scenarios of data censoring were simulated (Fig 7.12).

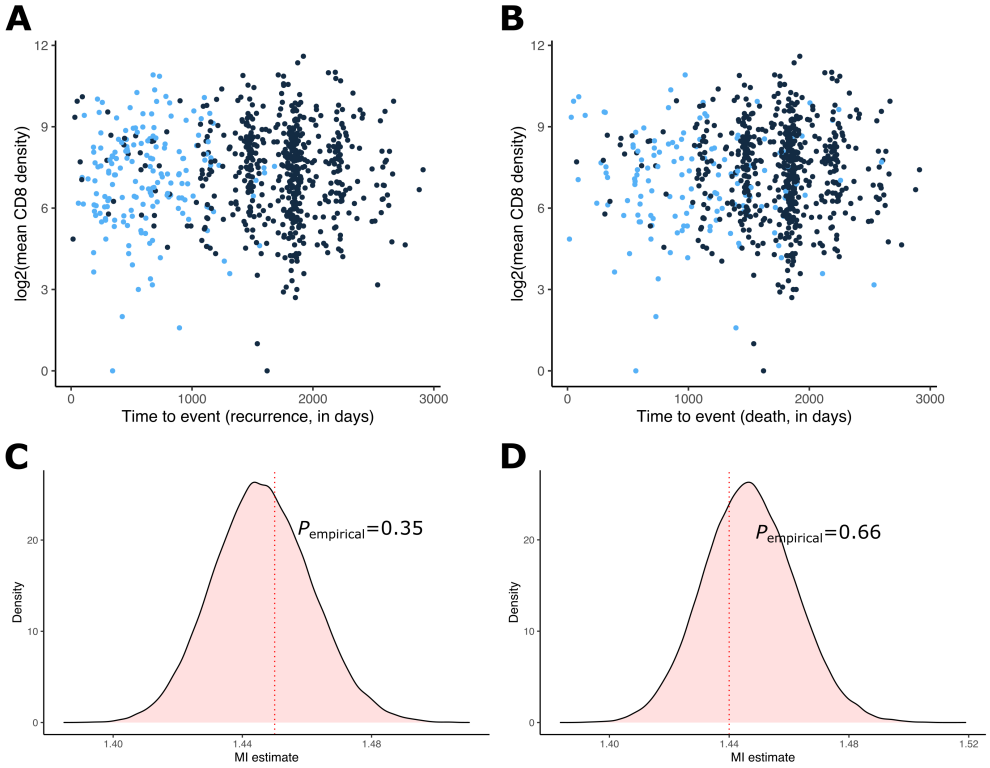

Figure 5.21. Assessing nonlinear relationship between the tumour infiltrating CD8+ T cell density and patient survival outcomes using mutual information.

(A, B) Scatter plots of the immune phenotype against time to event variables, where both observed and censored events are included: cancer recurrence (A) or death (B). Cases with events were highlighted in light blue, while censored cases are in navy. (C, D) Empirical distributions of MI estimates obtained via bootstrapping: CRC recurrence (C) or all-cause death (D). MI estimated from original datasets were marked by red dotted lines.

To assess the statistical significance of the MI estimates, I obtained the empirical distribution of these measures via bootstrapping (see Methods, section 2.3.11). Compared to the empirical distribution, neither of the two estimates supports the presence of nonlinear relationship in the dataset ($P_{empirical} > 0.05$; Fig 5.21C-D).

Moreover, the maximal information coefficients [197]) were measured to be in the range of 0.13-0.20, even when the data was stratified by pT and N stage (see Methods, section 2.3.11). Again, these results do not support the presence of nonlinear relationship in the dataset.

In summary, these results indicate a lack of detectable data structure between tumour infiltrating CD8+ T cell density and patient survival outcomes in the QUASAR2 cohort. Together, a nonlinear relationship between the immune trait and CRC clinical outcomes remains unproven.

5.3 Discussion

In this chapter, I performed a suite of association studies to evaluate the role of inherited genetic variants on prognosis and adaptive immune response in early-stage CRC. Utilising data from the QUASAR2 and SCOT clinical trial cohorts, totalling 3,858 high-risk stage II or stage III cases, I first conducted a genome-wide search for inherited genetic variants associated with TTR or OS. Through meta-analysis, I identified a locus at 15q26.2 and a locus at 15q22.2 that were significantly associated with TTR and OS, respective ($P <$ 5e-08; Fig 5.6, Fig 5.7, Fig 5.8). I also identified a number of loci associated with both TTR and OS at suggestive significance in the meta-analysis ($P <$ 1e-05; Fig 5.6, Fig 5.9, Fig 5.10). Further investigation would be of benefit to validate these findings and to provide biological insight into the associations.

Intriguingly, I observed that two candidate prognostic variants showed evidence of interacting with KRAS hotspot mutations to associate with CRC outcome. In particular, the minor alleles of rs184423256 and rs114887409 were associated with considerably shorter TTR and OS in patients with KRAS MUT tumours (with a HR of up to 12), as compared to those KRAS WT tumours (Fig 5.11, Fig 5.12, Fig 5.13, Fig 5.14). Together with the findings in chapter 4, these results suggest that germline by somatic interaction might be more prevalent, impacting on both CRC risk and prognosis. With genotype, molecular and clinical data from the QUASAR2 clinical trial cohort, the current study represent the first attempt to define such germline-somatic interactions. Furthermore, these associations have the potential

to aid in identifying a subset of patients that tend to perform worse in the setting of adjuvant chemotherapy. Additional medical intervention might be recommended for these individuals.

The candidate prognostic variant rs184423256 is an intronic variant in *RRM2B*, which encodes an essential enzyme involved in DNA synthesis. RRM2B is a member of the p53 pathway and the fluorouracil and platinum metabolism pathway (KEGG database; [369]. RRM2B expression is induced by 5-fluorouracil treatment in colon cancer cells, and its expression is correlated with cell cycle arrest and apoptosis [370]. RRM2B expression was associated with better prognosis in CRC, especially among patients with advanced disease [133]. It is plausible that the association observed of the variant rs184423256 and clinical outcomes in early-stage CRC might be mediated through RRM2B expression regulation. Currently, this variant is not associated with RRM2B expression in normal or cancerous tissues. Therefore, additional analysis is required to delineate the biological basis for the association.

This study did not support a prognostic role of known CRC risk loci, nor did it validate variants previously suggested to have prognostic value in previous genome-wide association studies in patient cohorts of advanced disease or mixed stages. This is in line with null findings from the SOCCS (n = 3,886; [307, 308]) and the NCCTG N0147 (n = 4,319; [309]) clinical trial cohorts, suggesting that the reported associations could be stage-specific or spurious findings.

Inherited genetic variants in the HLA genes could impact antigen and neoantigen presentation in cancer cells, thus possibly play a role in the interaction between tumour and adaptive immune cells. In turn, this could influence disease progression and cancer prognosis. Utilising data from the QUASAR2 and SCOT clinical trial cohorts, I also investigated the prognostic role of HLA genotypes on clinical outcomes in early-stage CRC. In contrast to the findings that HLA heterozygosity and certain HLA class I gene supertypes were associated with favourable prognosis in cohorts of advanced melanoma patients undergoing immunotherapy [352], the current study showed that HLA homozygosity and HLA class I gene supertypes were not associated with either TTR or OS in cohorts of early-stage CRC patients undergoing adjuvant chemotherapy (Fig 5.15, Fig 5.16, Fig 5.18).

The reason for this discrepancy can be two-fold. First, the prognostic role of HLA genotypes could be more pronounced during immunotherapies, where the interaction between tumour and infiltrating immune cells tends to be heightened, whereas the adaptive immune response could be suppressed otherwise, as commonly seen in CRC [352, 371]. Similar investigation among CRC patients undergoing immunotherapy could be more fruitful. Second, the prognostic role of HLA genotypes could be specific to advanced disease. A study showed that the prognostic value of adaptive immune response is the strongest in high-risk CRC cases but absent in low-risk cases, based on tumour and nodal stages [194]. This suggests that the effect of HLA in CRC prognosis may vary by disease stage. If true, it will be of interest to evaluate HLA genotypes in cohorts of advanced CRC.

Studies suggest that CRC with DNA mismatch repair deficiency (MMRd) tend to show better adaptive immune response in the TME [372], and MMRd phenotype was associated with favourable clinical outcomes [373]. HLA genotypes might play a more prominent role in this subset of patients. Considering that about 16% of CRC cases are MMRd, it requires additional samples for adequate statistical power for discovery. On the other hand, HLA loss has been suggested as a mechanism for immune evasion in lung cancer [351]. This can be a direction for further investigation.

In this chapter, I examined the association between inherited genetic variants and infiltrating CD8+ T cell density in CRC, which was measured by *in situ* immunohistochemistry staining on TMA samples of the QUASAR2 cohort. I identified one locus at genome-wide significance and 21 suggestive loci. In particular, the minor allele of the top variant, rs78959096 (at 14q23.1), was associates with 2.6-fold decrease in infiltrating CD8+ T cell density in CRC (Fig 5.19C). This variant is located in an intergenic region near the gene *DACT1*, which encodes a protein involved in WNT and GPCR signalling pathways. Activated WNT signalling was associated with T cell exclusion in CRC [374]. It is possible that this variant modifies nearby gene expression to facilitate T cell exclusion in the TME; however further investigation will be required to confirm or refute this possibility.

Using these 22 iQTLs for tumour-infiltrating CD8+ T cell density as the ge-

netic instrument, I studied the causal relationship between the tumour-infiltrating CD8+ T cell density and CRC outcomes in the QUASAR2 cohort using two-stage Mendelian randomisation. I detected a marginal causal association between the immune trait and OS, but I found no significant causal association between the immune trait and TTR. Moreover, mutual information and related measures did not support the presence of non-linear relationship between the immune trait and clinical outcomes in CRC. Collectively, this suite of statistical analyses constitutes an initial study of the role of common genetic variants on prognosis and adaptive immune trait in CRC. Together, a nonlinear relationship between the immune trait and CRC clinical outcomes remains unproven. Analysis in an additional cohort would help elucidate the causal relationship between adaptive immune response and CRC prognosis.

Chapter 6

Discussion

6.1 Summary

The last two decades have seen extensive efforts in cataloguing human genetic variation correlated with phenotypic differences in human populations (see also section 1.1.1). Most common inherited genetic variants have been assessed in GWASs for statistical associations with susceptibility to common cancers. However, only a limited fraction of the genetic component of cancer susceptibility has been identified, leaving the bulk of the genetic heritability unexplained (see also section 1.1.4). Moreover, it remains a challenge to elucidate the functional link between associated variants and cancer risk (see also section 1.1.5). Recent studies showed that inherited genetic variants were associated with somatic driver events and some could also influence tumour development (see also section 1.3). This poses the question whether inherited genetic variants interact with somatic driver events to modify cancer risk. If true, inherited cancer risk variants could be specific to cancers of certain molecular profiles but not others. Therefore, an investigation into germline by somatic interaction in cancer could potentially explain part of the missing heritability and help us better understand the influence of inherited genetic variants on risk, biology and prognosis of common cancers.

In this thesis, I analysed multi-cancer comprehensive datasets to investigate germline by somatic interaction in cancer and to evaluate the role of inherited

genetic variants on cancer risk, response to anti-cancer therapy, nearby gene expression, adaptive immune response, and prognosis. These analyses were aimed at defining inherited genetic variants associated with differential cancer risk by molecular subtype and somatic driver mutations, thus providing novel insights into the genetic basis of cancer susceptibility as well as practical implications for personalised disease prediction.

In chapter 3, I identified 110 inherited cancer risk variants that were associated with subtype-specific risk in breast and ovarian cancers. The pattern of associations was inversely correlated with p53 mutation frequency in individual cancer subtypes. One of these variants, the p53 poly(A) SNP rs78378222, directly tracked with p53 mutational status in tumours. It regulated p53 expression levels in both normal and cancerous tissues, regardless of p53 mutational status. Together, I found that this variant interacts with p53 mutational status in tumours to influence cancer risk, response to therapy and prognosis.

I extended the investigation on germline by somatic interaction from the p53 tumour suppressor pathway to the RAS-MAPK oncogenic pathway in chapter 4. I observed that a known CRC risk locus on chromosome 10p14 was associated specifically with risk for CRC with *KRAS* hotspot mutations. In particular, rs10795668 at this locus was associated with *KRAS* hotspot mutations across tumour types. It was also associated with *ATP5C1* expression levels in tumours, but contingent on KRAS mutational status. These findings suggest that germline by somatic interaction is not exclusive to the p53 pathway, inherited genetic variants also interact with somatic driver mutations in the RAS-MAPK pathway to associate with risk, nearby gene expression and potentially prognosis in CRC.

Chapter 5 was aimed at evaluating the role of inherited genetic variants on prognosis and adaptive immune response broadly in early-stage CRC. I identified two loci at 15q26.2 (rs72767781; *NR2F2-AS1*) and at 15q22.2 (rs11447843; *VPS13C*) associated with time to recurrence (TTR) and overall survival (OS) at genome-wide significance, respectively. In addition, I also identified a number of loci associated with both TTR and OS at suggestive significance. Interestingly, I observed that rs184423256 (8q22.3; *RRM2B*) and rs114887409 (4p16.1; *ABLIM2*) interact with

KRAS mutational status in tumours to associate with clinical outcomes. On the other hand, the current study did not support the prognostic role of known colorectal cancer risk variants, nor did it validate variants previously associated with colorectal cancer prognosis. This suggests that prognostic variants could be stage-specific, or that previous findings could be spurious due to limited statistical power.

The study on inherited genetic variants in the HLA region showed that HLA homozygosity and HLA class I supertypes were not associated with TTR or OS in early-stage CRC undergoing adjuvant chemotherapy, unlike the report on advanced melanoma undergoing immunotherapy [352]. This suggests that the effect of HLA genotype might vary by treatment type or disease stage. In the mapping of immune quantitative trait loci (iQTLs), I found that a locus on chromosome 14q23.1 (rs78959096) was associated with infiltrating CD8+ T cell density in CRC at genome-wide significance, together with 21 loci reaching suggestive significance. Using the 22 iQTLs to construct a genetic instrument, I also studied the causal relationship between tumour infiltrating CD8+ T cell density and clinical outcomes in CRC. I detected a marginal causal association between tumour infiltrating CD8+ T cell density and OS, but I found no significant causal association between tumour infiltrating CD8+ T cell density and TTR. Mutual information and related measure did not support the presence of nonlinear relationship between tumour infiltrating CD8+ T cell density and clinical outcomes in CRC.

The contribution of the work presented in this thesis is two fold. One is the integrated analysis of germline by somatic interaction in cancer, centred on two key pathways in cancer: the p53 pathway and the RAS-MAPK pathway. The other aspect of significance is that such a systematic evaluation of inherited genetic risk variants on risk, prognosis and adaptive immune response in common cancers have practical implications for personalised cancer risk stratification and prognosis prediction. These two aspects are discussed further in the following sections.

6.2 Germline by somatic interaction in common cancers

Standard genome-wide association studies on cancer risk did not distinguish the molecular features in tumours, therefore unable to discern genetic associations with cancers of specific molecular subtype or somatic driver mutations. A number of studies on breast cancer reported common inherited genetic variants that were associated with risk for specific cancer subtypes [30, 51, 120, 122], but a biological explanation for the observed subtype heterogeneity in cancer risk association was not provided. Studies also demonstrated that inherited genetic variants were associated with specific somatic driver mutation in multiple cancers [133, 135, 136, 138–140], but the relationship between the variants and cancer risk by somatic mutational status was not evaluated.

In this study, I leveraged germline and somatic information in large datasets on breast, ovarian and colorectal cancers to conduct an integrated analysis of germline by somatic interaction in cancer, to assess how the interaction could influence cancer risk, response to therapy, nearby gene expression and prognosis (chapter 3 and 4). By studying the biological functions of the variants, rs78378222 (17p13.1; *TP53*) and rs10795668 (10p14; *ATP5C1*), and their potential impact on the p53 and RAS-MAPK pathways, respectively, I was also able to provide plausible biological explanations for the observed statistical relationships (Fig 3.12, Fig 4.11). Both are examples of how an inherited genetic variants influence nearby gene expression to modulate pathway activity, to modify risk to develop tumours with specific mutational profile and to influence disease prognosis. Carriers of the minor allele of rs78378222 have lower p53 expression levels independent of p53 mutational status in tumours (Fig 3.6 and Fig 3.7); while the minor allele of rs10795668 was associated with increased *ATP5C1* gene expression only in KRAS MUT tumours (Fig 4.10). Although the biological mechanism of how the two variants interact with somatic driver mutations in tumours differ, they illustrated a common theme: inherited genetic variants can play an active role during tumour development. This germline by somatic interaction can be highly informative for understanding cancer risk and prognosis.

In the genome-wide survival analysis on early-stage CRC, I identified two variants, rs184423256 (8q22.3; *RRM2B*) and rs114887409 (4p16.1; *ABLIM2*), that interact with KRAS mutational status in tumours to associate with clinical outcomes (see section 5.2.3.5). Despite that the biological function of these two variants remains unclear, the statistical relationships of inherited genotype, somatic driver mutations and cancer prognosis suggest that the scope of germline by somatic interaction in cancer might be bigger than we expected. If true, this concept might also apply to other cancer hallmark pathways and be of great utility to uncover additional inherited genetic variants associated with cancer risk or prognosis.

6.3 Implications for personalised disease prediction

As discussed in section 1.1.5, inherited genetic variants associated with cancer can offer insight into the polygenic nature of cancer ethology, but also hold great promise for personalised risk prediction. Genotype information of cancer risk variants in an individual can be combined to produce a genetic risk score (PRS), which is predictive of relative cancer risk compared to the population average [59] and can be used to enhance standard cancer screening strategy [61]. A recent study demonstrated that cancer-specific PRS can be integrated with family history and modifiable risk factors to improve cancer risk prediction for common cancers [60].

Existing schemes for constructing cancer-specific PRSs rely on the genotype-risk association reported by standard genome-wide association studies, which did not account for the molecular features in tumours. Exemplified by the two variants, rs78378222 (17p13.1; *TP53*) and rs10795668 (10p14; *ATP5C1*), the current study demonstrated that the effect size of genotype-risk association can depend on the mutational status of cancer driver genes in tumours. This suggests that both germline and somatic information might be necessary for accurate risk prediction. Moreover, I showed that when cancer cases were stratified by molecular subtypes or somatic mutational status in tumours, the estimated effect sizes tend to be larger than those obtained in standard association studies where cases were mixed (see

also section 3.2.1 and 4.2.3). Therefore, stratified association study might represent a powerful method to refine genotype-risk association, which would benefit accurate PRS construction and personalised risk stratification.

Besides personalised risk stratification, the current study also showed that inherited genetic variants can be predictive of response to anti-cancer therapy, cancer prognosis and adaptive immune trait in tumours. These aspects are much less explored in cancer. It would be interesting to evaluate how the multifaceted information could be integrated into the existing schemes of cancer risk stratification to better inform clinical management.

6.4 Future perspectives

In this study, I presented evidence for germline by somatic interaction in cancer on risk, prognosis, nearby gene expression and response to anti-cancer therapy. It would of wide scientific interest to examine such interactions in key cancer pathways other than the p53 and RAS-MAPK pathways. As mentioned above, I developed biological hypotheses to explain the observed statistical relationships about rs78378222 (17p13.1; *TP53*) and rs10795668 (10p14; *ATP5C1*) (Fig 3.12, Fig 4.12). It would be valuable to test these hypotheses in experimental models and study germline by somatic interaction in cancer further.

In this thesis, I also broadly evaluated the influence of inherited genetic variants on cancer prognosis and adaptive immune response in tumours. Multiple associational findings in the current would benefit from validation in additional datasets, and further investigation is warranted to understand the biological functions of the prognostic variants and iQTLs. As above, integrating information of inherited genetic variants, somatic driver mutations and adaptive immune response to model cancer risk and prognosis could represent an emerging research direction to better inform personalised disease prediction and precision medicine.

Chapter 7

Appendix

Cox model	CRC recurrence		All-cause death	
	Causal Estimate[b]	*P*-value	Causal Estimate[b]	*P*-value
Univariate	HR=0.90 (0.76-1.07)	0.23	HR=0.82 (0.68-0.98)	0.032
Multivariate[a]	HR=0.92 (0.78-1.09)	0.33	HR=0.81 (0.67-0.97)	0.023

Table 7.1. Main results of two-stage Mendelian randomisation analysis using individual-level data in the QUASAR2 clinical trial cohort when four individuals with outlier GRSs (< 0.7) were removed.

[a] *Multivariate Cox regression model adjusted by age, sex, tumour location (rectum vs. colon), pT stage (p4 vs. p1-3), nodal stage (N1/2 vs. N0), and treatment (see Methods, section 2.3.8).* [b] *The causal estimate of exposure on survival outcome refers to change in hazard ratio per unit increase in log2-transformed tumour-infiltrating CD8+ T cell density.*

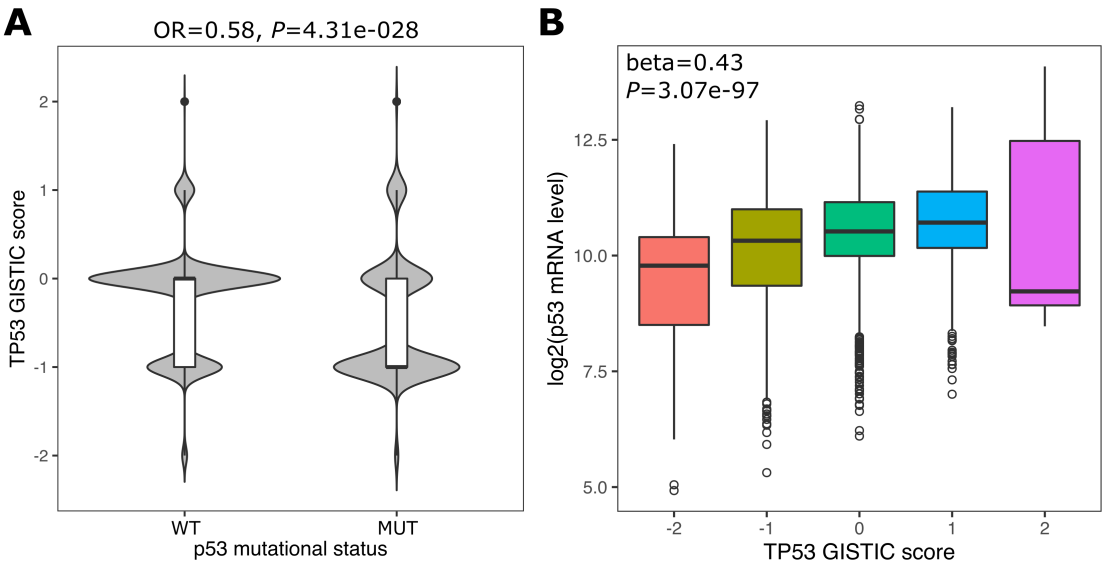

Figure 7.1. CN loss in TP53 is correlated with p53 driver mutations and lower p53 expression levels in tumours.

(A) A violin plot of TP53 GISTIC score against p53 mutational status in the TCGA pan-cancer cohort. OR and P-value were calculated in a logistic regression model, adjusted for tumour types. (B) A box plot showing the relationship between TP53 GISTIC score and p53 expression levels in the TCGA pan-cancer or CRC cohort. Beta coefficient and P-value were calculated by linear regression, adjusted for tumour types.

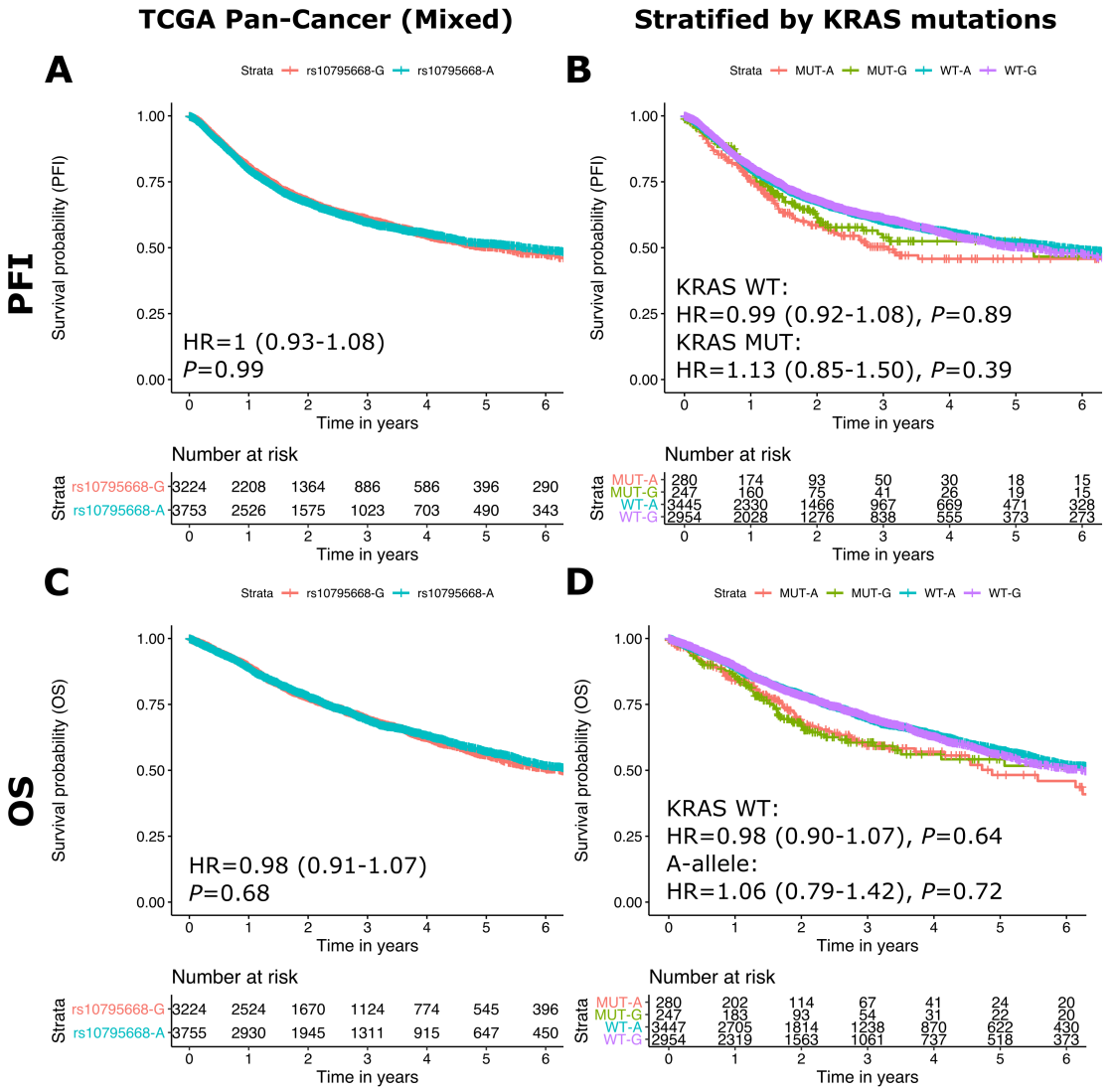

Figure 7.2. Evaluating the relationship between rs10795668 genotype and patient survival outcomes in the TCGA pan-cancer cohort.

(A, C) Kaplan-Meier curves for rs10795668 genotype (assuming a dominant model for the minor allele A) against progression-free interval (PFI) (A) and overall survival (OS) (C), where tumours were not stratified by KRAS mutational status. (B, D) Kaplan-Meier curves for rs10795668 genotype (assuming a dominant model for the minor allele A) against PFI (B) and OS (D), where tumours were stratified by KRAS mutational status. The relationship between genotype and patient survival outcomes was evaluated in cases with KRAS WT or KRAS MUT separately. The hazard ratio (HR) and the 95% CI were estimated from multivariate Cox regression, adjusted for age at diagnosis and gender; the nominal P-value was calculated by log-rank test.

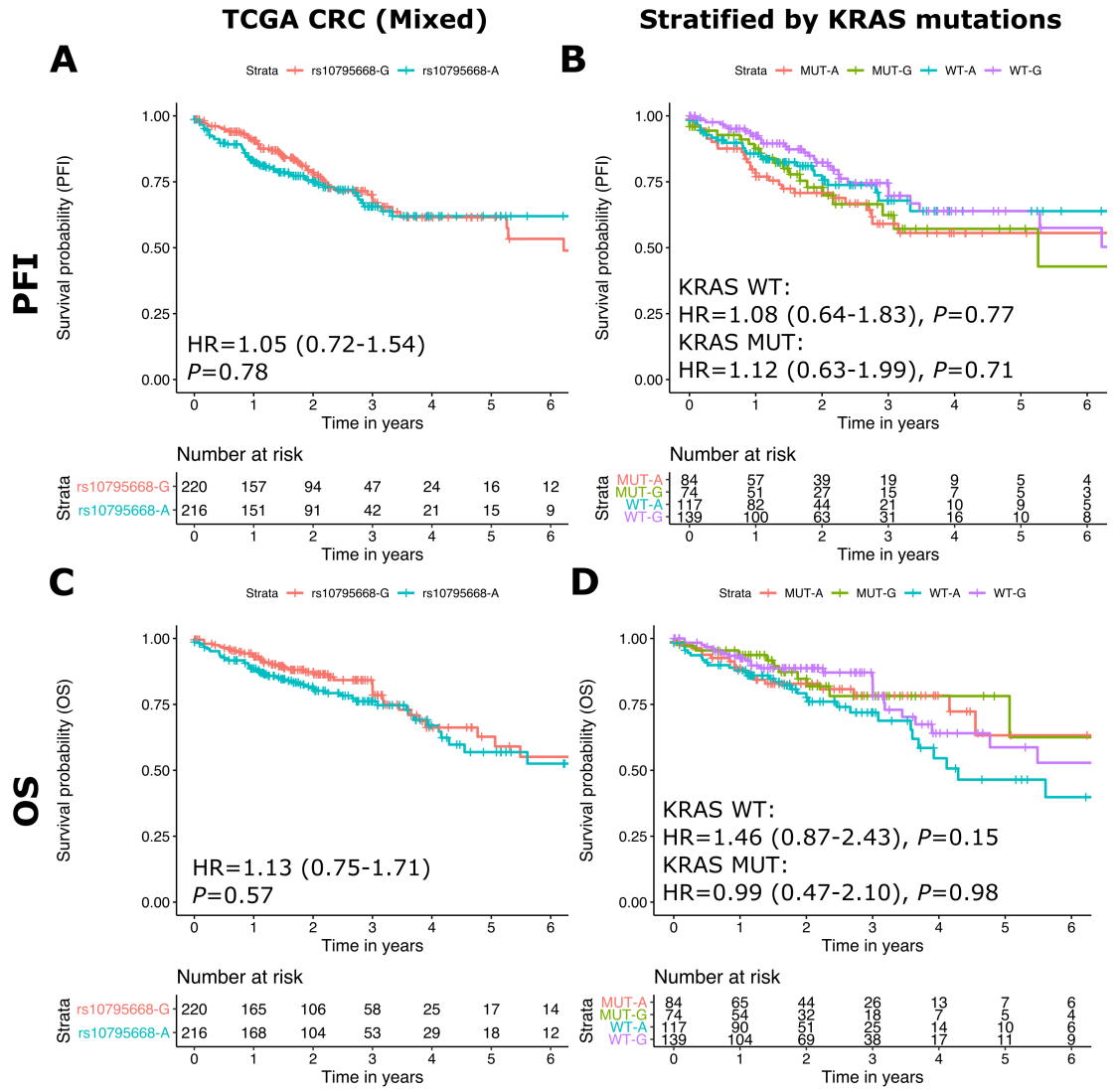

Figure 7.3. Evaluating the relationship between rs10795668 genotype and patient survival outcomes in the TCGA CRC cohort.

(A, C) Kaplan-Meier curves for rs10795668 genotype (assuming a dominant model for the minor allele A) against progression-free interval (PFI) (A) and overall survival (OS) (C), where tumours were not stratified by KRAS mutational status. (B, D) Kaplan-Meier curves for rs10795668 genotype (assuming a dominant model for the minor allele A) against PFI (B) and OS (D), where tumours were stratified by KRAS mutational status. The relationship between genotype and patient survival outcomes was evaluated in cases with KRAS WT or KRAS MUT separately. The hazard ratio (HR) and the 95% CI were estimated from multivariate Cox regression, adjusted for age at diagnosis and gender; the nominal P-value was calculated by log-rank test.

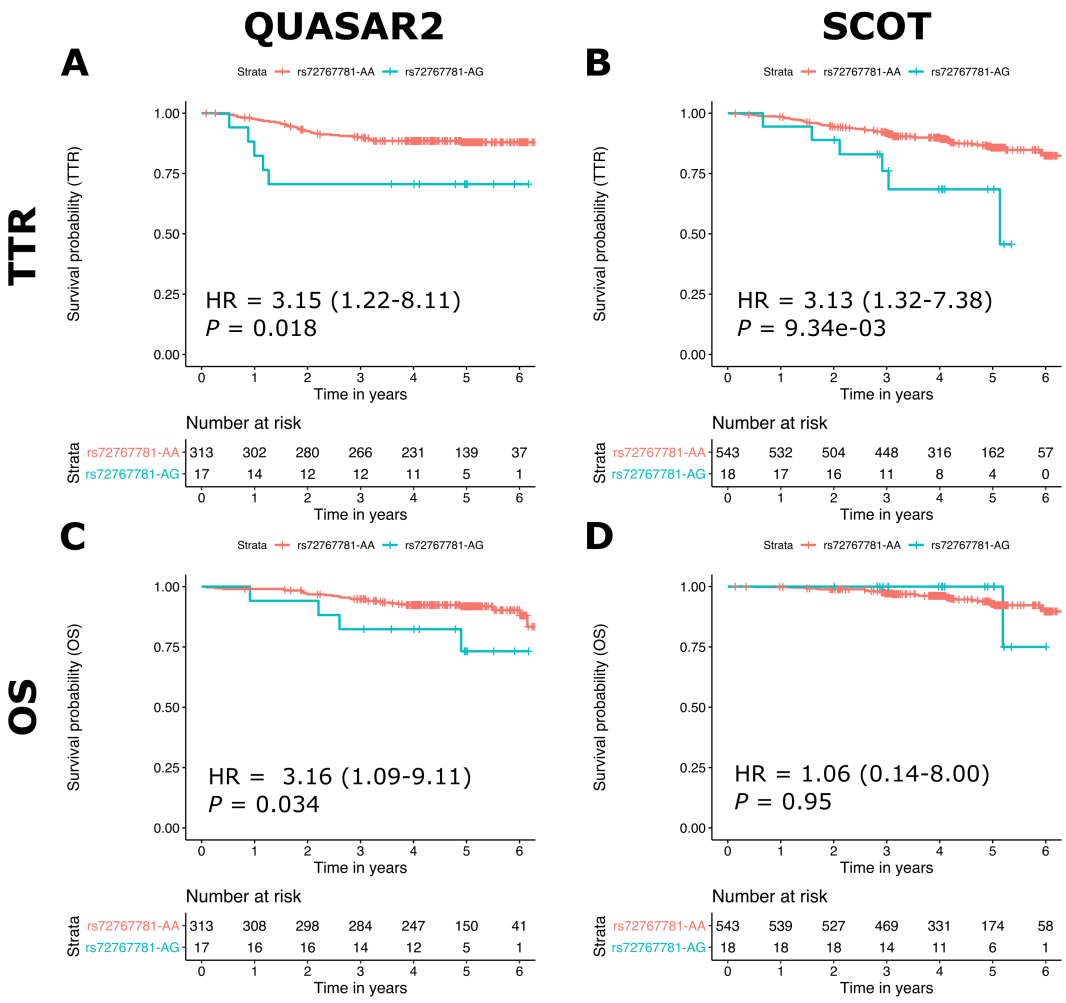

Figure 7.4. The relationship between rs72767781 (15q26.2, MAF = 0.03) and clinical outcomes among patient with high-risk stage II CRC in QUASAR2 (A, C) and SCOT (B, D).

(A, B) For TTR. (C, D) For OS. The hazard ratio (HR) and the 95% CI were estimated from multivariate Cox regression, adjusted for known confounders; the nominal P-value was calculated by log-rank test (see Methods, section 2.3.8).

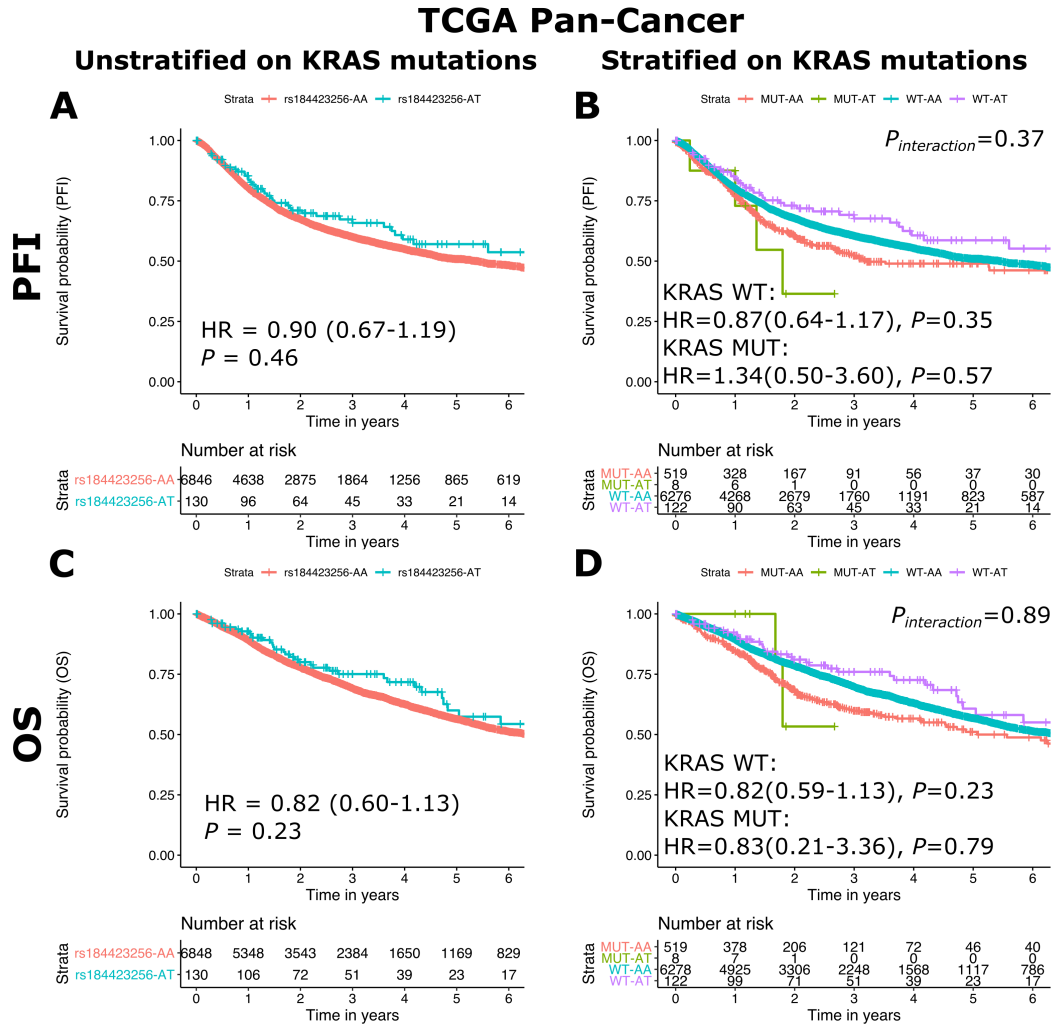

Figure 7.5. The relationship between rs184423256 (8q22.3, MAF = 0.009) and patient survival outcomes in the TCGA pan-cancer cohort by KRAS mutational status.

(A, B) Association of variant genotype with progression-free interval (PFI) when patients were unstratified (A) or when patients were stratified by KRAS mutational status (B). (C, D) Association of variant genotype with OS when patients were unstratified (C) or when patients were stratified by KRAS mutational status (D). The hazard ratio (HR) and the 95% confidence interval were estimated from multivariate Cox regression, adjusted for age and sex; the nominal P-value was calculated by log-rank test.

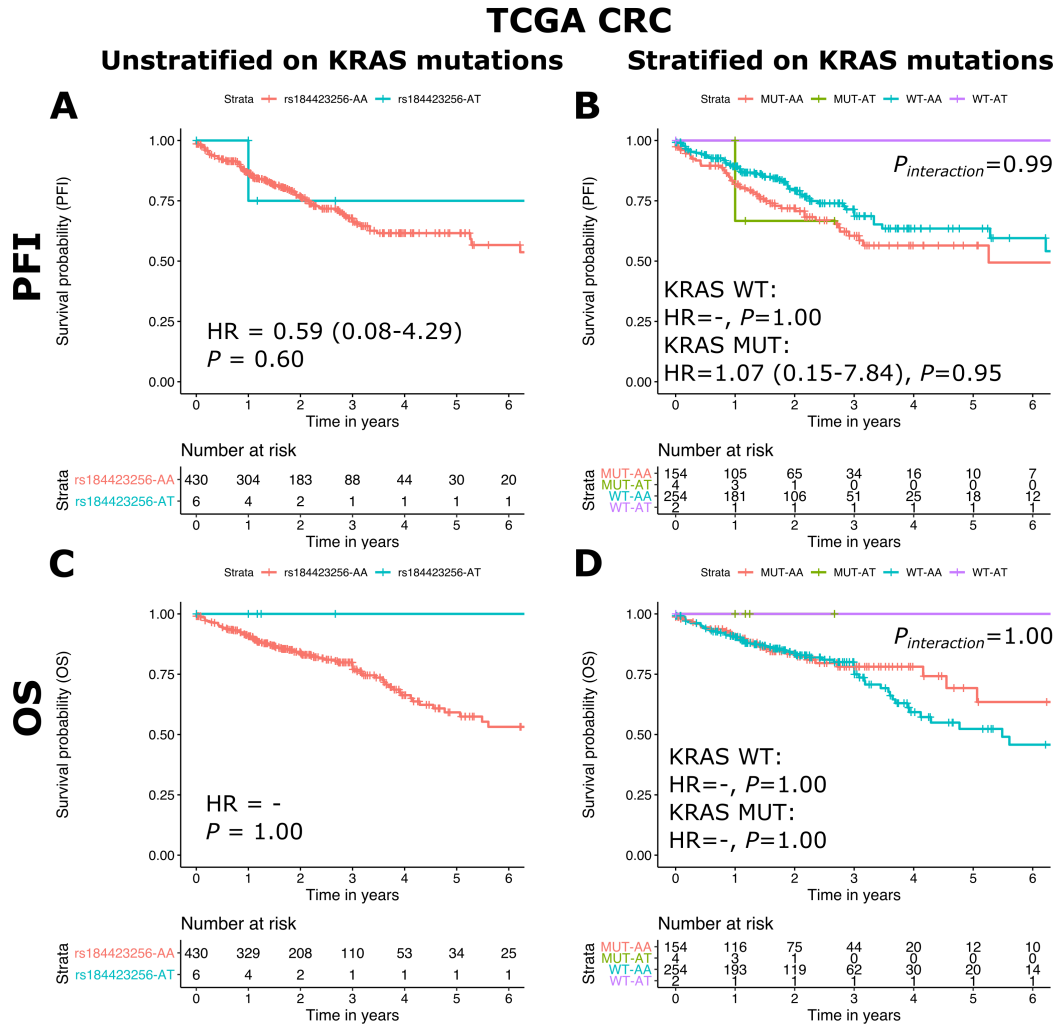

Figure 7.6. The relationship between rs184423256 (8q22.3, MAF = 0.009) and patient survival outcomes in the TCGA CRC cohort by KRAS mutational status.

(A, B) Association of variant genotype with progression-free interval (PFI) when patients were unstratified (A) or when patients were stratified by KRAS mutational status (B). (C, D) Association of variant genotype with OS when patients were unstratified (C) or when patients were stratified by KRAS mutational status (D). The hazard ratio (HR) and the 95% confidence interval were estimated from multivariate Cox regression whenever possible, adjusted for age and sex; the nominal P-value was calculated by log-rank test.

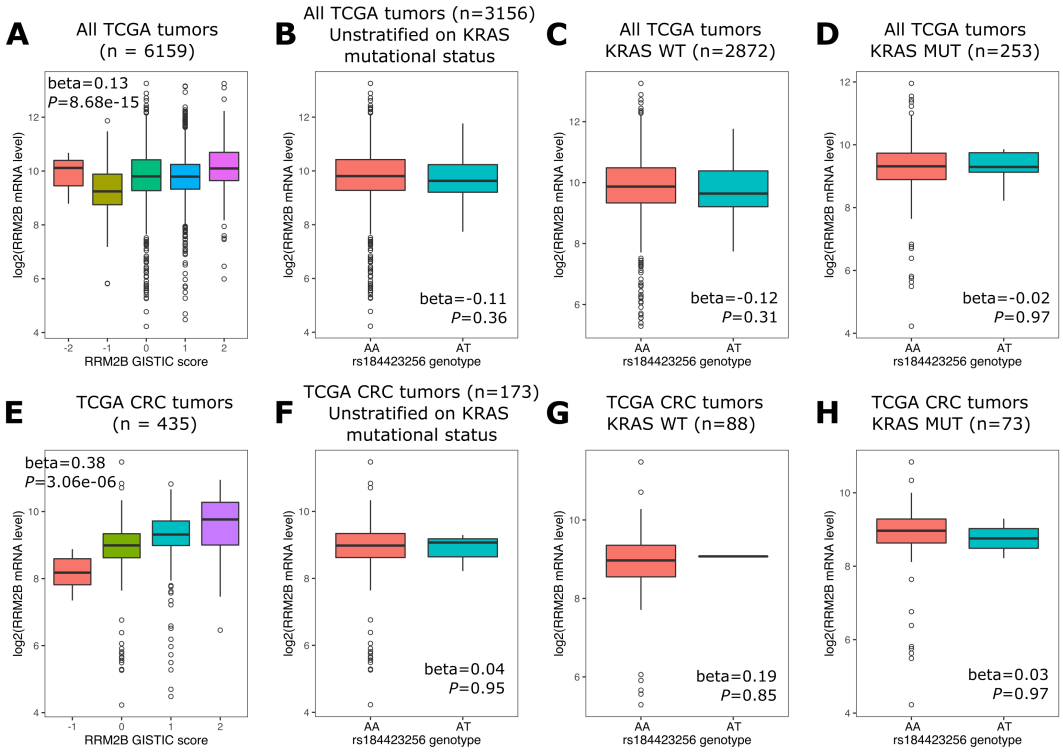

Figure 7.7. The relationship of rs184423256 genotype, KRAS mutational status and RRM2B gene expression in TCGA tumours.

(A, E) Box plots show the correlation between RRM2B GISTIC score and RRM2B mRNA levels in TCGA pan-cancer tumours (A) or TCGA CRC tumours (E). (B-D) Box plots show the correlation between rs184423256 genotype and RRM2B mRNA levels in all TCGA pan-cancer tumours (B), tumours with KRAS WT (C), and tumours with KRAS MUT (D). (F-H) Box plots show the correlation between rs184423256 genotype and RRM2B mRNA levels in all TCGA CRC tumours (F), CRC tumours with KRAS WT (G), and CRC tumours with KRAS MUT (H). Only samples with GISTIC score 0 (i.e., diploid in RRM2B) are included in the analyses.

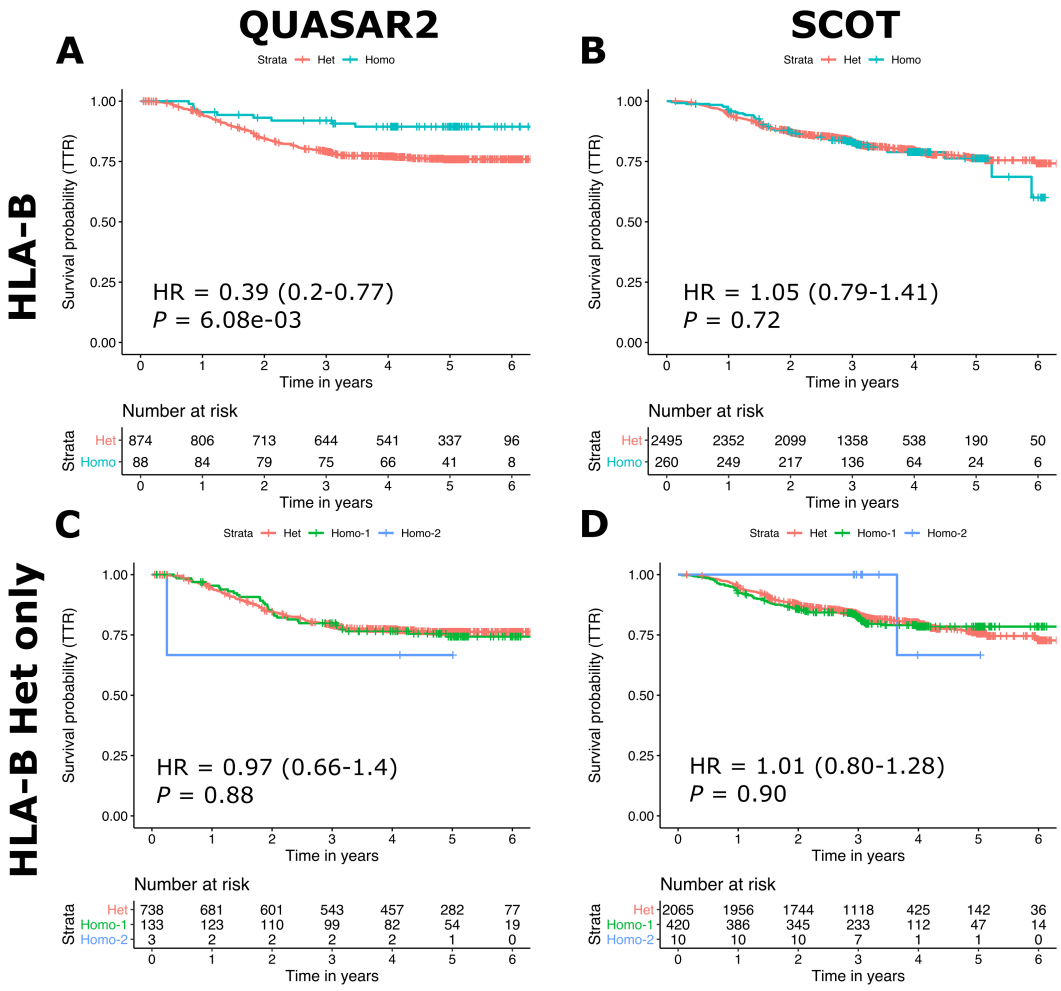

Figure 7.8. The relationship between homozygosity in HLA class I genes (HLA-A, -B, and -C) on TTR in QUASAR2 (A, C) and SCOT (B, D).

(A, B) Association between homozygosity in HLA-B gene and TTR. (C, D) Association between homozygosity in HLA-A and HLA-C and TTR among those with heterozygote alleles in HLA-B. The hazard ratio (HR) and the 95% CI were estimated from multivariate Cox regression, adjusted for known confounders; the nominal P-value was calculated by log-rank test (see Methods, section 2.3.8).

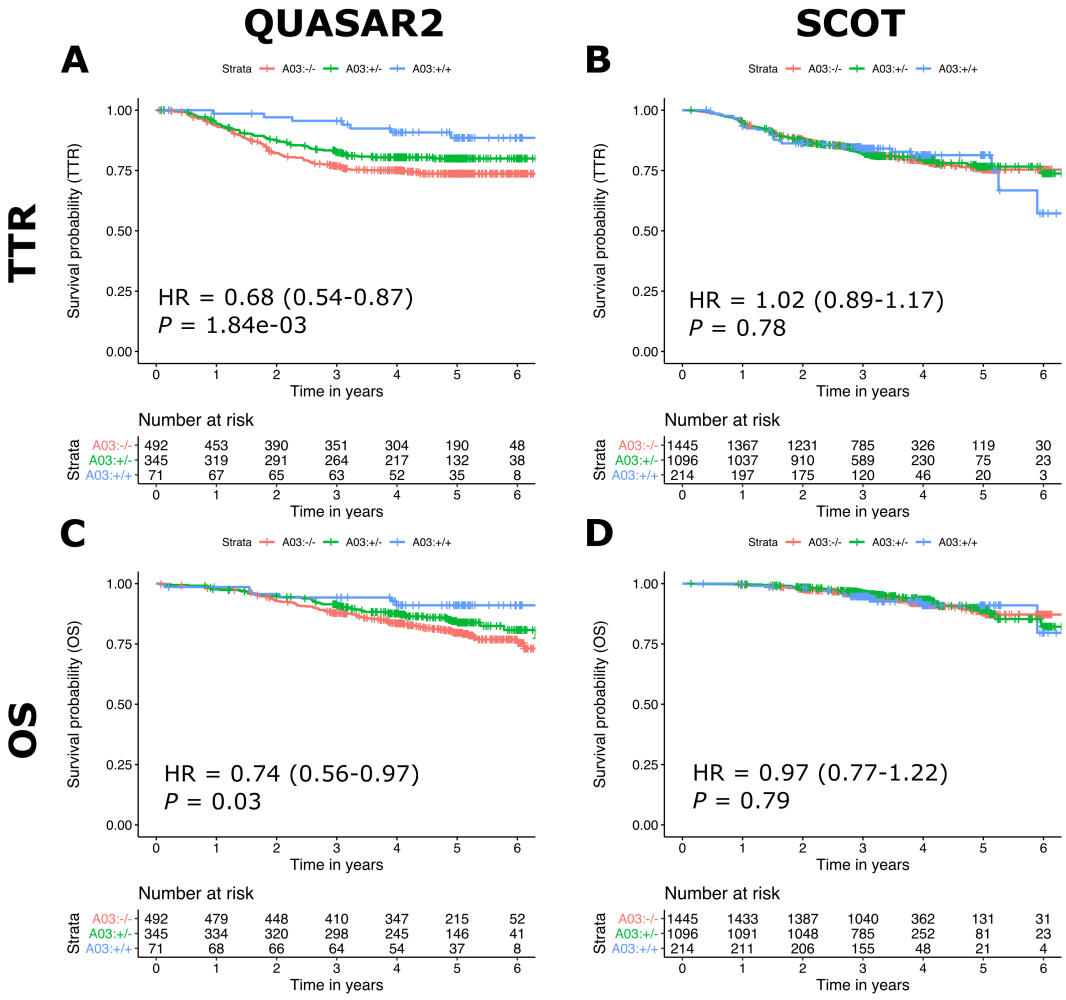

Figure 7.9. The relationship between HLA-A supertype A03 with patient survival outcomes in QUASAR2 (A, C) and SCOT (B, D).

(A, B) For TTR. (C, D) For OS. The hazard ratio (HR) and the 95% CI were estimated from multivariate Cox regression, adjusted for known confounders; the nominal P-value was calculated by log-rank test (see Methods, section 2.3.8).

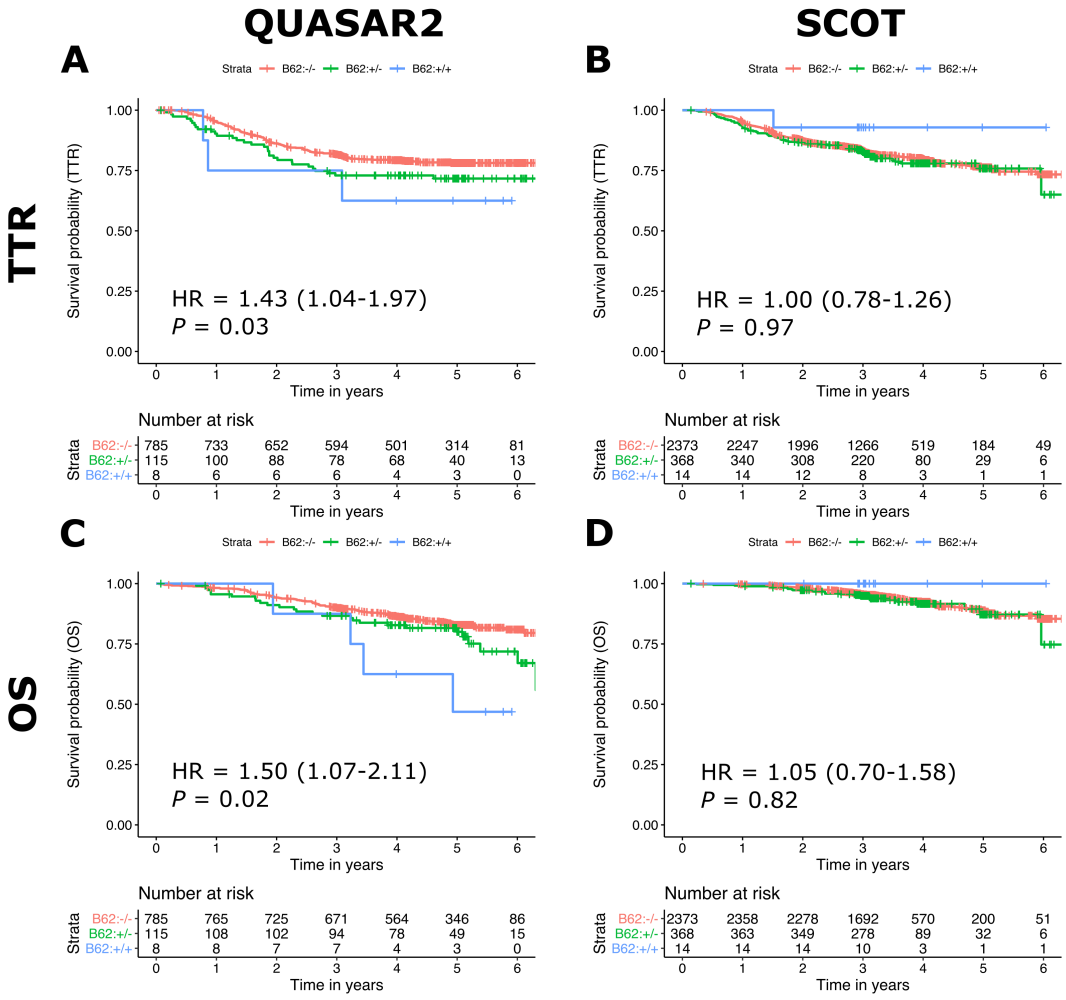

Figure 7.10. The relationship between HLA-B supertype B62 with patient survival outcomes in QUASAR2 (A, C) and SCOT (B, D).

(A, B) For TTR. (C, D) For OS. The hazard ratio (HR) and the 95% CI were estimated from multivariate Cox regression, adjusted for known confounders; the nominal P-value was calculated by log-rank test (see Methods, section 2.3.8).

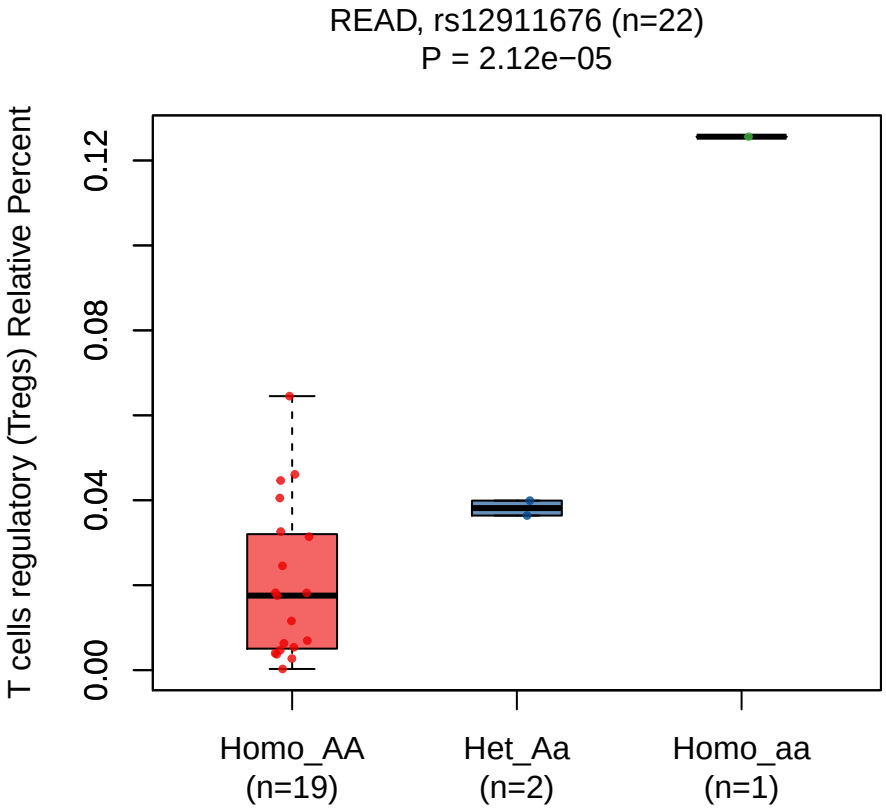

Figure 7.11. Box plot of Tregs relative percentage against rs12911676 genotype in rectal cancer cases in TCGA.

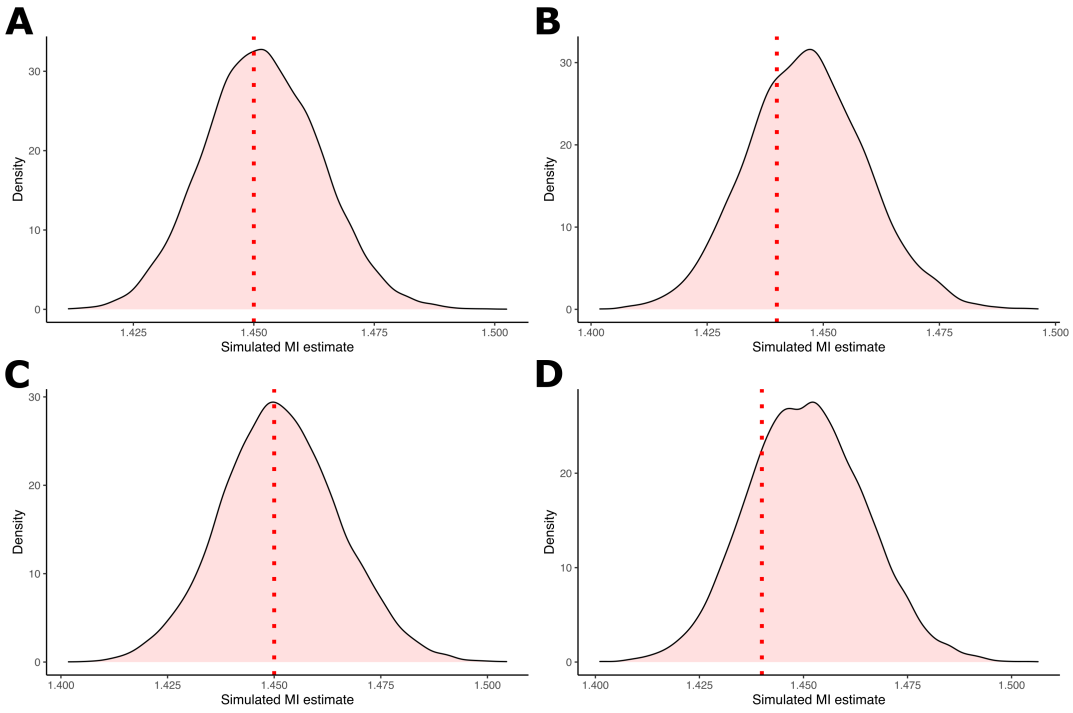

Figure 7.12. Sensitivity of mutual information estimate between tumour infiltrating CD8+ T cell density and time to recurrence (A, C) or all-cause death (B, D) due to right-censoring.

(A, B) Simulated distribution of mutual information estimate assuming a random delay between 0 and 100 days among right-censored cases. (C, D) Simulated distribution of mutual information estimate assuming a random delay between 0 and 500 days among right-censored cases. Red dotted lines indicate the mutual information estimate for each variable.

References

[1] Kelly A. Frazer, Sarah S. Murray, Nicholas J. Schork, and Eric J. Topol. Human genetic variation and its contribution to complex traits. *Nature Reviews Genetics*, 10(4):241–251, April 2009. ISSN 1471-0056, 1471-0064. doi: 10.1038/nrg2554. URL http://www.nature.com/articles/nrg2554.

[2] 1000 Genomes Project Consortium, Adam Auton, Lisa D. Brooks, Richard M. Durbin, Erik P. Garrison, Hyun Min Kang, Jan O. Korbel, Jonathan L. Marchini, Shane McCarthy, Gil A. McVean, and Gonçalo R. Abecasis. A global reference for human genetic variation. *Nature*, 526(7571):68–74, October 2015. ISSN 1476-4687. doi: 10.1038/nature15393.

[3] Aravinda Chakravarti. Genomic contributions to Mendelian disease. *Genome Research*, 21(5):643–644, May 2011. ISSN 1549-5469. doi: 10.1101/gr.123554.111.

[4] Lorelei A. Mucci, Jacob B. Hjelmborg, Jennifer R. Harris, Kamila Czene, David J. Havelick, Thomas Scheike, Rebecca E. Graff, Klaus Holst, Sören Möller, Robert H. Unger, Christina McIntosh, Elizabeth Nuttall, Ingunn Brandt, Kathryn L. Penney, Mikael Hartman, Peter Kraft, Giovanni Parmigiani, Kaare Christensen, Markku Koskenvuo, Niels V. Holm, Kauko Heikkilä, Eero Pukkala, Axel Skytthe, Hans-Olov Adami, Jaakko Kaprio, and for the Nordic Twin Study of Cancer (NorTwinCan) Collaboration. Familial Risk and Heritability of Cancer Among Twins in Nordic Countries. *JAMA*, 315(1):68, January 2016. ISSN 0098-7484. doi: 10.1001/jama.2015.17703. URL http://jama.jamanetwork.com/article.aspx?doi=10.1001/jama.2015.17703.

[5] N. Risch and K. Merikangas. The future of genetic studies of complex human diseases. *Science (New York, N.Y.)*, 273(5281):1516–1517, September 1996. ISSN 0036-8075. doi: 10.1126/science.273.5281.1516.

[6] H. A. Risch, J. R. McLaughlin, D. E. Cole, B. Rosen, L. Bradley, E. Kwan, E. Jack, D. J. Vesprini, G. Kuperstein, J. L. Abrahamson, I. Fan, B. Wong, and S. A. Narod. Prevalence and penetrance of germline BRCA1 and BRCA2 mutations in a population series of 649 women with ovarian cancer. *American Journal of Human Genetics*, 68(3):700–710, March 2001. ISSN 0002-9297. doi: 10.1086/318787.

[7] Allan Balmain. Cancer as a complex genetic trait: tumor susceptibility in humans and mouse models. *Cell*, 108(2):145–152, January 2002. ISSN 0092-8674. doi: 10.1016/s0092-8674(02)00622-0.

[8] Jose L. Badano and Nicholas Katsanis. Beyond Mendel: an evolving view of human genetic disease transmission. *Nature Reviews Genetics*, 3(10):779–789, October 2002. ISSN 1471-0056, 1471-0064. doi: 10.1038/nrg910. URL http://www.nature.com/articles/nrg910.

[9] Daniel J. M. Crouch and Walter F. Bodmer. Polygenic inheritance, GWAS, polygenic risk scores, and the search for functional variants. *Proceedings of the National Academy of Sciences*, 117(32):18924–18933, August 2020. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.2005634117. URL http://www.pnas.org/lookup/doi/10.1073/pnas.2005634117.

[10] Nicholas J. Timpson, Celia M. T. Greenwood, Nicole Soranzo, Daniel J. Lawson, and J. Brent Richards. Genetic architecture: the shape of the genetic contribution to human traits and disease. *Nature Reviews. Genetics*, 19(2):110–124, February 2018. ISSN 1471-0064. doi: 10.1038/nrg.2017.101.

[11] Melina Claussnitzer, Judy H. Cho, Rory Collins, Nancy J. Cox, Emmanouil T. Dermitzakis, Matthew E. Hurles, Sekar Kathiresan, Eimear E. Kenny, Cecilia M. Lindgren, Daniel G. MacArthur, Kathryn N. North, Sharon E. Plon, Heidi L. Rehm, Neil Risch, Charles N. Rotimi, Jay Shendure, Nicole Soranzo, and Mark I. McCarthy. A brief history of human disease genetics. *Nature*, 577(7789):179–189, January 2020. ISSN 0028-0836, 1476-4687. doi: 10.1038/s41586-019-1879-7. URL http://www.nature.com/articles/s41586-019-1879-7.

[12] Aditya Addepalli, Sakhare Kalyani, Minali Singh, Debashree Bandyopadhyay, and K. Naga Mohan. CalPen (Calculator of Penetrance), a web-based tool to estimate penetrance in complex genetic disorders. *PLOS ONE*, 15(1):e0228156, January 2020. ISSN 1932-6203. doi: 10.1371/journal.pone.0228156. URL https://dx.plos.org/10.1371/journal.pone.0228156.

[13] Nils Rahner and Verena Steinke. Hereditary Cancer Syndromes. *Deutsches Aerzteblatt Online*, October 2008. ISSN 1866-0452. doi: 10.3238/arztebl.2008.0706. URL https://www.aerzteblatt.de/10.3238/arztebl.2008.0706.

[14] Deborah K. Mayer, Larissa Nekhlyudov, Claire F. Snyder, Janette K. Merrill, Dana S. Wollins, and Lawrence N. Shulman. American Society of Clinical Oncology Clinical Expert Statement on Cancer Survivorship Care Planning. *Journal of Oncology Practice*, 10(6):345–351, November 2014. ISSN 1554-7477, 1935-469X. doi: 10.1200/JOP.2014.001321. URL http://ascopubs.org/doi/10.1200/JOP.2014.001321.

[15] A. G. Knudson. Mutation and Cancer: Statistical Study of Retinoblastoma. *Proceedings of the National Academy of Sciences*, 68(4):820–823, April 1971. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.68.4.820. URL http://www.pnas.org/cgi/doi/10.1073/pnas.68.4.820.

[16] Matthew B. Yurgelun, Anu B. Chittenden, Vicente Morales-Oyarvide, Douglas A. Rubinson, Richard F. Dunne, Margaret M. Kozak, Zhi Rong Qian, Marisa W. Welch, Lauren K. Brais, Annacarolina Da Silva, Justin L. Bui, Chen Yuan, Tingting Li, Wanwan Li, Atsuhiro Masuda, Mancang Gu, Andrea J. Bullock, Daniel T. Chang, Thomas E. Clancy, David C. Linehan, Jennifer J. Findeis-Hosey, Leona A. Doyle,

Aaron R. Thorner, Matthew D. Ducar, Bruce M. Wollison, Natalia Khalaf, Kimberly Perez, Sapna Syngal, Andrew J. Aguirre, William C. Hahn, Matthew L. Meyerson, Charles S. Fuchs, Shuji Ogino, Jason L. Hornick, Aram F. Hezel, Albert C. Koong, Jonathan A. Nowak, and Brian M. Wolpin. Germline cancer susceptibility gene variants, somatic second hits, and survival outcomes in patients with resected pancreatic cancer. *Genetics in Medicine: Official Journal of the American College of Medical Genetics*, 21(1):213–223, 2019. ISSN 1530-0366. doi: 10.1038/s41436-018-0009-5.

[17] Nancie Petrucelli, Mary B. Daly, and Tuya Pal. BRCA1- and BRCA2-Associated Hereditary Breast and Ovarian Cancer. In Margaret P. Adam, Holly H. Ardinger, Roberta A. Pagon, Stephanie E. Wallace, Lora JH Bean, Karen Stephens, and Anne Amemiya, editors, *GeneReviews®*. University of Washington, Seattle, Seattle (WA), 1993. URL `http://www.ncbi.nlm.nih.gov/books/NBK1247/`.

[18] Michael P. Lux, Peter A. Fasching, and Matthias W. Beckmann. Hereditary breast and ovarian cancer: review and future perspectives. *Journal of Molecular Medicine (Berlin, Germany)*, 84(1):16–28, January 2006. ISSN 0946-2716. doi: 10.1007/s00109-005-0696-7.

[19] Frank A. Sinicrope. Lynch Syndrome-Associated Colorectal Cancer. *The New England Journal of Medicine*, 379(8):764–773, August 2018. ISSN 1533-4406. doi: 10.1056/NEJMcp1714533.

[20] Eric R. Fearon. Molecular Genetics of Colorectal Cancer. *Annual Review of Pathology: Mechanisms of Disease*, 6(1):479–507, February 2011. ISSN 1553-4006, 1553-4014. doi: 10.1146/annurev-pathol-011110-130235. URL `http://www.annualreviews.org/doi/10.1146/annurev-pathol-011110-130235`.

[21] W. F. Bodmer, C. J. Bailey, J. Bodmer, H. J. Bussey, A. Ellis, P. Gorman, F. C. Lucibello, V. A. Murday, S. H. Rider, and P. Scambler. Localization of the gene for familial adenomatous polyposis on chromosome 5. *Nature*, 328(6131):614–616, August 1987. ISSN 0028-0836. doi: 10.1038/328614a0.

[22] E. Solomon, R. Voss, V. Hall, W. F. Bodmer, J. R. Jass, A. J. Jeffreys, F. C. Lucibello, I. Patel, and S. H. Rider. Chromosome 5 allele loss in human colorectal carcinomas. *Nature*, 328(6131):616–619, August 1987. ISSN 0028-0836. doi: 10.1038/328616a0.

[23] Gaëlle Bougeard, Mariette Renaux-Petel, Jean-Michel Flaman, Camille Charbonnier, Pierre Fermey, Muriel Belotti, Marion Gauthier-Villars, Dominique Stoppa-Lyonnet, Emilie Consolino, Laurence Brugières, Olivier Caron, Patrick R. Benusiglio, Brigitte Bressac-de Paillerets, Valérie Bonadona, Catherine Bonaïti-Pellié, Julie Tinat, Stéphanie Baert-Desurmont, and Thierry Frebourg. Revisiting Li-Fraumeni Syndrome From TP53 Mutation Carriers. *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology*, 33(21):2345–2352, July 2015. ISSN 1527-7755. doi: 10.1200/JCO.2014.59.5728.

[24] The European Reference Network GENTURIS, Thierry Frebourg, Svetlana Bajalica Lagercrantz, Carla Oliveira, Rita Magenheim, and D. Gareth Evans. Guidelines

for the Li–Fraumeni and heritable TP53-related cancer syndromes. *European Journal of Human Genetics*, 28(10):1379–1386, October 2020. ISSN 1018-4813, 1476-5438. doi: 10.1038/s41431-020-0638-4. URL `http://www.nature.com/articles/s41431-020-0638-4`.

[25] Glen W. Barrisford, Eric A. Singer, Inger L. Rosner, W. Marston Linehan, and Gennady Bratslavsky. Familial renal cancer: molecular genetics and surgical management. *International Journal of Surgical Oncology*, 2011:658767, 2011. ISSN 2090-1410. doi: 10.1155/2011/658767.

[26] Robert Pilarski, Randall Burt, Wendy Kohlman, Lana Pho, Kristen M. Shannon, and Elizabeth Swisher. Cowden syndrome and the PTEN hamartoma tumor syndrome: systematic review and revised diagnostic criteria. *Journal of the National Cancer Institute*, 105(21):1607–1616, November 2013. ISSN 1460-2105. doi: 10.1093/jnci/djt277.

[27] Rohini Roy, Jarin Chun, and Simon N. Powell. BRCA1 and BRCA2: different roles in a common pathway of genome protection. *Nature Reviews. Cancer*, 12(1):68–78, December 2011. ISSN 1474-1768. doi: 10.1038/nrc3181.

[28] The Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast tumours. *Nature*, 490(7418):61–70, October 2012. ISSN 0028-0836, 1476-4687. doi: 10.1038/nature11412. URL `http://www.nature.com/articles/nature11412`.

[29] Kathryn P. Pennington, Tom Walsh, Maria I. Harrell, Ming K. Lee, Christopher C. Pennil, Mara H. Rendi, Anne Thornton, Barbara M. Norquist, Silvia Casadei, Alexander S. Nord, Kathy J. Agnew, Colin C. Pritchard, Sheena Scroggins, Rochelle L. Garcia, Mary-Claire King, and Elizabeth M. Swisher. Germline and somatic mutations in homologous recombination genes predict platinum response and survival in ovarian, fallopian tube, and peritoneal carcinomas. *Clinical Cancer Research: An Official Journal of the American Association for Cancer Research*, 20(3): 764–775, February 2014. ISSN 1557-3265. doi: 10.1158/1078-0432.CCR-13-2287.

[30] Breast Cancer Family Registry, EMBRACE, GEMO Study Collaborators, HEBON, kConFab Investigators, Ontario Cancer Genetics Network, SWE-BRCA, CIMBA, Anna Marie Mulligan, Fergus J Couch, Daniel Barrowdale, Susan M Domchek, Diana Eccles, Heli Nevanlinna, Susan J Ramus, Mark Robson, Mark Sherman, Amanda B Spurdle, Barbara Wappenschmidt, Andrew Lee, Lesley McGuffog, Sue Healey, Olga M Sinilnikova, Ramunas Janavicius, Thomas vO Hansen, Finn C Nielsen, Bent Ejlertsen, Ana Osorio, Iván Muñoz-Repeto, Mercedes Durán, Javier Godino, Maroulio Pertesi, Javier Benítez, Paolo Peterlongo, Siranoush Manoukian, Bernard Peissel, Daniela Zaffaroni, Elisa Cattaneo, Bernardo Bonanni, Alessandra Viel, Barbara Pasini, Laura Papi, Laura Ottini, Antonella Savarese, Loris Bernard, Paolo Radice, Ute Hamann, Martijn Verheus, Hanne EJ Meijers-Heijboer, Juul Wijnen, Encarna B Gómez García, Marcel R Nelen, C Marleen Kets, Caroline Seynaeve, Madeleine MA Tilanus-Linthorst, Rob B van der Luijt, Theo van Os, Matti Rookus, Debra Frost, J Louise Jones, D Gareth Evans, Fiona Lalloo, Ros Eeles, Louise Izatt, Julian Adlard, Rosemarie Davidson, Jackie Cook, Alan Donaldson,

Huw Dorkins, Helen Gregory, Jacqueline Eason, Catherine Houghton, Julian Barwell, Lucy E Side, Emma McCann, Alex Murray, Susan Peock, Andrew K Godwin, Rita K Schmutzler, Kerstin Rhiem, Christoph Engel, Alfons Meindl, Ina Ruehl, Norbert Arnold, Dieter Niederacher, Christian Sutter, Helmut Deissler, Dorothea Gadzicki, Karin Kast, Sabine Preisler-Adams, Raymonda Varon-Mateeva, Ines Schoenbuchner, Britta Fiebig, Wolfram Heinritz, Dieter Schäfer, Heidrun Gevensleben, Virginie Caux-Moncoutier, Marion Fassy-Colcombet, François Cornelis, Sylvie Mazoyer, Mélanie Léoné, Nadia Boutry-Kryza, Agnès Hardouin, Pascaline Berthet, Danièle Muller, Jean-Pierre Fricker, Isabelle Mortemousque, Pascal Pujol, Isabelle Coupier, Marine Lebrun, Caroline Kientz, Michel Longy, Nicolas Sevenet, Dominique Stoppa-Lyonnet, Claudine Isaacs, Trinidad Caldes, Miguel de la Hoya, Tuomas Heikkinen, Kristiina Aittomäki, Ignacio Blanco, Conxi Lazaro, Rosa B Barkardottir, Penny Soucy, Martine Dumont, Jacques Simard, Marco Montagna, Silvia Tognazzo, Emma D'Andrea, Stephen Fox, Max Yan, Tim Rebbeck, Olufunmilayo I Olopade, Jeffrey N Weitzel, Henry T Lynch, Patricia A Ganz, Gail E Tomlinson, Xianshu Wang, Zachary Fredericksen, Vernon S Pankratz, Noralane M Lindor, Csilla Szabo, Kenneth Offit, Rita Sakr, Mia Gaudet, Jasmine Bhatia, Noah Kauff, Christian F Singer, Muy-Kheng Tea, Daphne Gschwantler-Kaulich, Anneliese Fink-Retter, Phuong L Mai, Mark H Greene, Evgeny Imyanitov, Frances P O'Malley, Hilmi Ozcelik, Gordon Glendon, Amanda E Toland, Anne-Marie Gerdes, Mads Thomassen, Torben A Kruse, Uffe Birk Jensen, Anne-Bine Skytte, Maria A Caligo, Maria Soller, Karin Henriksson, von Anna Wachenfeldt, Brita Arver, Marie Stenmark-Askmalm, Per Karlsson, Yuan Chun Ding, Susan L Neuhausen, Mary Beattie, Paul DP Pharoah, Kirsten B Moysich, Katherine L Nathanson, Beth Y Karlan, Jenny Gross, Esther M John, Mary B Daly, Saundra M Buys, Melissa C Southey, John L Hopper, Mary Beth Terry, Wendy Chung, Alexander F Miron, David Goldgar, Georgia Chenevix-Trench, Douglas F Easton, Irene L Andrulis, and Antonis C Antoniou. Common breast cancer susceptibility alleles are associated with tumour subtypes in BRCA1 and BRCA2 mutation carriers: results from the Consortium of Investigators of Modifiers of BRCA1/2. *Breast Cancer Research*, 13 (6):R110, December 2011. ISSN 1465-542X. doi: 10.1186/bcr3052. URL `http://breast-cancer-research.biomedcentral.com/articles/10.1186/bcr3052`.

[31] Ang Li, Rong Xie, Qihuan Zhi, Yixiao Deng, Yangming Wu, Weiwei Li, Lu Yang, Zinan Jiao, Jiaqi Luo, Yi Zi, Gang Sun, Jiajia Zhang, Yujian Shi, and Jian Liu. BRCA germline mutations in an unselected nationwide cohort of Chinese patients with ovarian cancer and healthy controls. *Gynecologic Oncology*, 151(1):145–152, October 2018. ISSN 00908258. doi: 10.1016/j.ygyno.2018.07.024. URL `https://linkinghub.elsevier.com/retrieve/pii/S0090825818310886`.

[32] Emilia Rogoża-Janiszewska, Karolina Malińska, Cezary Cybulski, Anna Jakubowska, Jacek Gronwald, Tomasz Huzarski, Marcin Lener, Bohdan Górski, Wojciech Kluźniak, Helena Rudnicka, Mohammad R. Akbari, Aniruddh Kashyap, Steven A. Narod, Jan Lubiński, Tadeusz Debniak, and null On Behalf Of The Polish Hereditary Breast Cancer Consortium. Prevalence of Recurrent Mutations

Predisposing to Breast Cancer in Early-Onset Breast Cancer Patients from Poland. *Cancers*, 12(8), August 2020. ISSN 2072-6694. doi: 10.3390/cancers12082321.

[33] Ht Lynch, Pm Lynch, Sj Lanspa, Cl Snyder, Jf Lynch, and Cr Boland. Review of the Lynch syndrome: history, molecular genetics, screening, differential diagnosis, and medicolegal ramifications. *Clinical Genetics*, 76(1):1–18, July 2009. ISSN 00099163, 13990004. doi: 10.1111/j.1399-0004.2009.01230.x. URL http://doi.wiley.com/10.1111/j.1399-0004.2009.01230.x.

[34] Katarzyna Tutlewska, Jan Lubinski, and Grzegorz Kurzawski. Germline deletions in the EPCAM gene as a cause of Lynch syndrome – literature review. *Hereditary Cancer in Clinical Practice*, 11(1):9, December 2013. ISSN 1897-4287. doi: 10.1186/1897-4287-11-9. URL http://hccpjournal.biomedcentral.com/articles/10.1186/1897-4287-11-9.

[35] Alicia Latham, Preethi Srinivasan, Yelena Kemel, Jinru Shia, Chaitanya Bandlamudi, Diana Mandelker, Sumit Middha, Jaclyn Hechtman, Ahmet Zehir, Marianne Dubard-Gault, Christina Tran, Carolyn Stewart, Margaret Sheehan, Alexander Penson, Deborah DeLair, Rona Yaeger, Joseph Vijai, Semanti Mukherjee, Jesse Galle, Mark A. Dickson, Yelena Janjigian, Eileen M. O'Reilly, Neil Segal, Leonard B. Saltz, Diane Reidy-Lagunes, Anna M. Varghese, Dean Bajorin, Maria I. Carlo, Karen Cadoo, Michael F. Walsh, Martin Weiser, Julio Garcia Aguilar, David S. Klimstra, Luis A. Diaz, Jose Baselga, Liying Zhang, Marc Ladanyi, David M. Hyman, David B. Solit, Mark E. Robson, Barry S. Taylor, Kenneth Offit, Michael F. Berger, and Zsofia K. Stadler. Microsatellite Instability Is Associated With the Presence of Lynch Syndrome Pan-Cancer. *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology*, 37(4):286–295, 2019. ISSN 1527-7755. doi: 10.1200/JCO.18.00283.

[36] J M D Wheeler and W F Bodmer. DNA mismatch repair genes and colorectal cancer. *Gut*, 47(1):148–153, July 2000. ISSN 00175749. doi: 10.1136/gut.47.1.148. URL https://gut.bmj.com/lookup/doi/10.1136/gut.47.1.148.

[37] S. V. Litvinov, M. P. Velders, H. A. Bakker, G. J. Fleuren, and S. O. Warnaar. Ep-CAM: a human epithelial antigen is a homophilic cell-cell adhesion molecule. *The Journal of Cell Biology*, 125(2):437–446, April 1994. ISSN 0021-9525. doi: 10.1083/jcb.125.2.437.

[38] Guo-Min Li. Mechanisms and functions of DNA mismatch repair. *Cell Research*, 18(1):85–98, January 2008. ISSN 1001-0602, 1748-7838. doi: 10.1038/cr.2007.115. URL http://www.nature.com/articles/cr2007115.

[39] Cristian Tomasetti and Bert Vogelstein. Cancer etiology. Variation in cancer risk among tissues can be explained by the number of stem cell divisions. *Science (New York, N.Y.)*, 347(6217):78–81, January 2015. ISSN 1095-9203. doi: 10.1126/science.1260825.

[40] Maurice P.A. Zeegers, Annemarie Jellema, and Harry Ostrer. Empiric risk of prostate carcinoma for relatives of patients with prostate carcinoma. *Cancer*, 97 (8):1894–1903, April 2003. ISSN 0008543X, 10970142. doi: 10.1002/cncr.11262. URL http://doi.wiley.com/10.1002/cncr.11262.

[41] Collaborative Group on Hormonal Factors in Breast Cancer. Familial breast cancer: collaborative reanalysis of individual data from 52 epidemiological studies including 58 209 women with breast cancer and 101 986 women without the disease. *The Lancet*, 358(9291):1389–1399, October 2001. ISSN 01406736. doi: 10.1016/S0140-6736(01)06524-2. URL https://linkinghub.elsevier.com/retrieve/pii/S0140673601065242.

[42] Charles S. Fuchs, Edward L. Giovannucci, Graham A. Colditz, David J. Hunter, Frank E. Speizer, and Walter C. Willett. A Prospective Study of Family History and the Risk of Colorectal Cancer. *New England Journal of Medicine*, 331(25):1669–1674, December 1994. ISSN 0028-4793, 1533-4406. doi: 10.1056/NEJM199412223312501. URL http://www.nejm.org/doi/abs/10.1056/NEJM199412223312501.

[43] Paul Lichtenstein, Niels V. Holm, Pia K. Verkasalo, Anastasia Iliadou, Jaakko Kaprio, Markku Koskenvuo, Eero Pukkala, Axel Skytthe, and Kari Hemminki. Environmental and Heritable Factors in the Causation of Cancer — Analyses of Cohorts of Twins from Sweden, Denmark, and Finland. *New England Journal of Medicine*, 343(2):78–85, July 2000. ISSN 0028-4793, 1533-4406. doi: 10.1056/NEJM200007133430201. URL http://www.nejm.org/doi/abs/10.1056/NEJM200007133430201.

[44] Kamila Czene, Paul Lichtenstein, and Kari Hemminki. Environmental and heritable causes of cancer among 9.6 million individuals in the Swedish family-cancer database. *International Journal of Cancer*, 99(2):260–266, May 2002. ISSN 0020-7136, 1097-0215. doi: 10.1002/ijc.10332. URL http://doi.wiley.com/10.1002/ijc.10332.

[45] Peter M. Visscher, Naomi R. Wray, Qian Zhang, Pamela Sklar, Mark I. McCarthy, Matthew A. Brown, and Jian Yang. 10 Years of GWAS Discovery: Biology, Function, and Translation. *The American Journal of Human Genetics*, 101(1):5–22, July 2017. ISSN 00029297. doi: 10.1016/j.ajhg.2017.06.005. URL https://linkinghub.elsevier.com/retrieve/pii/S0002929717302409.

[46] William S. Bush and Jason H. Moore. Chapter 11: Genome-Wide Association Studies. *PLoS Computational Biology*, 8(12):e1002822, December 2012. ISSN 1553-7358. doi: 10.1371/journal.pcbi.1002822. URL https://dx.plos.org/10.1371/journal.pcbi.1002822.

[47] Baiqiang Liang, Hongrong Ding, Lianfang Huang, Haiqing Luo, and Xiao Zhu. GWAS in cancer: progress and challenges. *Molecular Genetics and Genomics*, 295(3):537–561, May 2020. ISSN 1617-4615, 1617-4623. doi: 10.1007/s00438-020-01647-z. URL http://link.springer.com/10.1007/s00438-020-01647-z.

[48] Mark M. Pomerantz, Yashaswi Shrestha, Richard J. Flavin, Meredith M. Regan, Kathryn L. Penney, Lorelei A. Mucci, Meir J. Stampfer, David J. Hunter, Stephen J. Chanock, Eric J. Schafer, Jennifer A. Chan, Josep Tabernero, José Baselga, Andrea L. Richardson, Massimo Loda, William K. Oh, Philip W. Kantoff, William C. Hahn, and Matthew L. Freedman. Analysis of the 10q11 cancer risk locus implicates MSMB and NCOA4 in human prostate tumorigenesis. *PLoS genetics*, 6(11):e1001204, November 2010. ISSN 1553-7404. doi: 10.1371/journal.pgen.1001204.

[49] Christine Q Chang, Ajay Yesupriya, Jessica L Rowell, Camilla B Pimentel, Melinda Clyne, Marta Gwinn, Muin J Khoury, Anja Wulf, and Sheri D Schully. A systematic

review of cancer GWAS and candidate gene meta-analyses reveals limited overlap but similar effect sizes. *European Journal of Human Genetics*, 22(3):402–408, March 2014. ISSN 1018-4813, 1476-5438. doi: 10.1038/ejhg.2013.161. URL `http://www.nature.com/doifinder/10.1038/ejhg.2013.161`.

[50] Teri A. Manolio, Francis S. Collins, Nancy J. Cox, David B. Goldstein, Lucia A. Hindorff, David J. Hunter, Mark I. McCarthy, Erin M. Ramos, Lon R. Cardon, Aravinda Chakravarti, Judy H. Cho, Alan E. Guttmacher, Augustine Kong, Leonid Kruglyak, Elaine Mardis, Charles N. Rotimi, Montgomery Slatkin, David Valle, Alice S. Whittemore, Michael Boehnke, Andrew G. Clark, Evan E. Eichler, Greg Gibson, Jonathan L. Haines, Trudy F. C. Mackay, Steven A. McCarroll, and Peter M. Visscher. Finding the missing heritability of complex diseases. *Nature*, 461(7265): 747–753, October 2009. ISSN 0028-0836, 1476-4687. doi: 10.1038/nature08494. URL `http://www.nature.com/articles/nature08494`.

[51] kConFab Investigators, ABCTB Investigators, EMBRACE Study, GEMO Study Collaborators, Haoyu Zhang, Thomas U. Ahearn, Julie Lecarpentier, Daniel Barnes, Jonathan Beesley, Guanghao Qi, Xia Jiang, Tracy A. O'Mara, Ni Zhao, Manjeet K. Bolla, Alison M. Dunning, Joe Dennis, Qin Wang, Zumuruda Abu Ful, Kristiina Aittomäki, Irene L. Andrulis, Hoda Anton-Culver, Volker Arndt, Kristan J. Aronson, Banu K. Arun, Paul L. Auer, Jacopo Azzollini, Daniel Barrowdale, Heiko Becher, Matthias W. Beckmann, Sabine Behrens, Javier Benitez, Marina Bermisheva, Katarzyna Bialkowska, Ana Blanco, Carl Blomqvist, Natalia V. Bogdanova, Stig E. Bojesen, Bernardo Bonanni, Davide Bondavalli, Ake Borg, Hiltrud Brauch, Hermann Brenner, Ignacio Briceno, Annegien Broeks, Sara Y. Brucker, Thomas Brüning, Barbara Burwinkel, Saundra S. Buys, Helen Byers, Trinidad Caldés, Maria A. Caligo, Mariarosaria Calvello, Daniele Campa, Jose E. Castelao, Jenny Chang-Claude, Stephen J. Chanock, Melissa Christiaens, Hans Christiansen, Wendy K. Chung, Kathleen B. M. Claes, Christine L. Clarke, Sten Cornelissen, Fergus J. Couch, Angela Cox, Simon S. Cross, Kamila Czene, Mary B. Daly, Peter Devilee, Orland Diez, Susan M. Domchek, Thilo Dörk, Miriam Dwek, Diana M. Eccles, Arif B. Ekici, D. Gareth Evans, Peter A. Fasching, Jonine Figueroa, Lenka Foretova, Florentia Fostira, Eitan Friedman, Debra Frost, Manuela Gago-Dominguez, Susan M. Gapstur, Judy Garber, José A. García-Sáenz, Mia M. Gaudet, Simon A. Gayther, Graham G. Giles, Andrew K. Godwin, Mark S. Goldberg, David E. Goldgar, Anna González-Neira, Mark H. Greene, Jacek Gronwald, Pascal Guénel, Lothar Häberle, Eric Hahnen, Christopher A. Haiman, Christopher R. Hake, Per Hall, Ute Hamann, Elaine F. Harkness, Bernadette A. M. Heemskerk-Gerritsen, Peter Hillemanns, Frans B. L. Hogervorst, Bernd Holleczek, Antoinette Hollestelle, Maartje J. Hooning, Robert N. Hoover, John L. Hopper, Anthony Howell, Hanna Huebner, Peter J. Hulick, Evgeny N. Imyanitov, Claudine Isaacs, Louise Izatt, Agnes Jager, Milena Jakimovska, Anna Jakubowska, Paul James, Ramunas Janavicius, Wolfgang Janni, Esther M. John, Michael E. Jones, Audrey Jung, Rudolf Kaaks, Pooja Middha Kapoor, Beth Y. Karlan, Renske Keeman, Sofia Khan, Elza Khusnutdinova, Cari M. Kitahara, Yon-Dschun Ko, Irene Konstantopoulou, Linetta B. Kop-

pert, Stella Koutros, Vessela N. Kristensen, Anne-Vibeke Laenkholm, Diether Lambrechts, Susanna C. Larsson, Pierre Laurent-Puig, Conxi Lazaro, Emilija Lazarova, Flavio Lejbkowicz, Goska Leslie, Fabienne Lesueur, Annika Lindblom, Jolanta Lissowska, Wing-Yee Lo, Jennifer T. Loud, Jan Lubinski, Alicja Lukomska, Robert J. MacInnis, Arto Mannermaa, Mehdi Manoochehri, Siranoush Manoukian, Sara Margolin, Maria Elena Martinez, Laura Matricardi, Lesley McGuffog, Catriona McLean, Noura Mebirouk, Alfons Meindl, Usha Menon, Austin Miller, Elvira Mingazheva, Marco Montagna, Anna Marie Mulligan, Claire Mulot, Taru A. Muranen, Katherine L. Nathanson, Susan L. Neuhausen, Heli Nevanlinna, Patrick Neven, William G. Newman, Finn C. Nielsen, Liene Nikitina-Zake, Jesse Nodora, Kenneth Offit, Edith Olah, Olufunmilayo I. Olopade, Håkan Olsson, Nick Orr, Laura Papi, Janos Papp, Tjoung-Won Park-Simon, Michael T. Parsons, Bernard Peissel, Ana Peixoto, Beth Peshkin, Paolo Peterlongo, Julian Peto, Kelly-Anne Phillips, Marion Piedmonte, Dijana Plaseska-Karanfilska, Karolina Prajzendanc, Ross Prentice, Darya Prokofyeva, Brigitte Rack, Paolo Radice, Susan J. Ramus, Johanna Rantala, Muhammad U. Rashid, Gad Rennert, Hedy S. Rennert, Harvey A. Risch, Atocha Romero, Matti A. Rookus, Matthias Rübner, Thomas Rüdiger, Emmanouil Saloustros, Sarah Sampson, Dale P. Sandler, Elinor J. Sawyer, Maren T. Scheuner, Rita K. Schmutzler, Andreas Schneeweiss, Minouk J. Schoemaker, Ben Schöttker, Peter Schürmann, Leigha Senter, Priyanka Sharma, Mark E. Sherman, Xiao-Ou Shu, Christian F. Singer, Snezhana Smichkoska, Penny Soucy, Melissa C. Southey, John J. Spinelli, Jennifer Stone, Dominique Stoppa-Lyonnet, Anthony J. Swerdlow, Csilla I. Szabo, Rulla M. Tamimi, William J. Tapper, Jack A. Taylor, Manuel R. Teixeira, MaryBeth Terry, Mads Thomassen, Darcy L. Thull, Marc Tischkowitz, Amanda E. Toland, Rob A. E. M. Tollenaar, Ian Tomlinson, Diana Torres, Melissa A. Troester, Thérèse Truong, Nadine Tung, Michael Untch, Celine M. Vachon, Ans M. W. van den Ouweland, Lizet E. van der Kolk, Elke M. van Veen, Elizabeth J. vanRensburg, Ana Vega, Barbara Wappenschmidt, Clarice R. Weinberg, Jeffrey N. Weitzel, Hans Wildiers, Robert Winqvist, Alicja Wolk, Xiaohong R. Yang, Drakoulis Yannoukakos, Wei Zheng, Kristin K. Zorn, Roger L. Milne, Peter Kraft, Jacques Simard, Paul D. P. Pharoah, Kyriaki Michailidou, Antonis C. Antoniou, Marjanka K. Schmidt, Georgia Chenevix-Trench, Douglas F. Easton, Nilanjan Chatterjee, and Montserrat García-Closas. Genome-wide association study identifies 32 novel breast cancer susceptibility loci from overall and subtype-specific analyses. *Nature Genetics*, 52(6):572–581, June 2020. ISSN 1061-4036, 1546-1718. doi: 10.1038/s41588-020-0609-2. URL `http://www.nature.com/articles/s41588-020-0609-2`.

[52] The PRACTICAL consortium, Philip J. Law, Maria Timofeeva, Ceres Fernandez-Rozadilla, Peter Broderick, James Studd, Juan Fernandez-Tajes, Susan Farrington, Victoria Svinti, Claire Palles, Giulia Orlando, Amit Sud, Amy Holroyd, Steven Penegar, Evropi Theodoratou, Peter Vaughan-Shaw, Harry Campbell, Lina Zgaga, Caroline Hayward, Archie Campbell, Sarah Harris, Ian J. Deary, John Starr, Laura Gatcombe, Maria Pinna, Sarah Briggs, Lynn Martin, Emma Jaeger, Archana Sharma-Oates, James East, Simon Leedham, Roland Arnold, Elaine Johnstone, Haitao

Wang, David Kerr, Rachel Kerr, Tim Maughan, Richard Kaplan, Nada Al-Tassan, Kimmo Palin, Ulrika A. Hänninen, Tatiana Cajuso, Tomas Tanskanen, Johanna Kondelin, Eevi Kaasinen, Antti-Pekka Sarin, Johan G. Eriksson, Harri Rissanen, Paul Knekt, Eero Pukkala, Pekka Jousilahti, Veikko Salomaa, Samuli Ripatti, Aarno Palotie, Laura Renkonen-Sinisalo, Anna Lepistö, Jan Böhm, Jukka-Pekka Mecklin, Daniel D. Buchanan, Aung-Ko Win, John Hopper, Mark E. Jenkins, Noralane M. Lindor, Polly A. Newcomb, Steven Gallinger, David Duggan, Graham Casey, Per Hoffmann, Markus M. Nöthen, Karl-Heinz Jöckel, Douglas F. Easton, Paul D. P. Pharoah, Julian Peto, Federico Canzian, Anthony Swerdlow, Rosalind A. Eeles, Zsofia Kote-Jarai, Kenneth Muir, Nora Pashayan, Andrea Harkin, Karen Allan, John McQueen, James Paul, Timothy Iveson, Mark Saunders, Katja Butterbach, Jenny Chang-Claude, Michael Hoffmeister, Hermann Brenner, Iva Kirac, Petar Matošević, Philipp Hofer, Stefanie Brezina, Andrea Gsur, Jeremy P. Cheadle, Lauri A. Aaltonen, Ian Tomlinson, Richard S. Houlston, and Malcolm G. Dunlop. Association analyses identify 31 new risk loci for colorectal cancer susceptibility. *Nature Communications*, 10(1):2154, December 2019. ISSN 2041-1723. doi: 10.1038/s41467-019-09775-w. URL `http://www.nature.com/articles/s41467-019-09775-w`.

[53] John Lonsdale, Jeffrey Thomas, Mike Salvatore, Rebecca Phillips, Edmund Lo, Saboor Shad, Richard Hasz, Gary Walters, Fernando Garcia, Nancy Young, Barbara Foster, Mike Moser, Ellen Karasik, Bryan Gillard, Kimberley Ramsey, Susan Sullivan, Jason Bridge, Harold Magazine, John Syron, Johnelle Fleming, Laura Siminoff, Heather Traino, Maghboeba Mosavel, Laura Barker, Scott Jewell, Dan Rohrer, Dan Maxim, Dana Filkins, Philip Harbach, Eddie Cortadillo, Bree Berghuis, Lisa Turner, Eric Hudson, Kristin Feenstra, Leslie Sobin, James Robb, Phillip Branton, Greg Korzeniewski, Charles Shive, David Tabor, Liqun Qi, Kevin Groch, Sreenath Nampally, Steve Buia, Angela Zimmerman, Anna Smith, Robin Burges, Karna Robinson, Kim Valentino, Deborah Bradbury, Mark Cosentino, Norma Diaz-Mayoral, Mary Kennedy, Theresa Engel, Penelope Williams, Kenyon Erickson, Kristin Ardlie, Wendy Winckler, Gad Getz, David DeLuca, Daniel MacArthur, Manolis Kellis, Alexander Thomson, Taylor Young, Ellen Gelfand, Molly Donovan, Yan Meng, George Grant, Deborah Mash, Yvonne Marcus, Margaret Basile, Jun Liu, Jun Zhu, Zhidong Tu, Nancy J Cox, Dan L Nicolae, Eric R Gamazon, Hae Kyung Im, Anuar Konkashbaev, Jonathan Pritchard, Matthew Stevens, Timothèe Flutre, Xiaoquan Wen, Emmanouil T Dermitzakis, Tuuli Lappalainen, Roderic Guigo, Jean Monlong, Michael Sammeth, Daphne Koller, Alexis Battle, Sara Mostafavi, Mark McCarthy, Manual Rivas, Julian Maller, Ivan Rusyn, Andrew Nobel, Fred Wright, Andrey Shabalin, Mike Feolo, Nataliya Sharopova, Anne Sturcke, Justin Paschal, James M Anderson, Elizabeth L Wilder, Leslie K Derr, Eric D Green, Jeffery P Struewing, Gary Temple, Simona Volpi, Joy T Boyer, Elizabeth J Thomson, Mark S Guyer, Cathy Ng, Assya Abdallah, Deborah Colantuoni, Thomas R Insel, Susan E Koester, A Roger Little, Patrick K Bender, Thomas Lehner, Yin Yao, Carolyn C Compton, Jimmie B Vaught, Sherilyn Sawyer, Nicole C Lockhart, Joanne Demchok, and He-

len F Moore. The Genotype-Tissue Expression (GTEx) project. *Nature Genetics*, 45(6):580–585, June 2013. ISSN 1061-4036, 1546-1718. doi: 10.1038/ng.2653. URL http://www.nature.com/articles/ng.2653.

[54] David J. Hunter, Peter Kraft, Kevin B. Jacobs, David G. Cox, Meredith Yeager, Susan E. Hankinson, Sholom Wacholder, Zhaoming Wang, Robert Welch, Amy Hutchinson, Junwen Wang, Kai Yu, Nilanjan Chatterjee, Nick Orr, Walter C. Willett, Graham A. Colditz, Regina G. Ziegler, Christine D. Berg, Saundra S. Buys, Catherine A. McCarty, Heather Spencer Feigelson, Eugenia E. Calle, Michael J. Thun, Richard B. Hayes, Margaret Tucker, Daniela S. Gerhard, Joseph F. Fraumeni, Robert N. Hoover, Gilles Thomas, and Stephen J. Chanock. A genome-wide association study identifies alleles in FGFR2 associated with risk of sporadic postmenopausal breast cancer. *Nature Genetics*, 39(7):870–874, July 2007. ISSN 1061-4036. doi: 10.1038/ng2075.

[55] Michael N. C. Fletcher, Mauro A. A. Castro, Xin Wang, Ines de Santiago, Martin O'Reilly, Suet-Feung Chin, Oscar M. Rueda, Carlos Caldas, Bruce A. J. Ponder, Florian Markowetz, and Kerstin B. Meyer. Master regulators of FGFR2 signalling and breast cancer risk. *Nature Communications*, 4(1):2464, December 2013. ISSN 2041-1723. doi: 10.1038/ncomms3464. URL http://www.nature.com/articles/ncomms3464.

[56] Thomas M. Campbell, Mauro A.A. Castro, Ines de Santiago, Michael N.C. Fletcher, Silvia Halim, Radhika Prathalingam, Bruce A.J. Ponder, and Kerstin B. Meyer. FGFR2 risk SNPs confer breast cancer risk by augmenting oestrogen responsiveness. *Carcinogenesis*, 37(8):741–750, August 2016. ISSN 0143-3334, 1460-2180. doi: 10.1093/carcin/bgw065. URL https://academic.oup.com/carcin/article-lookup/doi/10.1093/carcin/bgw065.

[57] Matthew B. Yurgelun, Georgia Chenevix-Trench, and Scott M. Lippman. Translating Germline Cancer Risk into Precision Prevention. *Cell*, 168(4):566–570, February 2017. ISSN 00928674. doi: 10.1016/j.cell.2017.01.031. URL https://linkinghub.elsevier.com/retrieve/pii/S0092867417301162.

[58] Douglas Hanahan and Robert A. Weinberg. Hallmarks of cancer: the next generation. *Cell*, 144(5):646–674, March 2011. ISSN 1097-4172. doi: 10.1016/j.cell.2011.02.013.

[59] S. Lilly Zheng, Jielin Sun, Fredrik Wiklund, Shelly Smith, Pär Stattin, Ge Li, Hans-Olov Adami, Fang-Chi Hsu, Yi Zhu, Katarina Bälter, A. Karim Kader, Aubrey R. Turner, Wennuan Liu, Eugene R. Bleecker, Deborah A. Meyers, David Duggan, John D. Carpten, Bao-Li Chang, William B. Isaacs, Jianfeng Xu, and Henrik Grönberg. Cumulative Association of Five Genetic Variants with Prostate Cancer. *New England Journal of Medicine*, 358(9):910–919, February 2008. ISSN 0028-4793, 1533-4406. doi: 10.1056/NEJMoa075819. URL http://www.nejm.org/doi/abs/10.1056/NEJMoa075819.

[60] Linda Kachuri, Rebecca E. Graff, Karl Smith-Byrne, Travis J. Meyers, Sara R. Rashkin, Elad Ziv, John S. Witte, and Mattias Johansson. Pan-cancer analysis

demonstrates that integrating polygenic risk scores with modifiable risk factors improves risk prediction. *Nature Communications*, 11(1):6084, December 2020. ISSN 2041-1723. doi: 10.1038/s41467-020-19600-4. URL http://www.nature.com/articles/s41467-020-19600-4.

[61] M. J. E. Frampton, P. Law, K. Litchfield, E. J. Morris, D. Kerr, C. Turnbull, I. P. Tomlinson, and R. S. Houlston. Implications of polygenic risk for personalised colorectal cancer screening. *Annals of Oncology: Official Journal of the European Society for Medical Oncology*, 27(3):429–434, March 2016. ISSN 1569-8041. doi: 10.1093/annonc/mdv540.

[62] Karoline B Kuchenbaecker, Lesley McGuffog, Daniel Barrowdale, Andrew Lee, Penny Soucy, Joe Dennis, Susan M Domchek, Mark Robson, Amanda B Spurdle, Susan J Ramus, Nasim Mavaddat, Mary Beth Terry, Susan L Neuhausen, Rita Katharina Schmutzler, Jacques Simard, Paul D P Pharoah, Kenneth Offit, Fergus J Couch, Georgia Chenevix-Trench, Douglas F Easton, and Antonis C Antoniou. Evaluation of Polygenic Risk Scores for Breast and Ovarian Cancer Risk Prediction in BRCA1 and BRCA2 Mutation Carriers. *JNCI: Journal of the National Cancer Institute*, 109 (7), July 2017. ISSN 0027-8874, 1460-2105. doi: 10.1093/jnci/djw302. URL https://academic.oup.com/jnci/article/doi/10.1093/jnci/djw302/3064534.

[63] Matthew R Nelson, Hannah Tipney, Jeffery L Painter, Judong Shen, Paola Nicoletti, Yufeng Shen, Aris Floratos, Pak Chung Sham, Mulin Jun Li, Junwen Wang, Lon R Cardon, John C Whittaker, and Philippe Sanseau. The support of human genetic evidence for approved drug indications. *Nature Genetics*, 47(8):856–860, August 2015. ISSN 1061-4036, 1546-1718. doi: 10.1038/ng.3314. URL http://www.nature.com/articles/ng.3314.

[64] International Agency for Research on Cancer and International Association of Cancer Registries. *Cancer incidence in five continents. Volume X Volume X*. 2014. ISBN 978-3-642-85851-2. URL http://www.iarc.fr/en/publications/pdfs-online/epi/sp164/. OCLC: 1027788733.

[65] Michael R. Stratton, Peter J. Campbell, and P. Andrew Futreal. The cancer genome. *Nature*, 458(7239):719–724, April 2009. ISSN 0028-0836, 1476-4687. doi: 10.1038/nature07943. URL http://www.nature.com/articles/nature07943.

[66] B. Vogelstein, N. Papadopoulos, V. E. Velculescu, S. Zhou, L. A. Diaz, and K. W. Kinzler. Cancer Genome Landscapes. *Science*, 339(6127):1546–1558, March 2013. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.1235122. URL http://www.sciencemag.org/cgi/doi/10.1126/science.1235122.

[67] Matthew H. Bailey, Collin Tokheim, Eduard Porta-Pardo, Sohini Sengupta, Denis Bertrand, Amila Weerasinghe, Antonio Colaprico, Michael C. Wendl, Jaegil Kim, Brendan Reardon, Patrick Kwok-Shing Ng, Kang Jin Jeong, Song Cao, Zixing Wang, Jianjiong Gao, Qingsong Gao, Fang Wang, Eric Minwei Liu, Loris Mularoni, Carlota Rubio-Perez, Niranjan Nagarajan, Isidro Cortés-Ciriano, Daniel Cui Zhou, Wen-Wei Liang, Julian M. Hess, Venkata D. Yellapantula, David Tamborero, Abel Gonzalez-Perez, Chayaporn Suphavilai, Jia Yu Ko, Ekta Khurana, Peter J. Park, Eliezer M. Van Allen, Han Liang, Michael S. Lawrence, Adam Godzik, Nuria

Lopez-Bigas, Josh Stuart, David Wheeler, Gad Getz, Ken Chen, Alexander J. Lazar, Gordon B. Mills, Rachel Karchin, Li Ding, Samantha J. Caesar-Johnson, John A. Demchok, Ina Felau, Melpomeni Kasapi, Martin L. Ferguson, Carolyn M. Hutter, Heidi J. Sofia, Roy Tarnuzzer, Zhining Wang, Liming Yang, Jean C. Zenklusen, Jiashan (Julia) Zhang, Sudha Chudamani, Jia Liu, Laxmi Lolla, Rashi Naresh, Todd Pihl, Qiang Sun, Yunhu Wan, Ye Wu, Juok Cho, Timothy DeFreitas, Scott Frazer, Nils Gehlenborg, Gad Getz, David I. Heiman, Jaegil Kim, Michael S. Lawrence, Pei Lin, Sam Meier, Michael S. Noble, Gordon Saksena, Doug Voet, Hailei Zhang, Brady Bernard, Nyasha Chambwe, Varsha Dhankani, Theo Knijnenburg, Roger Kramer, Kalle Leinonen, Yuexin Liu, Michael Miller, Sheila Reynolds, Ilya Shmulevich, Vesteinn Thorsson, Wei Zhang, Rehan Akbani, Bradley M. Broom, Apurva M. Hegde, Zhenlin Ju, Rupa S. Kanchi, Anil Korkut, Jun Li, Han Liang, Shiyun Ling, Wenbin Liu, Yiling Lu, Gordon B. Mills, Kwok-Shing Ng, Arvind Rao, Michael Ryan, Jing Wang, John N. Weinstein, Jiexin Zhang, Adam Abeshouse, Joshua Armenia, Debyani Chakravarty, Walid K. Chatila, Ino de Bruijn, Jianjiong Gao, Benjamin E. Gross, Zachary J. Heins, Ritika Kundra, Konnor La, Marc Ladanyi, Augustin Luna, Moriah G. Nissan, Angelica Ochoa, Sarah M. Phillips, Ed Reznik, Francisco Sanchez-Vega, Chris Sander, Nikolaus Schultz, Robert Sheridan, S. Onur Sumer, Yichao Sun, Barry S. Taylor, Jioajiao Wang, Hongxin Zhang, Pavana Anur, Myron Peto, Paul Spellman, Christopher Benz, Joshua M. Stuart, Christopher K. Wong, Christina Yau, D. Neil Hayes, Joel S. Parker, Matthew D. Wilkerson, Adrian Ally, Miruna Balasundaram, Reanne Bowlby, Denise Brooks, Rebecca Carlsen, Eric Chuah, Noreen Dhalla, Robert Holt, Steven J.M. Jones, Katayoon Kasaian, Darlene Lee, Yussanne Ma, Marco A. Marra, Michael Mayo, Richard A. Moore, Andrew J. Mungall, Karen Mungall, A. Gordon Robertson, Sara Sadeghi, Jacqueline E. Schein, Payal Sipahimalani, Angela Tam, Nina Thiessen, Kane Tse, Tina Wong, Ashton C. Berger, Rameen Beroukhim, Andrew D. Cherniack, Carrie Cibulskis, Stacey B. Gabriel, Galen F. Gao, Gavin Ha, Matthew Meyerson, Steven E. Schumacher, Juliann Shih, Melanie H. Kucherlapati, Raju S. Kucherlapati, Stephen Baylin, Leslie Cope, Ludmila Danilova, Moiz S. Bootwalla, Phillip H. Lai, Dennis T. Maglinte, David J. Van Den Berg, Daniel J. Weisenberger, J. Todd Auman, Saianand Balu, Tom Bodenheimer, Cheng Fan, Katherine A. Hoadley, Alan P. Hoyle, Stuart R. Jefferys, Corbin D. Jones, Shaowu Meng, Piotr A. Mieczkowski, Lisle E. Mose, Amy H. Perou, Charles M. Perou, Jeffrey Roach, Yan Shi, Janae V. Simons, Tara Skelly, Matthew G. Soloway, Donghui Tan, Umadevi Veluvolu, Huihui Fan, Toshinori Hinoue, Peter W. Laird, Hui Shen, Wanding Zhou, Michelle Bellair, Kyle Chang, Kyle Covington, Chad J. Creighton, Huyen Dinh, HarshaVardhan Doddapaneni, Lawrence A. Donehower, Jennifer Drummond, Richard A. Gibbs, Robert Glenn, Walker Hale, Yi Han, Jianhong Hu, Viktoriya Korchina, Sandra Lee, Lora Lewis, Wei Li, Xiuping Liu, Margaret Morgan, Donna Morton, Donna Muzny, Jireh Santibanez, Margi Sheth, Eve Shinbrot, Linghua Wang, Min Wang, David A. Wheeler, Liu Xi, Fengmei Zhao, Julian Hess, Elizabeth L. Appelbaum, Matthew Bailey, Matthew G. Cordes, Li Ding, Catrina C. Fronick, Lucinda A.

Fulton, Robert S. Fulton, Cyriac Kandoth, Elaine R. Mardis, Michael D. McLellan, Christopher A. Miller, Heather K. Schmidt, Richard K. Wilson, Daniel Crain, Erin Curley, Johanna Gardner, Kevin Lau, David Mallery, Scott Morris, Joseph Paulauskis, Robert Penny, Candace Shelton, Troy Shelton, Mark Sherman, Eric Thompson, Peggy Yena, Jay Bowen, Julie M. Gastier-Foster, Mark Gerken, Kristen M. Leraas, Tara M. Lichtenberg, Nilsa C. Ramirez, Lisa Wise, Erik Zmuda, Niall Corcoran, Tony Costello, Christopher Hovens, Andre L. Carvalho, Ana C. de Carvalho, José H. Fregnani, Adhemar Longatto-Filho, Rui M. Reis, Cristovam Scapulatempo-Neto, Henrique C.S. Silveira, Daniel O. Vidal, Andrew Burnette, Jennifer Eschbacher, Beth Hermes, Ardene Noss, Rosy Singh, Matthew L. Anderson, Patricia D. Castro, Michael Ittmann, David Huntsman, Bernard Kohl, Xuan Le, Richard Thorp, Chris Andry, Elizabeth R. Duffy, Vladimir Lyadov, Oxana Paklina, Galiya Setdikova, Alexey Shabunin, Mikhail Tavobilov, Christopher McPherson, Ronald Warnick, Ross Berkowitz, Daniel Cramer, Colleen Feltmate, Neil Horowitz, Adam Kibel, Michael Muto, Chandrajit P. Raut, Andrei Malykh, Jill S. Barnholtz-Sloan, Wendi Barrett, Karen Devine, Jordonna Fulop, Quinn T. Ostrom, Kristen Shimmel, Yingli Wolinsky, Andrew E. Sloan, Agostino De Rose, Felice Giuliante, Marc Goodman, Beth Y. Karlan, Curt H. Hagedorn, John Eckman, Jodi Harr, Jerome Myers, Kelinda Tucker, Leigh Anne Zach, Brenda Deyarmin, Hai Hu, Leonid Kvecher, Caroline Larson, Richard J. Mural, Stella Somiari, Ales Vicha, Tomas Zelinka, Joseph Bennett, Mary Iacocca, Brenda Rabeno, Patricia Swanson, Mathieu Latour, Louis Lacombe, Bernard Têtu, Alain Bergeron, Mary McGraw, Susan M. Staugaitis, John Chabot, Hanina Hibshoosh, Antonia Sepulveda, Tao Su, Timothy Wang, Olga Potapova, Olga Voronina, Laurence Desjardins, Odette Mariani, Sergio Roman-Roman, Xavier Sastre, Marc-Henri Stern, Feixiong Cheng, Sabina Signoretti, Andrew Berchuck, Darell Bigner, Eric Lipp, Jeffrey Marks, Shannon McCall, Roger McLendon, Angeles Secord, Alexis Sharp, Madhusmita Behera, Daniel J. Brat, Amy Chen, Keith Delman, Seth Force, Fadlo Khuri, Kelly Magliocca, Shishir Maithel, Jeffrey J. Olson, Taofeek Owonikoko, Alan Pickens, Suresh Ramalingam, Dong M. Shin, Gabriel Sica, Erwin G. Van Meir, Hongzheng Zhang, Wil Eijckenboom, Ad Gillis, Esther Korpershoek, Leendert Looijenga, Wolter Oosterhuis, Hans Stoop, Kim E. van Kessel, Ellen C. Zwarthoff, Chiara Calatozzolo, Lucia Cuppini, Stefania Cuzzubbo, Francesco DiMeco, Gaetano Finocchiaro, Luca Mattei, Alessandro Perin, Bianca Pollo, Chu Chen, John Houck, Pawadee Lohavanichbutr, Arndt Hartmann, Christine Stoehr, Robert Stoehr, Helge Taubert, Sven Wach, Bernd Wullich, Witold Kycler, Dawid Murawa, Maciej Wiznerowicz, Ki Chung, W. Jeffrey Edenfield, Julie Martin, Eric Baudin, Glenn Bubley, Raphael Bueno, Assunta De Rienzo, William G. Richards, Steven Kalkanis, Tom Mikkelsen, Houtan Noushmehr, Lisa Scarpace, Nicolas Girard, Marta Aymerich, Elias Campo, Eva Giné, Armando López Guillermo, Nguyen Van Bang, Phan Thi Hanh, Bui Duc Phu, Yufang Tang, Howard Colman, Kimberley Evason, Peter R. Dottino, John A. Martignetti, Hani Gabra, Hartmut Juhl, Teniola Akeredolu, Serghei Stepa, Dave Hoon, Keunsoo Ahn, Koo Jeong Kang, Felix Beuschlein, Anne Breggia, Michael Birrer,

Debra Bell, Mitesh Borad, Alan H. Bryce, Erik Castle, Vishal Chandan, John Cheville, John A. Copland, Michael Farnell, Thomas Flotte, Nasra Giama, Thai Ho, Michael Kendrick, Jean-Pierre Kocher, Karla Kopp, Catherine Moser, David Nagorney, Daniel O'Brien, Brian Patrick O'Neill, Tushar Patel, Gloria Petersen, Florencia Que, Michael Rivera, Lewis Roberts, Robert Smallridge, Thomas Smyrk, Melissa Stanton, R. Houston Thompson, Michael Torbenson, Ju Dong Yang, Lizhi Zhang, Fadi Brimo, Jaffer A. Ajani, Ana Maria Angulo Gonzalez, Carmen Behrens, Jolanta Bondaruk, Russell Broaddus, Bogdan Czerniak, Bita Esmaeli, Junya Fujimoto, Jeffrey Gershenwald, Charles Guo, Alexander J. Lazar, Christopher Logothetis, Funda Meric-Bernstam, Cesar Moran, Lois Ramondetta, David Rice, Anil Sood, Pheroze Tamboli, Timothy Thompson, Patricia Troncoso, Anne Tsao, Ignacio Wistuba, Candace Carter, Lauren Haydu, Peter Hersey, Valerie Jakrot, Hojabr Kakavand, Richard Kefford, Kenneth Lee, Georgina Long, Graham Mann, Michael Quinn, Robyn Saw, Richard Scolyer, Kerwin Shannon, Andrew Spillane, Jonathan Stretch, Maria Synott, John Thompson, James Wilmott, Hikmat Al-Ahmadie, Timothy A. Chan, Ronald Ghossein, Anuradha Gopalan, Douglas A. Levine, Victor Reuter, Samuel Singer, Bhuvanesh Singh, Nguyen Viet Tien, Thomas Broudy, Cyrus Mirsaidi, Praveen Nair, Paul Drwiega, Judy Miller, Jennifer Smith, Howard Zaren, Joong-Won Park, Nguyen Phi Hung, Electron Kebebew, W. Marston Linehan, Adam R. Metwalli, Karel Pacak, Peter A. Pinto, Mark Schiffman, Laura S. Schmidt, Cathy D. Vocke, Nicolas Wentzensen, Robert Worrell, Hannah Yang, Marc Moncrieff, Chandra Goparaju, Jonathan Melamed, Harvey Pass, Natalia Botnariuc, Irina Caraman, Mircea Cernat, Inga Chemencedji, Adrian Clipca, Serghei Doruc, Ghenadie Gorincioi, Sergiu Mura, Maria Pirtac, Irina Stancul, Diana Tcaciuc, Monique Albert, Iakovina Alexopoulou, Angel Arnaout, John Bartlett, Jay Engel, Sebastien Gilbert, Jeremy Parfitt, Harman Sekhon, George Thomas, Doris M. Rassl, Robert C. Rintoul, Carlo Bifulco, Raina Tamakawa, Walter Urba, Nicholas Hayward, Henri Timmers, Anna Antenucci, Francesco Facciolo, Gianluca Grazi, Mirella Marino, Roberta Merola, Ronald de Krijger, Anne-Paule Gimenez-Roqueplo, Alain Piché, Simone Chevalier, Ginette McKercher, Kivanc Birsoy, Gene Barnett, Cathy Brewer, Carol Farver, Theresa Naska, Nathan A. Pennell, Daniel Raymond, Cathy Schilero, Kathy Smolenski, Felicia Williams, Carl Morrison, Jeffrey A. Borgia, Michael J. Liptay, Mark Pool, Christopher W. Seder, Kerstin Junker, Larsson Omberg, Mikhail Dinkin, George Manikhas, Domenico Alvaro, Maria Consiglia Bragazzi, Vincenzo Cardinale, Guido Carpino, Eugenio Gaudio, David Chesla, Sandra Cottingham, Michael Dubina, Fedor Moiseenko, Renumathy Dhanasekaran, Karl-Friedrich Becker, Klaus-Peter Janssen, Julia Slotta-Huspenina, Mohamed H. Abdel-Rahman, Dina Aziz, Sue Bell, Colleen M. Cebulla, Amy Davis, Rebecca Duell, J. Bradley Elder, Joe Hilty, Bahavna Kumar, James Lang, Norman L. Lehman, Randy Mandt, Phuong Nguyen, Robert Pilarski, Karan Rai, Lynn Schoenfield, Kelly Senecal, Paul Wakely, Paul Hansen, Ronald Lechan, James Powers, Arthur Tischler, William E. Grizzle, Katherine C. Sexton, Alison Kastl, Joel Henderson, Sima Porten, Jens Waldmann, Martin Fassnacht, Sylvia L. Asa, Dirk Schaden-

dorf, Marta Couce, Markus Graefen, Hartwig Huland, Guido Sauter, Thorsten Schlomm, Ronald Simon, Pierre Tennstedt, Oluwole Olabode, Mark Nelson, Oliver Bathe, Peter R. Carroll, June M. Chan, Philip Disaia, Pat Glenn, Robin K. Kelley, Charles N. Landen, Joanna Phillips, Michael Prados, Jeffry Simko, Karen Smith-McCune, Scott VandenBerg, Kevin Roggin, Ashley Fehrenbach, Ady Kendler, Suzanne Sifri, Ruth Steele, Antonio Jimeno, Francis Carey, Ian Forgie, Massimo Mannelli, Michael Carney, Brenda Hernandez, Benito Campos, Christel Herold-Mende, Christin Jungk, Andreas Unterberg, Andreas von Deimling, Aaron Bossler, Joseph Galbraith, Laura Jacobus, Michael Knudson, Tina Knutson, Deqin Ma, Mohammed Milhem, Rita Sigmund, Andrew K. Godwin, Rashna Madan, Howard G. Rosenthal, Clement Adebamowo, Sally N. Adebamowo, Alex Boussioutas, David Beer, Thomas Giordano, Anne-Marie Mes-Masson, Fred Saad, Therese Bocklage, Lisa Landrum, Robert Mannel, Kathleen Moore, Katherine Moxley, Russel Postier, Joan Walker, Rosemary Zuna, Michael Feldman, Federico Valdivieso, Rajiv Dhir, James Luketich, Edna M. Mora Pinero, Mario Quintero-Aguilo, Carlos Gilberto Carlotti, Jose Sebastião Dos Santos, Rafael Kemp, Ajith Sankarankuty, Daniela Tirapelli, James Catto, Kathy Agnew, Elizabeth Swisher, Jenette Creaney, Bruce Robinson, Carl Simon Shelley, Eryn M. Godwin, Sara Kendall, Cassaundra Shipman, Carol Bradford, Thomas Carey, Andrea Haddad, Jeffey Moyer, Lisa Peterson, Mark Prince, Laura Rozek, Gregory Wolf, Rayleen Bowman, Kwun M. Fong, Ian Yang, Robert Korst, W. Kimryn Rathmell, J. Leigh Fantacone-Campbell, Jeffrey A. Hooke, Albert J. Kovatich, Craig D. Shriver, John DiPersio, Bettina Drake, Ramaswamy Govindan, Sharon Heath, Timothy Ley, Brian Van Tine, Peter Westervelt, Mark A. Rubin, Jung Il Lee, Natália D. Aredes, and Armaz Mariamidze. Comprehensive Characterization of Cancer Driver Genes and Mutations. *Cell*, 173 (2):371–385.e18, April 2018. ISSN 00928674. doi: 10.1016/j.cell.2018.02.060. URL `https://linkinghub.elsevier.com/retrieve/pii/S009286741830237X`.

[68] Levi A. Garraway and Eric S. Lander. Lessons from the cancer genome. *Cell*, 153 (1):17–37, March 2013. ISSN 1097-4172. doi: 10.1016/j.cell.2013.03.002.

[69] The ICGC/TCGA Pan-Cancer Analysis of Whole Genomes Consortium. Pan-cancer analysis of whole genomes. *Nature*, 578(7793):82–93, February 2020. ISSN 0028-0836, 1476-4687. doi: 10.1038/s41586-020-1969-6. URL `http://www.nature.com/articles/s41586-020-1969-6`.

[70] Iñigo Martincorena, Keiran M. Raine, Moritz Gerstung, Kevin J. Dawson, Kerstin Haase, Peter Van Loo, Helen Davies, Michael R. Stratton, and Peter J. Campbell. Universal Patterns of Selection in Cancer and Somatic Tissues. *Cell*, 171(5):1029–1041.e21, November 2017. ISSN 1097-4172. doi: 10.1016/j.cell.2017.09.042.

[71] Ahrim Youn and Richard Simon. Identifying cancer driver genes in tumor genome sequencing studies. *Bioinformatics (Oxford, England)*, 27(2):175–181, January 2011. ISSN 1367-4811. doi: 10.1093/bioinformatics/btq630.

[72] Nathan D. Dees, Qunyuan Zhang, Cyriac Kandoth, Michael C. Wendl, William Schierding, Daniel C. Koboldt, Thomas B. Mooney, Matthew B. Callaway, David

Dooling, Elaine R. Mardis, Richard K. Wilson, and Li Ding. MuSiC: identifying mutational significance in cancer genomes. *Genome Research*, 22(8):1589–1598, August 2012. ISSN 1549-5469. doi: 10.1101/gr.134635.111.

[73] Michael S. Lawrence, Petar Stojanov, Paz Polak, Gregory V. Kryukov, Kristian Cibulskis, Andrey Sivachenko, Scott L. Carter, Chip Stewart, Craig H. Mermel, Steven A. Roberts, Adam Kiezun, Peter S. Hammerman, Aaron McKenna, Yotam Drier, Lihua Zou, Alex H. Ramos, Trevor J. Pugh, Nicolas Stransky, Elena Helman, Jaegil Kim, Carrie Sougnez, Lauren Ambrogio, Elizabeth Nickerson, Erica Shefler, Maria L. Cortés, Daniel Auclair, Gordon Saksena, Douglas Voet, Michael Noble, Daniel DiCara, Pei Lin, Lee Lichtenstein, David I. Heiman, Timothy Fennell, Marcin Imielinski, Bryan Hernandez, Eran Hodis, Sylvan Baca, Austin M. Dulak, Jens Lohr, Dan-Avi Landau, Catherine J. Wu, Jorge Melendez-Zajgla, Alfredo Hidalgo-Miranda, Amnon Koren, Steven A. McCarroll, Jaume Mora, Brian Crompton, Robert Onofrio, Melissa Parkin, Wendy Winckler, Kristin Ardlie, Stacey B. Gabriel, Charles W. M. Roberts, Jaclyn A. Biegel, Kimberly Stegmaier, Adam J. Bass, Levi A. Garraway, Matthew Meyerson, Todd R. Golub, Dmitry A. Gordenin, Shamil Sunyaev, Eric S. Lander, and Gad Getz. Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature*, 499(7457):214–218, July 2013. ISSN 1476-4687. doi: 10.1038/nature12213.

[74] Francisco Martínez-Jiménez, Ferran Muiños, Inés Sentís, Jordi Deu-Pons, Iker Reyes-Salazar, Claudia Arnedo-Pac, Loris Mularoni, Oriol Pich, Jose Bonet, Hanna Kranas, Abel Gonzalez-Perez, and Nuria Lopez-Bigas. A compendium of mutational cancer driver genes. *Nature Reviews Cancer*, 20(10):555–572, October 2020. ISSN 1474-175X, 1474-1768. doi: 10.1038/s41568-020-0290-x. URL `http://www.nature.com/articles/s41568-020-0290-x`.

[75] Chi V. Dang. MYC on the path to cancer. *Cell*, 149(1):22–35, March 2012. ISSN 1097-4172. doi: 10.1016/j.cell.2012.03.003.

[76] Claire M. Connell and Gary J. Doherty. Activating HER2 mutations as emerging targets in multiple solid cancers. *ESMO open*, 2(5):e000279, 2017. ISSN 2059-7029. doi: 10.1136/esmoopen-2017-000279.

[77] Ernst J. Kuipers, William M. Grady, David Lieberman, Thomas Seufferlein, Joseph J. Sung, Petra G. Boelens, Cornelis J. H. van de Velde, and Toshiaki Watanabe. Colorectal cancer. *Nature Reviews Disease Primers*, page 15065, November 2015. ISSN 2056-676X. doi: 10.1038/nrdp.2015.65. URL `http://www.nature.com/articles/nrdp201565`.

[78] Charlotte Philpott, Hannah Tovell, Ian M. Frayling, David N. Cooper, and Meena Upadhyaya. The NF1 somatic mutational landscape in sporadic human cancers. *Human Genomics*, 11(1):13, December 2017. ISSN 1479-7364. doi: 10.1186/s40246-017-0109-3. URL `http://humgenomics.biomedcentral.com/articles/10.1186/s40246-017-0109-3`.

[79] Francisco Sanchez-Vega, Marco Mina, Joshua Armenia, Walid K. Chatila, Augustin Luna, Konnor C. La, Sofia Dimitriadoy, David L. Liu, Havish S. Kantheti, Sadegh Saghafinia, Debyani Chakravarty, Foysal Daian, Qingsong Gao, Matthew H. Bailey,

Wen-Wei Liang, Steven M. Foltz, Ilya Shmulevich, Li Ding, Zachary Heins, Angelica Ochoa, Benjamin Gross, Jianjiong Gao, Hongxin Zhang, Ritika Kundra, Cyriac Kandoth, Istemi Bahceci, Leonard Dervishi, Ugur Dogrusoz, Wanding Zhou, Hui Shen, Peter W. Laird, Gregory P. Way, Casey S. Greene, Han Liang, Yonghong Xiao, Chen Wang, Antonio Iavarone, Alice H. Berger, Trever G. Bivona, Alexander J. Lazar, Gary D. Hammer, Thomas Giordano, Lawrence N. Kwong, Grant McArthur, Chenfei Huang, Aaron D. Tward, Mitchell J. Frederick, Frank McCormick, Matthew Meyerson, Eliezer M. Van Allen, Andrew D. Cherniack, Giovanni Ciriello, Chris Sander, Nikolaus Schultz, Samantha J. Caesar-Johnson, John A. Demchok, Ina Felau, Melpomeni Kasapi, Martin L. Ferguson, Carolyn M. Hutter, Heidi J. Sofia, Roy Tarnuzzer, Zhining Wang, Liming Yang, Jean C. Zenklusen, Jiashan (Julia) Zhang, Sudha Chudamani, Jia Liu, Laxmi Lolla, Rashi Naresh, Todd Pihl, Qiang Sun, Yunhu Wan, Ye Wu, Juok Cho, Timothy DeFreitas, Scott Frazer, Nils Gehlenborg, Gad Getz, David I. Heiman, Jaegil Kim, Michael S. Lawrence, Pei Lin, Sam Meier, Michael S. Noble, Gordon Saksena, Doug Voet, Hailei Zhang, Brady Bernard, Nyasha Chambwe, Varsha Dhankani, Theo Knijnenburg, Roger Kramer, Kalle Leinonen, Yuexin Liu, Michael Miller, Sheila Reynolds, Ilya Shmulevich, Vesteinn Thorsson, Wei Zhang, Rehan Akbani, Bradley M. Broom, Apurva M. Hegde, Zhenlin Ju, Rupa S. Kanchi, Anil Korkut, Jun Li, Han Liang, Shiyun Ling, Wenbin Liu, Yiling Lu, Gordon B. Mills, Kwok-Shing Ng, Arvind Rao, Michael Ryan, Jing Wang, John N. Weinstein, Jiexin Zhang, Adam Abeshouse, Joshua Armenia, Debyani Chakravarty, Walid K. Chatila, Ino de Bruijn, Jianjiong Gao, Benjamin E. Gross, Zachary J. Heins, Ritika Kundra, Konnor La, Marc Ladanyi, Augustin Luna, Moriah G. Nissan, Angelica Ochoa, Sarah M. Phillips, Ed Reznik, Francisco Sanchez-Vega, Chris Sander, Nikolaus Schultz, Robert Sheridan, S. Onur Sumer, Yichao Sun, Barry S. Taylor, Jioajiao Wang, Hongxin Zhang, Pavana Anur, Myron Peto, Paul Spellman, Christopher Benz, Joshua M. Stuart, Christopher K. Wong, Christina Yau, D. Neil Hayes, Joel S. Parker, Matthew D. Wilkerson, Adrian Ally, Miruna Balasundaram, Reanne Bowlby, Denise Brooks, Rebecca Carlsen, Eric Chuah, Noreen Dhalla, Robert Holt, Steven J.M. Jones, Katayoon Kasaian, Darlene Lee, Yussanne Ma, Marco A. Marra, Michael Mayo, Richard A. Moore, Andrew J. Mungall, Karen Mungall, A. Gordon Robertson, Sara Sadeghi, Jacqueline E. Schein, Payal Sipahimalani, Angela Tam, Nina Thiessen, Kane Tse, Tina Wong, Ashton C. Berger, Rameen Beroukhim, Andrew D. Cherniack, Carrie Cibulskis, Stacey B. Gabriel, Galen F. Gao, Gavin Ha, Matthew Meyerson, Steven E. Schumacher, Juliann Shih, Melanie H. Kucherlapati, Raju S. Kucherlapati, Stephen Baylin, Leslie Cope, Ludmila Danilova, Moiz S. Bootwalla, Phillip H. Lai, Dennis T. Maglinte, David J. Van Den Berg, Daniel J. Weisenberger, J. Todd Auman, Saianand Balu, Tom Bodenheimer, Cheng Fan, Katherine A. Hoadley, Alan P. Hoyle, Stuart R. Jefferys, Corbin D. Jones, Shaowu Meng, Piotr A. Mieczkowski, Lisle E. Mose, Amy H. Perou, Charles M. Perou, Jeffrey Roach, Yan Shi, Janae V. Simons, Tara Skelly, Matthew G. Soloway, Donghui Tan, Umadevi Veluvolu, Huihui Fan, Toshinori Hinoue, Peter W. Laird, Hui Shen, Wanding Zhou, Michelle Bel-

lair, Kyle Chang, Kyle Covington, Chad J. Creighton, Huyen Dinh, HarshaVardhan Doddapaneni, Lawrence A. Donehower, Jennifer Drummond, Richard A. Gibbs, Robert Glenn, Walker Hale, Yi Han, Jianhong Hu, Viktoriya Korchina, Sandra Lee, Lora Lewis, Wei Li, Xiuping Liu, Margaret Morgan, Donna Morton, Donna Muzny, Jireh Santibanez, Margi Sheth, Eve Shinbrot, Linghua Wang, Min Wang, David A. Wheeler, Liu Xi, Fengmei Zhao, Julian Hess, Elizabeth L. Appelbaum, Matthew Bailey, Matthew G. Cordes, Li Ding, Catrina C. Fronick, Lucinda A. Fulton, Robert S. Fulton, Cyriac Kandoth, Elaine R. Mardis, Michael D. McLellan, Christopher A. Miller, Heather K. Schmidt, Richard K. Wilson, Daniel Crain, Erin Curley, Johanna Gardner, Kevin Lau, David Mallery, Scott Morris, Joseph Paulauskis, Robert Penny, Candace Shelton, Troy Shelton, Mark Sherman, Eric Thompson, Peggy Yena, Jay Bowen, Julie M. Gastier-Foster, Mark Gerken, Kristen M. Leraas, Tara M. Lichtenberg, Nilsa C. Ramirez, Lisa Wise, Erik Zmuda, Niall Corcoran, Tony Costello, Christopher Hovens, Andre L. Carvalho, Ana C. de Carvalho, José H. Fregnani, Adhemar Longatto-Filho, Rui M. Reis, Cristovam Scapulatempo-Neto, Henrique C.S. Silveira, Daniel O. Vidal, Andrew Burnette, Jennifer Eschbacher, Beth Hermes, Ardene Noss, Rosy Singh, Matthew L. Anderson, Patricia D. Castro, Michael Ittmann, David Huntsman, Bernard Kohl, Xuan Le, Richard Thorp, Chris Andry, Elizabeth R. Duffy, Vladimir Lyadov, Oxana Paklina, Galiya Setdikova, Alexey Shabunin, Mikhail Tavobilov, Christopher McPherson, Ronald Warnick, Ross Berkowitz, Daniel Cramer, Colleen Feltmate, Neil Horowitz, Adam Kibel, Michael Muto, Chandrajit P. Raut, Andrei Malykh, Jill S. Barnholtz-Sloan, Wendi Barrett, Karen Devine, Jordonna Fulop, Quinn T. Ostrom, Kristen Shimmel, Yingli Wolinsky, Andrew E. Sloan, Agostino De Rose, Felice Giuliante, Marc Goodman, Beth Y. Karlan, Curt H. Hagedorn, John Eckman, Jodi Harr, Jerome Myers, Kelinda Tucker, Leigh Anne Zach, Brenda Deyarmin, Hai Hu, Leonid Kvecher, Caroline Larson, Richard J. Mural, Stella Somiari, Ales Vicha, Tomas Zelinka, Joseph Bennett, Mary Iacocca, Brenda Rabeno, Patricia Swanson, Mathieu Latour, Louis Lacombe, Bernard Têtu, Alain Bergeron, Mary McGraw, Susan M. Staugaitis, John Chabot, Hanina Hibshoosh, Antonia Sepulveda, Tao Su, Timothy Wang, Olga Potapova, Olga Voronina, Laurence Desjardins, Odette Mariani, Sergio Roman-Roman, Xavier Sastre, Marc-Henri Stern, Feixiong Cheng, Sabina Signoretti, Andrew Berchuck, Darell Bigner, Eric Lipp, Jeffrey Marks, Shannon McCall, Roger McLendon, Angeles Secord, Alexis Sharp, Madhusmita Behera, Daniel J. Brat, Amy Chen, Keith Delman, Seth Force, Fadlo Khuri, Kelly Magliocca, Shishir Maithel, Jeffrey J. Olson, Taofeek Owonikoko, Alan Pickens, Suresh Ramalingam, Dong M. Shin, Gabriel Sica, Erwin G. Van Meir, Hongzheng Zhang, Wil Eijckenboom, Ad Gillis, Esther Korpershoek, Leendert Looijenga, Wolter Oosterhuis, Hans Stoop, Kim E. van Kessel, Ellen C. Zwarthoff, Chiara Calatozzolo, Lucia Cuppini, Stefania Cuzzubbo, Francesco DiMeco, Gaetano Finocchiaro, Luca Mattei, Alessandro Perin, Bianca Pollo, Chu Chen, John Houck, Pawadee Lohavanichbutr, Arndt Hartmann, Christine Stoehr, Robert Stoehr, Helge Taubert, Sven Wach, Bernd Wullich, Witold Kycler, Dawid Murawa, Maciej Wiznerowicz, Ki Chung,

W. Jeffrey Edenfield, Julie Martin, Eric Baudin, Glenn Bubley, Raphael Bueno, Assunta De Rienzo, William G. Richards, Steven Kalkanis, Tom Mikkelsen, Houtan Noushmehr, Lisa Scarpace, Nicolas Girard, Marta Aymerich, Elias Campo, Eva Giné, Armando López Guillermo, Nguyen Van Bang, Phan Thi Hanh, Bui Duc Phu, Yufang Tang, Howard Colman, Kimberley Evason, Peter R. Dottino, John A. Martignetti, Hani Gabra, Hartmut Juhl, Teniola Akeredolu, Serghei Stepa, Dave Hoon, Keunsoo Ahn, Koo Jeong Kang, Felix Beuschlein, Anne Breggia, Michael Birrer, Debra Bell, Mitesh Borad, Alan H. Bryce, Erik Castle, Vishal Chandan, John Cheville, John A. Copland, Michael Farnell, Thomas Flotte, Nasra Giama, Thai Ho, Michael Kendrick, Jean-Pierre Kocher, Karla Kopp, Catherine Moser, David Nagorney, Daniel O'Brien, Brian Patrick O'Neill, Tushar Patel, Gloria Petersen, Florencia Que, Michael Rivera, Lewis Roberts, Robert Smallridge, Thomas Smyrk, Melissa Stanton, R. Houston Thompson, Michael Torbenson, Ju Dong Yang, Lizhi Zhang, Fadi Brimo, Jaffer A. Ajani, Ana Maria Angulo Gonzalez, Carmen Behrens, Jolanta Bondaruk, Russell Broaddus, Bogdan Czerniak, Bita Esmaeli, Junya Fujimoto, Jeffrey Gershenwald, Charles Guo, Alexander J. Lazar, Christopher Logothetis, Funda Meric-Bernstam, Cesar Moran, Lois Ramondetta, David Rice, Anil Sood, Pheroze Tamboli, Timothy Thompson, Patricia Troncoso, Anne Tsao, Ignacio Wistuba, Candace Carter, Lauren Haydu, Peter Hersey, Valerie Jakrot, Hojabr Kakavand, Richard Kefford, Kenneth Lee, Georgina Long, Graham Mann, Michael Quinn, Robyn Saw, Richard Scolyer, Kerwin Shannon, Andrew Spillane, Jonathan Stretch, Maria Synott, John Thompson, James Wilmott, Hikmat Al-Ahmadie, Timothy A. Chan, Ronald Ghossein, Anuradha Gopalan, Douglas A. Levine, Victor Reuter, Samuel Singer, Bhuvanesh Singh, Nguyen Viet Tien, Thomas Broudy, Cyrus Mirsaidi, Praveen Nair, Paul Drwiega, Judy Miller, Jennifer Smith, Howard Zaren, Joong-Won Park, Nguyen Phi Hung, Electron Kebebew, W. Marston Linehan, Adam R. Metwalli, Karel Pacak, Peter A. Pinto, Mark Schiffman, Laura S. Schmidt, Cathy D. Vocke, Nicolas Wentzensen, Robert Worrell, Hannah Yang, Marc Moncrieff, Chandra Goparaju, Jonathan Melamed, Harvey Pass, Natalia Botnariuc, Irina Caraman, Mircea Cernat, Inga Chemencedji, Adrian Clipca, Serghei Doruc, Ghenadie Gorincioi, Sergiu Mura, Maria Pirtac, Irina Stancul, Diana Tcaciuc, Monique Albert, Iakovina Alexopoulou, Angel Arnaout, John Bartlett, Jay Engel, Sebastien Gilbert, Jeremy Parfitt, Harman Sekhon, George Thomas, Doris M. Rassl, Robert C. Rintoul, Carlo Bifulco, Raina Tamakawa, Walter Urba, Nicholas Hayward, Henri Timmers, Anna Antenucci, Francesco Facciolo, Gianluca Grazi, Mirella Marino, Roberta Merola, Ronald de Krijger, Anne-Paule Gimenez-Roqueplo, Alain Piché, Simone Chevalier, Ginette McKercher, Kivanc Birsoy, Gene Barnett, Cathy Brewer, Carol Farver, Theresa Naska, Nathan A. Pennell, Daniel Raymond, Cathy Schilero, Kathy Smolenski, Felicia Williams, Carl Morrison, Jeffrey A. Borgia, Michael J. Liptay, Mark Pool, Christopher W. Seder, Kerstin Junker, Larsson Omberg, Mikhail Dinkin, George Manikhas, Domenico Alvaro, Maria Consiglia Bragazzi, Vincenzo Cardinale, Guido Carpino, Eugenio Gaudio, David Chesla, Sandra Cottingham, Michael Dubina, Fedor Moiseenko, Renumathy Dhanasekaran, Karl-Friedrich Becker, Klaus-

Peter Janssen, Julia Slotta-Huspenina, Mohamed H. Abdel-Rahman, Dina Aziz, Sue Bell, Colleen M. Cebulla, Amy Davis, Rebecca Duell, J. Bradley Elder, Joe Hilty, Bahavna Kumar, James Lang, Norman L. Lehman, Randy Mandt, Phuong Nguyen, Robert Pilarski, Karan Rai, Lynn Schoenfield, Kelly Senecal, Paul Wakely, Paul Hansen, Ronald Lechan, James Powers, Arthur Tischler, William E. Grizzle, Katherine C. Sexton, Alison Kastl, Joel Henderson, Sima Porten, Jens Waldmann, Martin Fassnacht, Sylvia L. Asa, Dirk Schadendorf, Marta Couce, Markus Graefen, Hartwig Huland, Guido Sauter, Thorsten Schlomm, Ronald Simon, Pierre Tennstedt, Oluwole Olabode, Mark Nelson, Oliver Bathe, Peter R. Carroll, June M. Chan, Philip Disaia, Pat Glenn, Robin K. Kelley, Charles N. Landen, Joanna Phillips, Michael Prados, Jeffry Simko, Karen Smith-McCune, Scott VandenBerg, Kevin Roggin, Ashley Fehrenbach, Ady Kendler, Suzanne Sifri, Ruth Steele, Antonio Jimeno, Francis Carey, Ian Forgie, Massimo Mannelli, Michael Carney, Brenda Hernandez, Benito Campos, Christel Herold-Mende, Christin Jungk, Andreas Unterberg, Andreas von Deimling, Aaron Bossler, Joseph Galbraith, Laura Jacobus, Michael Knudson, Tina Knutson, Deqin Ma, Mohammed Milhem, Rita Sigmund, Andrew K. Godwin, Rashna Madan, Howard G. Rosenthal, Clement Adebamowo, Sally N. Adebamowo, Alex Boussioutas, David Beer, Thomas Giordano, Anne-Marie Mes-Masson, Fred Saad, Therese Bocklage, Lisa Landrum, Robert Mannel, Kathleen Moore, Katherine Moxley, Russel Postier, Joan Walker, Rosemary Zuna, Michael Feldman, Federico Valdivieso, Rajiv Dhir, James Luketich, Edna M. Mora Pinero, Mario Quintero-Aguilo, Carlos Gilberto Carlotti, Jose Sebastião Dos Santos, Rafael Kemp, Ajith Sankarankuty, Daniela Tirapelli, James Catto, Kathy Agnew, Elizabeth Swisher, Jenette Creaney, Bruce Robinson, Carl Simon Shelley, Eryn M. Godwin, Sara Kendall, Cassaundra Shipman, Carol Bradford, Thomas Carey, Andrea Haddad, Jeffey Moyer, Lisa Peterson, Mark Prince, Laura Rozek, Gregory Wolf, Rayleen Bowman, Kwun M. Fong, Ian Yang, Robert Korst, W. Kimryn Rathmell, J. Leigh Fantacone-Campbell, Jeffrey A. Hooke, Albert J. Kovatich, Craig D. Shriver, John DiPersio, Bettina Drake, Ramaswamy Govindan, Sharon Heath, Timothy Ley, Brian Van Tine, Peter Westervelt, Mark A. Rubin, Jung Il Lee, Natália D. Aredes, and Armaz Mariamidze. Oncogenic Signaling Pathways in The Cancer Genome Atlas. *Cell*, 173(2):321–337.e10, April 2018. ISSN 00928674. doi: 10.1016/j.cell.2018.03.035. URL `https://linkinghub.elsevier.com/retrieve/pii/S0092867418303593`.

[80] Hyuna Sung, Jacques Ferlay, Rebecca L. Siegel, Mathieu Laversanne, Isabelle Soerjomataram, Ahmedin Jemal, and Freddie Bray. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians*, page caac.21660, February 2021. ISSN 0007-9235, 1542-4863. doi: 10.3322/caac.21660. URL `https://onlinelibrary.wiley.com/doi/10.3322/caac.21660`.

[81] Maggie C. U. Cheang, Stephen K. Chia, David Voduc, Dongxia Gao, Samuel Leung, Jacqueline Snider, Mark Watson, Sherri Davies, Philip S. Bernard, Joel S. Parker, Charles M. Perou, Matthew J. Ellis, and Torsten O. Nielsen. Ki67 Index, HER2 Status, and Prognosis of Patients With Luminal B Breast Cancer. *JNCI: Journal of*

the National Cancer Institute, 101(10):736–750, May 2009. ISSN 0027-8874, 1460-2105. doi: 10.1093/jnci/djp082. URL https://academic.oup.com/jnci/article-lookup/doi/10.1093/jnci/djp082.

[82] C. M. Perou, T. Sørlie, M. B. Eisen, M. van de Rijn, S. S. Jeffrey, C. A. Rees, J. R. Pollack, D. T. Ross, H. Johnsen, L. A. Akslen, O. Fluge, A. Pergamenschikov, C. Williams, S. X. Zhu, P. E. Lønning, A. L. Børresen-Dale, P. O. Brown, and D. Botstein. Molecular portraits of human breast tumours. *Nature*, 406(6797): 747–752, August 2000. ISSN 0028-0836. doi: 10.1038/35021093.

[83] Joel S. Parker, Michael Mullins, Maggie C. U. Cheang, Samuel Leung, David Voduc, Tammi Vickery, Sherri Davies, Christiane Fauron, Xiaping He, Zhiyuan Hu, John F. Quackenbush, Inge J. Stijleman, Juan Palazzo, J. S. Marron, Andrew B. Nobel, Elaine Mardis, Torsten O. Nielsen, Matthew J. Ellis, Charles M. Perou, and Philip S. Bernard. Supervised risk predictor of breast cancer based on intrinsic subtypes. *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology*, 27(8):1160–1167, March 2009. ISSN 1527-7755. doi: 10.1200/JCO.2008.18.1370.

[84] Christos Sotiriou and Lajos Pusztai. Gene-Expression Signatures in Breast Cancer. *New England Journal of Medicine*, 360(8):790–800, February 2009. ISSN 0028-4793, 1533-4406. doi: 10.1056/NEJMra0801289. URL http://www.nejm.org/doi/abs/10.1056/NEJMra0801289.

[85] Saber Fallahpour, Tanya Navaneelan, Prithwish De, and Alessia Borgo. Breast cancer survival by molecular subtype: a population-based analysis of cancer registry data. *CMAJ open*, 5(3):E734–E739, September 2017. ISSN 2291-0026. doi: 10.9778/cmajo.20170030.

[86] Nadia Howlader, Kathleen A. Cronin, Allison W. Kurian, and Rebecca Andridge. Differences in Breast Cancer Survival by Molecular Subtypes in the United States. *Cancer Epidemiology Biomarkers & Prevention*, 27(6):619–626, June 2018. ISSN 1055-9965, 1538-7755. doi: 10.1158/1055-9965.EPI-17-0627. URL http://cebp.aacrjournals.org/lookup/doi/10.1158/1055-9965.EPI-17-0627.

[87] Jaime Prat. Ovarian carcinomas: five distinct diseases with different origins, genetic alterations, and clinicopathological features. *Virchows Archiv: An International Journal of Pathology*, 460(3):237–249, March 2012. ISSN 1432-2307. doi: 10.1007/s00428-012-1203-5.

[88] W. Glenn McCluggage. Morphological subtypes of ovarian carcinoma: a review with emphasis on new developments and pathogenesis. *Pathology*, 43(5):420–432, August 2011. ISSN 1465-3931. doi: 10.1097/PAT.0b013e328348a6e7.

[89] Russell Vang, Ie-Ming Shih, and Robert J. Kurman. Fallopian tube precursors of ovarian low- and high-grade serous neoplasms. *Histopathology*, 62(1):44–58, January 2013. ISSN 1365-2559. doi: 10.1111/his.12046.

[90] Zhiyuan Hu, Mara Artibani, Abdulkhaliq Alsaadi, Nina Wietek, Matteo Morotti, Tingyan Shi, Zhe Zhong, Laura Santana Gonzalez, Salma El-Sahhar, Mohammad KaramiNejadRanjbar, Garry Mallett, Yun Feng, Kenta Masuda, Yiyan

Zheng, Kay Chong, Stephen Damato, Sunanda Dhar, Leticia Campo, Riccardo Garruto Campanile, Hooman Soleymani majd, Vikram Rai, David Maldonado-Perez, Stephanie Jones, Vincenzo Cerundolo, Tatjana Sauka-Spengler, Christopher Yau, and Ahmed Ashour Ahmed. The Repertoire of Serous Ovarian Cancer Non-genetic Heterogeneity Revealed by Single-Cell Sequencing of Normal Fallopian Tube Epithelial Cells. *Cancer Cell*, 37(2):226–242.e7, February 2020. ISSN 15356108. doi: 10.1016/j.ccell.2020.01.003. URL `https://linkinghub.elsevier.com/retrieve/pii/S1535610820300428`.

[91] Ioannis Gounaris and James D. Brenton. Molecular pathogenesis of ovarian clear cell carcinoma. *Future Oncology (London, England)*, 11(9):1389–1405, 2015. ISSN 1744-8301. doi: 10.2217/fon.15.45.

[92] Yiying Wang, Maggie Mang, Yue Wang, Lijie Wang, Robert Klein, Beihua Kong, and Wenxin Zheng. Tubal origin of ovarian endometriosis and clear cell and endometrioid carcinoma. *American Journal of Cancer Research*, 5(3):869–879, 2015. ISSN 2156-6976.

[93] Jubilee Brown and Michael Frumovitz. Mucinous tumors of the ovary: current thoughts on diagnosis and management. *Current Oncology Reports*, 16(6):389, June 2014. ISSN 1534-6269. doi: 10.1007/s11912-014-0389-x.

[94] D. Bell, A. Berchuck, M. Birrer, J. Chien, D. W. Cramer, F. Dao, R. Dhir, P. DiSaia, H. Gabra, P. Glenn, A. K. Godwin, J. Gross, L. Hartmann, M. Huang, D. G. Huntsman, M. Iacocca, M. Imielinski, S. Kalloger, B. Y. Karlan, D. A. Levine, G. B. Mills, C. Morrison, D. Mutch, N. Olvera, S. Orsulic, K. Park, N. Petrelli, B. Rabeno, J. S. Rader, B. I. Sikic, K. Smith-McCune, A. K. Sood, D. Bowtell, R. Penny, J. R. Testa, K. Chang, H. H. Dinh, J. A. Drummond, G. Fowler, P. Gunaratne, A. C. Hawes, C. L. Kovar, L. R. Lewis, M. B. Morgan, I. F. Newsham, J. Santibanez, J. G. Reid, L. R. Trevino, Y.-Q. Wu, M. Wang, D. M. Muzny, D. A. Wheeler, R. A. Gibbs, G. Getz, M. S. Lawrence, K. Cibulskis, A. Y. Sivachenko, C. Sougnez, D. Voet, J. Wilkinson, T. Bloom, K. Ardlie, T. Fennell, J. Baldwin, S. Gabriel, E. S. Lander, L. Ding, R. S. Fulton, D. C. Koboldt, M. D. McLellan, T. Wylie, J. Walker, M. O'Laughlin, D. J. Dooling, L. Fulton, R. Abbott, N. D. Dees, Q. Zhang, C. Kandoth, M. Wendl, W. Schierding, D. Shen, C. C. Harris, H. Schmidt, J. Kalicki, K. D. Delehaunty, C. C. Fronick, R. Demeter, L. Cook, J. W. Wallis, L. Lin, V. J. Magrini, J. S. Hodges, J. M. Eldred, S. M. Smith, C. S. Pohl, F. Vandin, B. J. Raphael, G. M. Weinstock, E. R. Mardis, R. K. Wilson, M. Meyerson, W. Winckler, G. Getz, R. G. W. Verhaak, S. L. Carter, C. H. Mermel, G. Saksena, H. Nguyen, R. C. Onofrio, M. S. Lawrence, D. Hubbard, S. Gupta, A. Crenshaw, A. H. Ramos, K. Ardlie, L. Chin, A. Protopopov, Juinhua Zhang, T. M. Kim, I. Perna, Y. Xiao, H. Zhang, G. Ren, N. Sathiamoorthy, R. W. Park, E. Lee, P. J. Park, R. Kucherlapati, D. M. Absher, L. Waite, G. Sherlock, J. D. Brooks, J. Z. Li, J. Xu, R. M. Myers, P. W. Laird, L. Cope, J. G. Herman, H. Shen, D. J. Weisenberger, H. Noushmehr, F. Pan, T. Triche Jr, B. P. Berman, D. J. Van Den Berg, J. Buckley, S. B. Baylin, P. T. Spellman, E. Purdom, P. Neuvial, H. Bengtsson, L. R. Jakkula, S. Durinck, J. Han, S. Dorton, H. Marr, Y. G. Choi,

V. Wang, N. J. Wang, J. Ngai, J. G. Conboy, B. Parvin, H. S. Feiler, T. P. Speed, J. W. Gray, D. A. Levine, N. D. Socci, Y. Liang, B. S. Taylor, N. Schultz, L. Borsu, A. E. Lash, C. Brennan, A. Viale, C. Sander, M. Ladanyi, K. A. Hoadley, S. Meng, Y. Du, Y. Shi, L. Li, Y. J. Turman, D. Zang, E. B. Helms, S. Balu, X. Zhou, J. Wu, M. D. Topal, D. N. Hayes, C. M. Perou, G. Getz, D. Voet, G. Saksena, Junihua Zhang, H. Zhang, C. J. Wu, S. Shukla, K. Cibulskis, M. S. Lawrence, A. Sivachenko, R. Jing, R. W. Park, Y. Liu, P. J. Park, M. Noble, L. Chin, H. Carter, D. Kim, R. Karchin, P. T. Spellman, E. Purdom, P. Neuvial, H. Bengtsson, S. Durinck, J. Han, J. E. Korkola, L. M. Heiser, R. J. Cho, Z. Hu, B. Parvin, T. P. Speed, J. W. Gray, N. Schultz, E. Cerami, B. S. Taylor, A. Olshen, B. Reva, Y. Antipin, R. Shen, P. Mankoo, R. Sheridan, G. Ciriello, W. K. Chang, J. A. Bernanke, L. Borsu, D. A. Levine, M. Ladanyi, C. Sander, D. Haussler, C. C. Benz, J. M. Stuart, S. C. Benz, J. Z. Sanborn, C. J. Vaske, J. Zhu, C. Szeto, G. K. Scott, C. Yau, K. A. Hoadley, Y. Du, S. Balu, D. N. Hayes, C. M. Perou, M. D. Wilkerson, N. Zhang, R. Akbani, K. A. Baggerly, W. K. Yung, G. B. Mills, J. N. Weinstein, R. Penny, T. Shelton, D. Grimm, M. Hatfield, S. Morris, P. Yena, P. Rhodes, M. Sherman, J. Paulauskis, S. Millis, A. Kahn, J. M. Greene, R. Sfeir, M. A. Jensen, J. Chen, J. Whitmore, S. Alonso, J. Jordan, A. Chu, Jinghui Zhang, A. Barker, C. Compton, G. Eley, M. Ferguson, P. Fielding, D. S. Gerhard, R. Myles, C. Schaefer, K. R. Mills Shaw, J. Vaught, J. B. Vockley, P. J. Good, M. S. Guyer, B. Ozenberger, J. Peterson, and E. Thomson. Integrated genomic analyses of ovarian carcinoma. *Nature*, 474(7353): 609–615, June 2011. ISSN 0028-0836, 1476-4687. doi: 10.1038/nature10166. URL `http://www.nature.com/doifinder/10.1038/nature10166`.

[95] Robert C. Bast, Ursula A. Matulonis, Anil K. Sood, Ahmed A. Ahmed, Adaobi E. Amobi, Frances R. Balkwill, Monicka Wielgos-Bonvallet, David D. L. Bowtell, James D. Brenton, Joan S. Brugge, Robert L. Coleman, Giulio F. Draetta, Kai Doberstein, Ronny I. Drapkin, Mark A. Eckert, Robert P. Edwards, Kevin M. Elias, Darren Ennis, Andrew Futreal, David M. Gershenson, Roger A. Greenberg, David G. Huntsman, Jennifer Xiao Ye Ji, Elise C. Kohn, Claudia Iavarone, Ernst R. Lengyel, Douglas A. Levine, Christopher J. Lord, Zhen Lu, Gordon B. Mills, Francesmary Modugno, Brad H. Nelson, Kunle Odunsi, Jessica A. Pilsworth, Robert K. Rottapel, Daniel J. Powell, Li Shen, Ie-Ming Shih, David R. Spriggs, Josephine Walton, Kaiyang Zhang, Rugang Zhang, and Lee Zou. Critical questions in ovarian cancer research and treatment: Report of an American Association for Cancer Research Special Conference. *Cancer*, 125(12):1963–1972, June 2019. ISSN 0008-543X, 1097-0142. doi: 10.1002/cncr.32004. URL `https://onlinelibrary.wiley.com/doi/abs/10.1002/cncr.32004`.

[96] Ahmed Ashour Ahmed, Dariush Etemadmoghadam, Jillian Temple, Andy G Lynch, Mohamed Riad, Raghwa Sharma, Colin Stewart, Sian Fereday, Carlos Caldas, Anna deFazio, David Bowtell, and James D Brenton. Driver mutations in *TP53* are ubiquitous in high grade serous carcinoma of the ovary: *TP53* mutation in high-grade pelvic serous carcinoma. *The Journal of Pathology*, 221(1):49–56, May 2010.

ISSN 00223417. doi: 10.1002/path.2696. URL http://doi.wiley.com/10.1002/path.2696.

[97] Gad Singer, Robert Stöhr, Leslie Cope, Reiko Dehari, Arndt Hartmann, Deng-Fan Cao, Tian-Li Wang, Robert J. Kurman, and Ie-Ming Shih. Patterns of p53 mutations separate ovarian serous borderline tumors and low- and high-grade carcinomas and provide support for a new model of ovarian carcinogenesis: a mutational analysis with immunohistochemical correlation. *The American Journal of Surgical Pathology*, 29(2):218–224, February 2005. ISSN 0147-5185. doi: 10.1097/01.pas.0000146025.91953.8d.

[98] Russell Vang, Douglas A. Levine, Robert A. Soslow, Charles Zaloudek, Ie-Ming Shih, and Robert J. Kurman. Molecular Alterations of TP53 are a Defining Feature of Ovarian High-Grade Serous Carcinoma: A Rereview of Cases Lacking TP53 Mutations in The Cancer Genome Atlas Ovarian Study. *International Journal of Gynecological Pathology: Official Journal of the International Society of Gynecological Pathologists*, 35(1):48–55, January 2016. ISSN 1538-7151. doi: 10.1097/PGP.0000000000000207.

[99] E. S. Ho, C. R. Lai, Y. T. Hsieh, J. T. Chen, A. J. Lin, M. H. Hung, and F. S. Liu. p53 mutation is infrequent in clear cell carcinoma of the ovary. *Gynecologic Oncology*, 80(2):189–193, February 2001. ISSN 0090-8258. doi: 10.1006/gyno.2000.6025.

[100] Kimberly C. Wiegand, Sohrab P. Shah, Osama M. Al-Agha, Yongjun Zhao, Kane Tse, Thomas Zeng, Janine Senz, Melissa K. McConechy, Michael S. Anglesio, Steve E. Kalloger, Winnie Yang, Alireza Heravi-Moussavi, Ryan Giuliany, Christine Chow, John Fee, Abdalnasser Zayed, Leah Prentice, Nataliya Melnyk, Gulisa Turashvili, Allen D. Delaney, Jason Madore, Stephen Yip, Andrew W. McPherson, Gavin Ha, Lynda Bell, Sian Fereday, Angela Tam, Laura Galletta, Patricia N. Tonin, Diane Provencher, Dianne Miller, Steven J. M. Jones, Richard A. Moore, Gregg B. Morin, Arusha Oloumi, Niki Boyd, Samuel A. Aparicio, Ie-Ming Shih, Anne-Marie Mes-Masson, David D. Bowtell, Martin Hirst, Blake Gilks, Marco A. Marra, and David G. Huntsman. ARID1A mutations in endometriosis-associated ovarian carcinomas. *The New England Journal of Medicine*, 363(16):1532–1543, October 2010. ISSN 1533-4406. doi: 10.1056/NEJMoa1008433.

[101] Georgina L. Ryland, Sally M. Hunter, Maria A. Doyle, Franco Caramia, Jason Li, Simone M. Rowley, Michael Christie, Prue E. Allan, Andrew N. Stephens, David D. L. Bowtell, Australian Ovarian Cancer Study Group, Ian G. Campbell, and Kylie L. Gorringe. Mutational landscape of mucinous ovarian carcinoma and its neoplastic precursors. *Genome Medicine*, 7(1):87, 2015. ISSN 1756-994X. doi: 10.1186/s13073-015-0210-y.

[102] Tania Moujaber, Dariush Etemadmoghadam, Cristina Mapagu, Catherine Kennedy, Yoke-Eng Chiew, Casina Kan, Nikilyn Nevins, Sivatharsny Srirangan, Sian Fereday, Nadia Traficante, Australian Ovarian Cancer Study group, David Bowtell, Rosemary Balleine, Paul Harnett, and Anna deFazio. Abstract 2584: Mutations in low-grade serous ovarian cancer and response to *BRAF* and *MEK* inhibitors. In *Clinical Research (Excluding Clinical Trials)*, pages 2584–2584. American Association for

Cancer Research, July 2018. doi: 10.1158/1538-7445.AM2018-2584. URL http://cancerres.aacrjournals.org/lookup/doi/10.1158/1538-7445.AM2018-2584.

[103] Jacques Ferlay, Isabelle Soerjomataram, Rajesh Dikshit, Sultan Eser, Colin Mathers, Marise Rebelo, Donald Maxwell Parkin, David Forman, and Freddie Bray. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *International Journal of Cancer*, 136(5):E359–386, March 2015. ISSN 1097-0215. doi: 10.1002/ijc.29210.

[104] Freddie Bray, Jacques Ferlay, Isabelle Soerjomataram, Rebecca L. Siegel, Lindsey A. Torre, and Ahmedin Jemal. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians*, 68(6):394–424, November 2018. ISSN 00079235. doi: 10.3322/caac.21492. URL http://doi.wiley.com/10.3322/caac.21492.

[105] The Cancer Genome Atlas Network. Comprehensive molecular characterization of human colon and rectal cancer. *Nature*, 487(7407):330–337, July 2012. ISSN 0028-0836, 1476-4687. doi: 10.1038/nature11252. URL http://www.nature.com/articles/nature11252.

[106] Justin Guinney, Rodrigo Dienstmann, Xin Wang, Aurélien de Reyniès, Andreas Schlicker, Charlotte Soneson, Laetitia Marisa, Paul Roepman, Gift Nyamundanda, Paolo Angelino, Brian M Bot, Jeffrey S Morris, Iris M Simon, Sarah Gerster, Evelyn Fessler, Felipe De Sousa E Melo, Edoardo Missiaglia, Hena Ramay, David Barras, Krisztian Homicsko, Dipen Maru, Ganiraju C Manyam, Bradley Broom, Valerie Boige, Beatriz Perez-Villamil, Ted Laderas, Ramon Salazar, Joe W Gray, Douglas Hanahan, Josep Tabernero, Rene Bernards, Stephen H Friend, Pierre Laurent-Puig, Jan Paul Medema, Anguraj Sadanandam, Lodewyk Wessels, Mauro Delorenzi, Scott Kopetz, Louis Vermeulen, and Sabine Tejpar. The consensus molecular subtypes of colorectal cancer. *Nature Medicine*, 21(11):1350–1356, November 2015. ISSN 1078-8956, 1546-170X. doi: 10.1038/nm.3967. URL http://www.nature.com/articles/nm.3967.

[107] Rodrigo Dienstmann, Louis Vermeulen, Justin Guinney, Scott Kopetz, Sabine Tejpar, and Josep Tabernero. Consensus molecular subtypes and the evolution of precision medicine in colorectal cancer. *Nature Reviews Cancer*, 17(2):79–92, February 2017. ISSN 1474-175X, 1474-1768. doi: 10.1038/nrc.2016.126. URL http://www.nature.com/articles/nrc.2016.126.

[108] R. Gryfe, H. Kim, E. T. Hsieh, M. D. Aronson, E. J. Holowaty, S. B. Bull, M. Redston, and S. Gallinger. Tumor microsatellite instability and clinical outcome in young patients with colorectal cancer. *The New England Journal of Medicine*, 342 (2):69–77, January 2000. ISSN 0028-4793. doi: 10.1056/NEJM200001133420201.

[109] Thomas F. Gajewski, Hans Schreiber, and Yang-Xin Fu. Innate and adaptive immune cells in the tumor microenvironment. *Nature Immunology*, 14(10):1014–1022, October 2013. ISSN 1529-2916. doi: 10.1038/ni.2703.

[110] A. Karolina Palucka and Lisa M. Coussens. The Basis of Oncoimmunology. *Cell*, 164(6):1233–1247, March 2016. ISSN 1097-4172. doi: 10.1016/j.cell.2016.01.049.

[111] J. Galon. Type, Density, and Location of Immune Cells Within Human Colorectal Tumors Predict Clinical Outcome. *Science*, 313(5795):1960–1964, September 2006. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.1129139. URL https://www.sciencemag.org/lookup/doi/10.1126/science.1129139.

[112] Jérôme Galon, Wolf-Herman Fridman, and Franck Pagès. The Adaptive Immunologic Microenvironment in Colorectal Cancer: A Novel Perspective: Figure 1. *Cancer Research*, 67(5):1883–1886, March 2007. ISSN 0008-5472, 1538-7445. doi: 10.1158/0008-5472.CAN-06-4806. URL http://cancerres.aacrjournals.org/lookup/doi/10.1158/0008-5472.CAN-06-4806.

[113] Bernhard Mlecnik, Gabriela Bindea, Helen K. Angell, Pauline Maby, Mihaela Angelova, David Tougeron, Sarah E. Church, Lucie Lafontaine, Maria Fischer, Tessa Fredriksen, Maristella Sasso, Amélie M. Bilocq, Amos Kirilovsky, Anna C. Obenauf, Mohamad Hamieh, Anne Berger, Patrick Bruneval, Jean-Jacques Tuech, Jean-Christophe Sabourin, Florence Le Pessot, Jacques Mauillon, Arash Rafii, Pierre Laurent-Puig, Michael R. Speicher, Zlatko Trajanoski, Pierre Michel, Richard Sesboüe, Thierry Frebourg, Franck Pagès, Viia Valge-Archer, Jean-Baptiste Latouche, and Jérôme Galon. Integrative Analyses of Colorectal Cancer Show Immunoscore Is a Stronger Predictor of Patient Survival Than Microsatellite Instability. *Immunity*, 44(3):698–711, March 2016. ISSN 10747613. doi: 10.1016/j.immuni.2016.02.025. URL https://linkinghub.elsevier.com/retrieve/pii/S1074761316300644.

[114] Franck Pagès, Bernhard Mlecnik, Florence Marliot, Gabriela Bindea, Fang-Shu Ou, Carlo Bifulco, Alessandro Lugli, Inti Zlobec, Tilman T Rau, Martin D Berger, Iris D Nagtegaal, Elisa Vink-Börger, Arndt Hartmann, Carol Geppert, Julie Kolwelter, Susanne Merkel, Robert Grützmann, Marc Van den Eynde, Anne Jouret-Mourin, Alex Kartheuser, Daniel Léonard, Christophe Remue, Julia Y Wang, Prashant Bavi, Michael H A Roehrl, Pamela S Ohashi, Linh T Nguyen, SeongJun Han, Heather L MacGregor, Sara Hafezi-Bakhtiari, Bradly G Wouters, Giuseppe V Masucci, Emilia K Andersson, Eva Zavadova, Michal Vocka, Jan Spacek, Lubos Petruzelka, Bohuslav Konopasek, Pavel Dundr, Helena Skalova, Kristyna Nemejcova, Gerardo Botti, Fabiana Tatangelo, Paolo Delrio, Gennaro Ciliberto, Michele Maio, Luigi Laghi, Fabio Grizzi, Tessa Fredriksen, Bénédicte Buttard, Mihaela Angelova, Angela Vasaturo, Pauline Maby, Sarah E Church, Helen K Angell, Lucie Lafontaine, Daniela Bruni, Carine El Sissy, Nacilla Haicheur, Amos Kirilovsky, Anne Berger, Christine Lagorce, Jeffrey P Meyers, Christopher Paustian, Zipei Feng, Carmen Ballesteros-Merino, Jeroen Dijkstra, Carlijn van de Water, Shannon van Lent-van Vliet, Nikki Knijn, Ana-Maria Mușină, Dragos-Viorel Scripcariu, Boryana Popivanova, Mingli Xu, Tomonobu Fujita, Shoichi Hazama, Nobuaki Suzuki, Hiroaki Nagano, Kiyotaka Okuno, Toshihiko Torigoe, Noriyuki Sato, Tomohisa Furuhata, Ichiro Takemasa, Kyogo Itoh, Prabhu S Patel, Hemangini H Vora, Birva Shah, Jayendrakumar B Patel, Kruti N Rajvik, Shashank J Pandya, Shilin N Shukla, Yili Wang, Guanjun Zhang, Yutaka Kawakami, Francesco M Marincola, Paolo A Ascierto, Daniel J Sargent, Bernard A Fox, and Jérôme Ga-

lon. International validation of the consensus Immunoscore for the classification of colon cancer: a prognostic and accuracy study. *The Lancet*, 391(10135):2128–2139, May 2018. ISSN 01406736. doi: 10.1016/S0140-6736(18)30789-X. URL https://linkinghub.elsevier.com/retrieve/pii/S014067361830789X.

[115] Johnny R. Ramroop, Madelyn M. Gerber, and Amanda Ewart Toland. Germline Variants Impact Somatic Events during Tumorigenesis. *Trends in Genetics*, 35(7): 515–526, July 2019. ISSN 01689525. doi: 10.1016/j.tig.2019.04.005. URL https://linkinghub.elsevier.com/retrieve/pii/S0168952519300769.

[116] Tao Qing, Hussein Mohsen, Michal Marczyk, Yixuan Ye, Tess O'Meara, Hongyu Zhao, Jeffrey P. Townsend, Mark Gerstein, Christos Hatzis, Yuval Kluger, and Lajos Pusztai. Germline variant burden in cancer genes correlates with age at diagnosis and somatic mutation burden. *Nature Communications*, 11(1):2438, December 2020. ISSN 2041-1723. doi: 10.1038/s41467-020-16293-7. URL http://www.nature.com/articles/s41467-020-16293-7.

[117] J. E. Armes, A. J. Egan, M. C. Southey, G. S. Dite, M. R. McCredie, G. G. Giles, J. L. Hopper, and D. J. Venter. The histologic phenotypes of breast carcinoma occurring before age 40 years in women with and without BRCA1 or BRCA2 germline mutations: a population-based study. *Cancer*, 83(11):2335–2345, December 1998. ISSN 0008-543X.

[118] S. R. Lakhani, J. Jacquemier, J. P. Sloane, B. A. Gusterson, T. J. Anderson, M. J. van de Vijver, L. M. Farid, D. Venter, A. Antoniou, A. Storfer-Isser, E. Smyth, C. M. Steel, N. Haites, R. J. Scott, D. Goldgar, S. Neuhausen, P. A. Daly, W. Ormiston, R. McManus, S. Scherneck, B. A. Ponder, D. Ford, J. Peto, D. Stoppa-Lyonnet, Y. J. Bignon, J. P. Struewing, N. K. Spurr, D. T. Bishop, J. G. Klijn, P. Devilee, C. J. Cornelisse, C. Lasset, G. Lenoir, R. B. Barkardottir, V. Egilsson, U. Hamann, J. Chang-Claude, H. Sobol, B. Weber, M. R. Stratton, and D. F. Easton. Multifactorial analysis of differences between sporadic breast cancers and cancers involving BRCA1 and BRCA2 mutations. *Journal of the National Cancer Institute*, 90(15): 1138–1145, August 1998. ISSN 0027-8874. doi: 10.1093/jnci/90.15.1138.

[119] M. C. Southey, S. J. Ramus, J. G. Dowty, L. D. Smith, A. A. Tesoriero, E. E. M. Wong, G. S. Dite, M. A. Jenkins, G. B. Byrnes, I. Winship, K.-A. Phillips, G. G. Giles, and J. L. Hopper. Morphological predictors of BRCA1 germline mutations in young women with breast cancer. *British Journal of Cancer*, 104(6):903–909, March 2011. ISSN 1532-1827. doi: 10.1038/bjc.2011.41.

[120] Núria Bonifaci, Bohdan Górski, Bartlomiej Masojć, Dominika Wokołorczyk, Anna Jakubowska, Tadeusz Debniak, Antoni Berenguer, Jordi Serra Musach, Joan Brunet, Joaquín Dopazo, Steven A. Narod, Jan Lubiński, Conxi Lázaro, Cezary Cybulski, and Miguel Angel Pujana. Exploring the Link between Germline and Somatic Genetic Alterations in Breast Carcinogenesis. *PLoS ONE*, 5(11):e14078, November 2010. ISSN 1932-6203. doi: 10.1371/journal.pone.0014078. URL https://dx.plos.org/10.1371/journal.pone.0014078.

[121] EMBRACE, GEMO Study Collaborators, HEBON, kConFab Investigators, Alison M Dunning, Kyriaki Michailidou, Karoline B Kuchenbaecker, Deborah Thomp-

son, Juliet D French, Jonathan Beesley, Catherine S Healey, Siddhartha Kar, Karen A Pooley, Elena Lopez-Knowles, Ed Dicks, Daniel Barrowdale, Nicholas A Sinnott-Armstrong, Richard C Sallari, Kristine M Hillman, Susanne Kaufmann, Haran Sivakumaran, Mahdi Moradi Marjaneh, Jason S Lee, Margaret Hills, Monika Jarosz, Suzie Drury, Sander Canisius, Manjeet K Bolla, Joe Dennis, Qin Wang, John L Hopper, Melissa C Southey, Annegien Broeks, Marjanka K Schmidt, Artitaya Lophatananon, Kenneth Muir, Matthias W Beckmann, Peter A Fasching, Isabel dos Santos-Silva, Julian Peto, Elinor J Sawyer, Ian Tomlinson, Barbara Burwinkel, Frederik Marme, Pascal Guénel, Thérèse Truong, Stig E Bojesen, Henrik Flyger, Anna González-Neira, Jose I A Perez, Hoda Anton-Culver, Lee Eunjung, Volker Arndt, Hermann Brenner, Alfons Meindl, Rita K Schmutzler, Hiltrud Brauch, Ute Hamann, Kristiina Aittomäki, Carl Blomqvist, Hidemi Ito, Keitaro Matsuo, Natasha Bogdanova, Thilo Dörk, Annika Lindblom, Sara Margolin, Veli-Matti Kosma, Arto Mannermaa, Chiu-chen Tseng, Anna H Wu, Diether Lambrechts, Hans Wildiers, Jenny Chang-Claude, Anja Rudolph, Paolo Peterlongo, Paolo Radice, Janet E Olson, Graham G Giles, Roger L Milne, Christopher A Haiman, Brian E Henderson, Mark S Goldberg, Soo H Teo, Cheng Har Yip, Silje Nord, Anne-Lise Borresen-Dale, Vessela Kristensen, Jirong Long, Wei Zheng, Katri Pylkäs, Robert Winqvist, Irene L Andrulis, Julia A Knight, Peter Devilee, Caroline Seynaeve, Jonine Figueroa, Mark E Sherman, Kamila Czene, Hatef Darabi, Antoinette Hollestelle, Ans M W van den Ouweland, Keith Humphreys, Yu-Tang Gao, Xiao-Ou Shu, Angela Cox, Simon S Cross, William Blot, Qiuyin Cai, Maya Ghoussaini, Barbara J Perkins, Mitul Shah, Ji-Yeob Choi, Daehee Kang, Soo Chin Lee, Mikael Hartman, Maria Kabisch, Diana Torres, Anna Jakubowska, Jan Lubinski, Paul Brennan, Suleeporn Sangrajrang, Christine B Ambrosone, Amanda E Toland, Chen-Yang Shen, Pei-Ei Wu, Nick Orr, Anthony Swerdlow, Lesley McGuffog, Sue Healey, Andrew Lee, Miroslav Kapuscinski, Esther M John, Mary Beth Terry, Mary B Daly, David E Goldgar, Saundra S Buys, Ramunas Janavicius, Laima Tihomirova, Nadine Tung, Cecilia M Dorfling, Elizabeth J van Rensburg, Susan L Neuhausen, Bent Ejlertsen, Thomas V O Hansen, Ana Osorio, Javier Benitez, Rachel Rando, Jeffrey N Weitzel, Bernardo Bonanni, Bernard Peissel, Siranoush Manoukian, Laura Papi, Laura Ottini, Irene Konstantopoulou, Paraskevi Apostolou, Judy Garber, Muhammad Usman Rashid, Debra Frost, Louise Izatt, Steve Ellis, Andrew K Godwin, Norbert Arnold, Dieter Niederacher, Kerstin Rhiem, Nadja Bogdanova-Markov, Charlotte Sagne, Dominique Stoppa-Lyonnet, Francesca Damiola, Olga M Sinilnikova, Sylvie Mazoyer, Claudine Isaacs, Kathleen B M Claes, Kim De Leeneer, Miguel de la Hoya, Trinidad Caldes, Heli Nevanlinna, Sofia Khan, Arjen R Mensenkamp, Maartje J Hooning, Matti A Rookus, Ava Kwong, Edith Olah, Orland Diez, Joan Brunet, Miquel Angel Pujana, Jacek Gronwald, Tomasz Huzarski, Rosa B Barkardottir, Rachel Laframboise, Penny Soucy, Marco Montagna, Simona Agata, Manuel R Teixeira, Sue Kyung Park, Noralane Lindor, Fergus J Couch, Marc Tischkowitz, Lenka Foretova, Joseph Vijai, Kenneth Offit, Christian F Singer, Christine Rappaport, Catherine M Phelan, Mark H Greene, Phuong L Mai, Gad Rennert, Evgeny N Imyanitov, Peter J

Hulick, Kelly-Anne Phillips, Marion Piedmonte, Anna Marie Mulligan, Gord Glendon, Anders Bojesen, Mads Thomassen, Maria A Caligo, Sook-Yee Yoon, Eitan Friedman, Yael Laitman, Ake Borg, Anna von Wachenfeldt, Hans Ehrencrona, Johanna Rantala, Olufunmilayo I Olopade, Patricia A Ganz, Robert L Nussbaum, Simon A Gayther, Katherine L Nathanson, Susan M Domchek, Banu K Arun, Gillian Mitchell, Beth Y Karlan, Jenny Lester, Gertraud Maskarinec, Christy Woolcott, Christopher Scott, Jennifer Stone, Carmel Apicella, Rulla Tamimi, Robert Luben, Kay-Tee Khaw, Åslaug Helland, Vilde Haakensen, Mitch Dowsett, Paul D P Pharoah, Jacques Simard, Per Hall, Montserrat García-Closas, Celine Vachon, Georgia Chenevix-Trench, Antonis C Antoniou, Douglas F Easton, and Stacey L Edwards. Breast cancer risk variants at 6q25 display different phenotype associations and regulate ESR1, RMND1 and CCDC170. *Nature Genetics*, 48(4): 374–386, April 2016. ISSN 1061-4036, 1546-1718. doi: 10.1038/ng.3521. URL `http://www.nature.com/articles/ng.3521`.

[122] K. N. Stevens, C. M. Vachon, A. M. Lee, S. Slager, T. Lesnick, C. Olswold, P. A. Fasching, P. Miron, D. Eccles, J. E. Carpenter, A. K. Godwin, C. Ambrosone, R. Winqvist, H. Brauch, on behalf of the GENICA consortium, M. K. Schmidt, A. Cox, S. S. Cross, E. Sawyer, A. Hartmann, M. W. Beckmann, R. Schulz-Wendtland, A. B. Ekici, W. J. Tapper, S. M. Gerty, L. Durcan, N. Graham, R. Hein, S. Nickels, D. Flesch-Janys, J. Heinz, H.-P. Sinn, I. Konstantopoulou, F. Fostira, D. Pectasides, A. M. Dimopoulos, G. Fountzilas, C. L. Clarke, R. Balleine, J. E. Olson, Z. Fredericksen, R. B. Diasio, H. Pathak, E. Ross, J. Weaver, T. Rudiger, A. Forsti, T. Dunnebier, F. Ademuyiwa, S. Kulkarni, K. Pylkas, A. Jukkola-Vuorinen, Y.-D. Ko, E. Van Limbergen, H. Janssen, J. Peto, O. Fletcher, G. G. Giles, L. Baglietto, S. Verhoef, I. Tomlinson, V.-M. Kosma, J. Beesley, D. Greco, C. Blomqvist, A. Irwanto, J. Liu, F. M. Blows, S.-J. Dawson, S. Margolin, A. Mannermaa, N. G. Martin, G. W. Montgomery, D. Lambrechts, I. dos Santos Silva, G. Severi, U. Hamann, P. Pharoah, D. F. Easton, J. Chang-Claude, D. Yannoukakos, H. Nevanlinna, X. Wang, and F. J. Couch. Common Breast Cancer Susceptibility Loci Are Associated with Triple-Negative Breast Cancer. *Cancer Research*, 71(19):6240–6249, October 2011. ISSN 0008-5472, 1538-7445. doi: 10.1158/0008-5472.CAN-11-1266. URL `http://cancerres.aacrjournals.org/cgi/doi/10.1158/0008-5472.CAN-11-1266`.

[123] Noralane M. Lindor, Melissa C. Larson, Melissa S. DeRycke, Shannon K. McDonnell, Saurabh Baheti, Zachary C. Fogarty, Aung Ko Win, John D. Potter, Daniel D. Buchanan, Mark Clendenning, Polly A. Newcomb, Graham Casey, Steven Gallinger, Loïc Le Marchand, John L. Hopper, Mark A. Jenkins, Ellen L. Goode, and Stephen N. Thibodeau. Germline miRNA DNA variants and the risk of colorectal cancer by subtype. *Genes, Chromosomes & Cancer*, 56(3):177–184, 2017. ISSN 1098-2264. doi: 10.1002/gcc.22420.

[124] Joice Kuroiwa-Trzmielina, Fan Wang, Robert W. Rapkins, Robyn L. Ward, Daniel D. Buchanan, Aung Ko Win, Mark Clendenning, Christophe Rosty, Melissa C. Southey, Ingrid M. Winship, John L. Hopper, Mark A. Jenkins, Jake

Olivier, Nicholas J. Hawkins, and Megan P. Hitchins. SNP rs16906252C>T Is an Expression and Methylation Quantitative Trait Locus Associated with an Increased Risk of Developing MGMT-Methylated Colorectal Cancer. *Clinical Cancer Research: An Official Journal of the American Association for Cancer Research*, 22 (24):6266–6277, December 2016. ISSN 1078-0432. doi: 10.1158/1078-0432.CCR-15 -2765.

[125] C. Richard Boland and Ajay Goel. Microsatellite instability in colorectal cancer. *Gastroenterology*, 138(6):2073–2087.e3, June 2010. ISSN 1528-0012. doi: 10.1053/j. gastro.2009.12.064.

[126] Serena Nik-Zainal, Helen Davies, Johan Staaf, Manasa Ramakrishna, Dominik Glodzik, Xueqing Zou, Inigo Martincorena, Ludmil B. Alexandrov, Sancha Martin, David C. Wedge, Peter Van Loo, Young Seok Ju, Marcel Smid, Arie B. Brinkman, Sandro Morganella, Miriam R. Aure, Ole Christian Lingjærde, Anita Langerød, Markus Ringnér, Sung-Min Ahn, Sandrine Boyault, Jane E. Brock, Annegien Broeks, Adam Butler, Christine Desmedt, Luc Dirix, Serge Dronov, Aquila Fatima, John A. Foekens, Moritz Gerstung, Gerrit K. J. Hooijer, Se Jin Jang, David R. Jones, Hyung-Yong Kim, Tari A. King, Savitri Krishnamurthy, Hee Jin Lee, Jeong-Yeon Lee, Yilong Li, Stuart McLaren, Andrew Menzies, Ville Mustonen, Sarah O'Meara, Iris Pauporté, Xavier Pivot, Colin A. Purdie, Keiran Raine, Kamna Ramakrishnan, F. Germán Rodríguez-González, Gilles Romieu, Anieta M. Sieuwerts, Peter T. Simpson, Rebecca Shepherd, Lucy Stebbings, Olafur A. Stefansson, Jon Teague, Stefania Tommasi, Isabelle Treilleux, Gert G. Van den Eynden, Peter Vermeulen, Anne Vincent-Salomon, Lucy Yates, Carlos Caldas, Laura van't Veer, Andrew Tutt, Stian Knappskog, Benita Kiat Tee Tan, Jos Jonkers, Åke Borg, Naoto T. Ueno, Christos Sotiriou, Alain Viari, P. Andrew Futreal, Peter J. Campbell, Paul N. Span, Steven Van Laere, Sunil R. Lakhani, Jorunn E. Eyfjord, Alastair M. Thompson, Ewan Birney, Hendrik G. Stunnenberg, Marc J. van de Vijver, John W. M. Martens, Anne-Lise Børresen-Dale, Andrea L. Richardson, Gu Kong, Gilles Thomas, and Michael R. Stratton. Landscape of somatic mutations in 560 breast cancer whole-genome sequences. *Nature*, 534(7605):47–54, June 2016. ISSN 0028-0836, 1476-4687. doi: 10.1038/nature17676. URL `http://www.nature.com/articles/nature17676`.

[127] Francesca Menghi, Koichiro Inaki, XingYi Woo, Pooja A. Kumar, Krzysztof R. Grzeda, Ankit Malhotra, Vinod Yadav, Hyunsoo Kim, Eladio J. Marquez, Duygu Ucar, Phung T. Shreckengast, Joel P. Wagner, George MacIntyre, Krishna R. Murthy Karuturi, Ralph Scully, James Keck, Jeffrey H. Chuang, and Edison T. Liu. The tandem duplicator phenotype as a distinct genomic configuration in cancer. *Proceedings of the National Academy of Sciences of the United States of America*, 113(17):E2373–2382, April 2016. ISSN 1091-6490. doi: 10.1073/pnas.1520010113.

[128] Carla Daniela Robles-Espinoza, Nicola D. Roberts, Shuyang Chen, Finbarr P. Leacy, Ludmil B. Alexandrov, Natapol Pornputtapong, Ruth Halaban, Michael Krauthammer, Rutao Cui, D. Timothy Bishop, and David J. Adams. Germline MC1R status influences somatic mutation burden in melanoma. *Nature Communications*, 7:12064, 2016. ISSN 2041-1723. doi: 10.1038/ncomms12064.

[129] Junichi Soh, Naoki Okumura, William W. Lockwood, Hiromasa Yamamoto, Hisayuki Shigematsu, Wei Zhang, Raj Chari, David S. Shames, Ximing Tang, Calum MacAulay, Marileila Varella-Garcia, Tõnu Vooder, Ignacio I. Wistuba, Stephen Lam, Rolf Brekken, Shinichi Toyooka, John D. Minna, Wan L. Lam, and Adi F. Gazdar. Oncogene mutations, copy number gains and mutant allele specific imbalance (MASI) frequently occur together in tumor cells. *PloS One*, 4(10):e7464, October 2009. ISSN 1932-6203. doi: 10.1371/journal.pone.0007464.

[130] Chih-Chieh Yu, Wanglong Qiu, Caroline S. Juang, Mahesh M. Mansukhani, Balazs Halmos, and Gloria H. Su. Mutant allele specific imbalance in oncogenes with copy number alterations: Occurrence, mechanisms, and potential clinical implications. *Cancer Letters*, 384:86–93, January 2017. ISSN 03043835. doi: 10.1016/j.canlet.2016.10.013. URL https://linkinghub.elsevier.com/retrieve/pii/S0304383516306255.

[131] YongKiat Wee, TianFang Wang, Yining Liu, Xiaoyan Li, and Min Zhao. A pan-cancer study of copy number gain and up-regulation in human oncogenes. *Life Sciences*, 211:206–214, October 2018. ISSN 1879-0631. doi: 10.1016/j.lfs.2018.09.032.

[132] Damla Olcaydu, Ashot Harutyunyan, Roland Jäger, Tiina Berg, Bettina Gisslinger, Ingrid Pabinger, Heinz Gisslinger, and Robert Kralovics. A common JAK2 haplotype confers susceptibility to myeloproliferative neoplasms. *Nature Genetics*, 41 (4):450–454, April 2009. ISSN 1061-4036, 1546-1718. doi: 10.1038/ng.341. URL http://www.nature.com/articles/ng.341.

[133] Wanqing Liu, Lijun He, Jacqueline Ramírez, Soundararajan Krishnaswamy, Rajani Kanteti, Yi-Ching Wang, Ravi Salgia, and Mark J. Ratain. Functional EGFR germline polymorphisms may confer risk for EGFR somatic mutations in non-small cell lung cancer, with a predominant effect on exon 19 microdeletions. *Cancer Research*, 71(7):2423–2427, April 2011. ISSN 1538-7445. doi: 10.1158/0008-5472.CAN-10-2689.

[134] Xiang Shu, Jianchun Gu, Maosheng Huang, Nizar M Tannir, Surena F Matin, Jose A Karam, Christopher G Wood, Xifeng Wu, and Yuanqing Ye. Germline genetic variants in somatically significantly mutated genes in tumors are associated with renal cell carcinoma risk and outcome. *Carcinogenesis*, 39(6):752–757, May 2018. ISSN 0143-3334, 1460-2180. doi: 10.1093/carcin/bgy021. URL https://academic.oup.com/carcin/article/39/6/752/4960094.

[135] Roberto Puzone and Ulrich Pfeffer. SNP variants at the MAP3K1/SETD9 locus 5q11.2 associate with somatic PIK3CA variants in breast cancers. *European Journal of Human Genetics*, 25(3):384–387, March 2017. ISSN 1018-4813, 1476-5438. doi: 10.1038/ejhg.2016.179. URL http://www.nature.com/articles/ejhg2016179.

[136] Xu Zhang, Yuzhuo Wang, Tian Tian, Gangqiao Zhou, and Guangfu Jin. Germline genetic variants were interactively associated with somatic alterations in gastric cancer. *Cancer Medicine*, 7(8):3912–3920, August 2018. ISSN 20457634. doi: 10.1002/cam4.1612. URL http://doi.wiley.com/10.1002/cam4.1612.

[137] Yuzhuo Wang, Cheng Wang, Jiahui Zhang, Meng Zhu, Xu Zhang, Zhihua Li, Juncheng Dai, Hongxia Ma, Zhibin Hu, Guangfu Jin, and Hongbing Shen. Interaction analysis between germline susceptibility loci and somatic alterations in lung cancer: Germline variants and somatic alterations. *International Journal of Cancer*, 143(4):878–885, August 2018. ISSN 00207136. doi: 10.1002/ijc.31351. URL http://doi.wiley.com/10.1002/ijc.31351.

[138] Nancy E. Thomas, Sharon N. Edmiston, Irene Orlow, Peter A. Kanetsky, Li Luo, David C. Gibbs, Eloise A. Parrish, Honglin Hao, Klaus J. Busam, Bruce K. Armstrong, Anne Kricker, Anne E. Cust, Hoda Anton-Culver, Stephen B. Gruber, Richard P. Gallagher, Roberto Zanetti, Stefano Rosso, Lidia Sacchetto, Terence Dwyer, David W. Ollila, Colin B. Begg, Marianne Berwick, Kathleen Conway, Marianne Berwick, Colin Begg, Irene Orlow, Klaus J. Busam, Pampa Roy, Anne Reiner, Siok Leong, Sergio Corrales Guerrero, Keimya Sadeghi, Marianne Berwick, Li Luo, Tawny W. Boyce, Anne E. Cust, Bruce K. Armstrong, Anne Kricker, Alison Venn, Terence Dwyer, Paul Tucker, Richard P. Gallagher, Loraine D. Marrett, Lynn From, Roberto Zanetti, Stefano Rosso, Hoda Anton-Culver, Stephen B. Gruber, Shu-Chen Huang, Nancy E. Thomas, Kathleen Conway, David W. Ollila, Pamela A. Groben, Sharon N. Edmiston, Honglin Hao, Eloise Parrish, Jill S. Frank, David C. Gibbs, Timothy R. Rebbeck, Peter A. Kanetsky, Julia Lee Taylor, and Sasha Madronich. Inherited Genetic Variants Associated with Melanoma BRAF/NRAS Subtypes. *Journal of Investigative Dermatology*, 138(11):2398–2404, November 2018. ISSN 0022202X. doi: 10.1016/j.jid.2018.04.025. URL https://linkinghub.elsevier.com/retrieve/pii/S0022202X18319626.

[139] Hannah Carter, Rachel Marty, Matan Hofree, Andrew M. Gross, James Jensen, Kathleen M. Fisch, Xingyu Wu, Christopher DeBoever, Eric L. Van Nostrand, Yan Song, Emily Wheeler, Jason F. Kreisberg, Scott M. Lippman, Gene W. Yeo, J. Silvio Gutkind, and Trey Ideker. Interaction Landscape of Inherited Polymorphisms with Somatic Events in Cancer. *Cancer Discovery*, 7(4):410–423, April 2017. ISSN 2159-8274, 2159-8290. doi: 10.1158/2159-8290.CD-16-1045. URL http://cancerdiscovery.aacrjournals.org/lookup/doi/10.1158/2159-8290.CD-16-1045.

[140] Nancy E. Thomas, Sharon N. Edmiston, Peter A. Kanetsky, Klaus J. Busam, Anne Kricker, Bruce K. Armstrong, Anne E. Cust, Hoda Anton-Culver, Stephen B. Gruber, Li Luo, Irene Orlow, Anne S. Reiner, Richard P. Gallagher, Roberto Zanetti, Stefano Rosso, Lidia Sacchetto, Terence Dwyer, Eloise A. Parrish, Honglin Hao, David C. Gibbs, Jill S. Frank, David W. Ollila, Colin B. Begg, Marianne Berwick, Kathleen Conway, Marianne Berwick, Colin B. Begg, Irene Orlow, Klaus J. Busam, Anne S. Reiner, Pampa Roy, Himali Patel, Marianne Berwick, Li Luo, Susan Paine, Anne E. Cust, Bruce K. Armstrong, Anne Kricker, Alison Venn, Terence Dwyer, Paul Tucker, Richard P. Gallagher, Loraine D. Marrett, Lynn From, Roberto Zanetti, Stefano Rosso, Hoda Anton-Culver, Stephen B. Gruber, Shu-Chen Huang, Nancy E. Thomas, David W. Ollila, Kathleen Conway, Pamela A. Groben, Sharon N. Edmiston, Honglin Hao, Eloise Parrish, Jill S. Frank, David C. Gibbs, Jennifer I. Bramson, Timothy R. Rebbeck, Peter A. Kanetsky, Julia Lee Taylor, and Sasha

Madronich. Associations of MC1R Genotype and Patient Phenotypes with BRAF and NRAS Mutations in Melanoma. *Journal of Investigative Dermatology*, 137(12): 2588–2598, December 2017. ISSN 0022202X. doi: 10.1016/j.jid.2017.07.832. URL `https://linkinghub.elsevier.com/retrieve/pii/S0022202X17327914`.

[141] Andrea Borrego, Wafa H. K. Cabrera, José R. Jensen, Mara Correa, Orlando G. Ribeiro, Nancy Starobinas, Marcelo De Franco, Angela Pettinicchio, Tommaso A. Dragani, Olga C. M. Ibañez, and Giacomo Manenti. Germline control of somatic Kras mutations in mouse lung tumors. *Molecular Carcinogenesis*, 57(6):745–751, 2018. ISSN 1098-2744. doi: 10.1002/mc.22796.

[142] F. C. Grumet, A. Coukell, J. G. Bodmer, W. F. Bodmer, and H. O. McDevitt. Histocompatibility (HL-A) antigens associated with systemic lupus erythematosus. A possible genetic predisposition to disease. *The New England Journal of Medicine*, 285(4):193–196, July 1971. ISSN 0028-4793. doi: 10.1056/NEJM197107222850403.

[143] Elizabeth Trachtenberg, Bette Korber, Cristina Sollars, Thomas B Kepler, Peter T Hraber, Elizabeth Hayes, Robert Funkhouser, Michael Fugate, James Theiler, Yen S Hsu, Kevin Kunstman, Samuel Wu, John Phair, Henry Erlich, and Steven Wolinsky. Advantage of rare HLA supertype in HIV disease progression. *Nature Medicine*, 9 (7):928–935, July 2003. ISSN 1078-8956, 1546-170X. doi: 10.1038/nm893. URL `http://www.nature.com/articles/nm893`.

[144] Philippe Goyette, Gabrielle Boucher, Dermot Mallon, Eva Ellinghaus, Luke Jostins, Hailiang Huang, Stephan Ripke, Elena S. Gusareva, Vito Annese, Stephen L. Hauser, Jorge R. Oksenberg, Ingo Thomsen, Stephen Leslie, International Inflammatory Bowel Disease Genetics Consortium, Australia and New Zealand IBDGC, Belgium IBD Genetics Consortium, Italian Group for IBD Genetic Consortium, NIDDK Inflammatory Bowel Disease Genetics Consortium, United Kingdom IBDGC, Wellcome Trust Case Control Consortium, Quebec IBD Genetics Consortium, Mark J. Daly, Kristel Van Steen, Richard H. Duerr, Jeffrey C. Barrett, Dermot P. B. McGovern, L. Philip Schumm, James A. Traherne, Mary N. Carrington, Vasilis Kosmoliaptsis, Tom H. Karlsen, Andre Franke, and John D. Rioux. High-density mapping of the MHC identifies a shared role for HLA-DRB1*01:03 in inflammatory bowel diseases and heterozygous advantage in ulcerative colitis. *Nature Genetics*, 47(2): 172–179, February 2015. ISSN 1546-1718. doi: 10.1038/ng.3176.

[145] Javier Gutierrez-Achury, Alexandra Zhernakova, Sara L. Pulit, Gosia Trynka, Karen A. Hunt, Jihane Romanos, Soumya Raychaudhuri, David A. van Heel, Cisca Wijmenga, and Paul I. W. de Bakker. Fine mapping in the MHC region accounts for 18% additional genetic risk for celiac disease. *Nature Genetics*, 47(6):577–578, June 2015. ISSN 1546-1718. doi: 10.1038/ng.3268.

[146] Xinli Hu, Aaron J. Deutsch, Tobias L. Lenz, Suna Onengut-Gumuscu, Buhm Han, Wei-Min Chen, Joanna M. M. Howson, John A. Todd, Paul I. W. de Bakker, Stephen S. Rich, and Soumya Raychaudhuri. Additive and interaction effects at three amino acid positions in HLA-DQ and HLA-DR molecules drive type 1 diabetes risk. *Nature Genetics*, 47(8):898–905, August 2015. ISSN 1546-1718. doi: 10.1038/ng.3353.

[147] Loukas Moutsianas, Luke Jostins, Ashley H. Beecham, Alexander T. Dilthey, Dionysia K. Xifara, Maria Ban, Tejas S. Shah, Nikolaos A. Patsopoulos, Lars Alfredsson, Carl A. Anderson, Katherine E. Attfield, Sergio E. Baranzini, Jeffrey Barrett, Thomas M. C. Binder, David Booth, Dorothea Buck, Elisabeth G. Celius, Chris Cotsapas, Sandra D'Alfonso, Calliope A. Dendrou, Peter Donnelly, Bénédicte Dubois, Bertrand Fontaine, Lars Fugger, An Goris, Pierre-Antoine Gourraud, Christiane Graetz, Bernhard Hemmer, Jan Hillert, International IBD Genetics Consortium (IIBDGC), Ingrid Kockum, Stephen Leslie, Christina M. Lill, Filippo Martinelli-Boneschi, Jorge R. Oksenberg, Tomas Olsson, Annette Oturai, Janna Saarela, Helle Bach Søndergaard, Anne Spurkland, Bruce Taylor, Juliane Winkelmann, Frauke Zipp, Jonathan L. Haines, Margaret A. Pericak-Vance, Chris C. A. Spencer, Graeme Stewart, David A. Hafler, Adrian J. Ivinson, Hanne F. Harbo, Stephen L. Hauser, Philip L. De Jager, Alastair Compston, Jacob L. McCauley, Stephen Sawcer, and Gil McVean. Class II HLA interactions modulate genetic risk for multiple sclerosis. *Nature Genetics*, 47(10):1107–1113, October 2015. ISSN 1546-1718. doi: 10.1038/ng.3395.

[148] Celi Sun, Julio E. Molineros, Loren L. Looger, Xu-Jie Zhou, Kwangwoo Kim, Yukinori Okada, Jianyang Ma, Yuan-Yuan Qi, Xana Kim-Howard, Prasenjeet Motghare, Krishna Bhattarai, Adam Adler, So-Young Bang, Hye-Soon Lee, Tae-Hwan Kim, Young Mo Kang, Chang-Hee Suh, Won Tae Chung, Yong-Beom Park, Jung-Yoon Choe, Seung Cheol Shim, Yuta Kochi, Akari Suzuki, Michiaki Kubo, Takayuki Sumida, Kazuhiko Yamamoto, Shin-Seok Lee, Young Jin Kim, Bok-Ghee Han, Mikhail Dozmorov, Kenneth M. Kaufman, Jonathan D. Wren, John B. Harley, Nan Shen, Kek Heng Chua, Hong Zhang, Sang-Cheol Bae, and Swapan K. Nath. High-density genotyping of immune-related loci identifies new SLE risk variants in individuals with Asian ancestry. *Nature Genetics*, 48(3):323–330, March 2016. ISSN 1546-1718. doi: 10.1038/ng.3496.

[149] Chao Tian, Bethann S. Hromatka, Amy K. Kiefer, Nicholas Eriksson, Suzanne M. Noble, Joyce Y. Tung, and David A. Hinds. Genome-wide association and HLA region fine-mapping studies identify susceptibility loci for multiple common infections. *Nature Communications*, 8(1):599, December 2017. ISSN 2041-1723. doi: 10.1038/s41467-017-00257-5. URL `http://www.nature.com/articles/s41467-017-00257-5`.

[150] Matteo D'Antonio, Joaquin Reyna, David Jakubosky, Margaret KR Donovan, Marc-Jan Bonder, Hiroko Matsui, Oliver Stegle, Naoki Nariai, Agnieszka D'Antonio-Chronowska, and Kelly A Frazer. Systematic genetic analysis of the MHC region reveals mechanistic underpinnings of HLA type associations with disease. *eLife*, 8:e48476, November 2019. ISSN 2050-084X. doi: 10.7554/eLife.48476. URL `https://elifesciences.org/articles/48476`.

[151] Rachel Marty, Saghar Kaabinejadian, David Rossell, Michael J. Slifker, Joris van de Haar, Hatice Billur Engin, Nicola de Prisco, Trey Ideker, William H. Hildebrand, Joan Font-Burgada, and Hannah Carter. MHC-I Genotype Restricts the Oncogenic Mutational Landscape. *Cell*, 171(6):1272–1283.e15, November 2017. ISSN 00928674.

doi: 10.1016/j.cell.2017.09.050. URL https://linkinghub.elsevier.com/retrieve/pii/S0092867417311443.

[152] Rachel Marty Pyke, Wesley Kurt Thompson, Rany M. Salem, Joan Font-Burgada, Maurizio Zanetti, and Hannah Carter. Evolutionary Pressure against MHC Class II Binding Cancer Mutations. *Cell*, 175(2):416–428.e13, October 2018. ISSN 00928674. doi: 10.1016/j.cell.2018.08.048. URL https://linkinghub.elsevier.com/retrieve/pii/S0092867418311097.

[153] Yoong Wearn Lim, Haiyin Chen-Harris, Oleg Mayba, Steve Lianoglou, Arthur Wuster, Tushar Bhangale, Zia Khan, Sanjeev Mariathasan, Anneleen Daemen, Jens Reeder, Peter M. Haverty, William F. Forrest, Matthew Brauer, Ira Mellman, and Matthew L. Albert. Germline genetic polymorphisms influence tumor gene expression and immune cell infiltration. *Proceedings of the National Academy of Sciences*, 115(50):E11701–E11710, December 2018. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.1804506115. URL http://www.pnas.org/lookup/doi/10.1073/pnas.1804506115.

[154] Sahar Shahamatdar, Meng Xiao He, Matthew A. Reyna, Alexander Gusev, Saud H. AlDubayan, Eliezer M. Van Allen, and Sohini Ramachandran. Germline Features Associated with Immune Infiltration in Solid Tumors. *Cell Reports*, 30(9):2900–2908.e4, March 2020. ISSN 22111247. doi: 10.1016/j.celrep.2020.02.039. URL https://linkinghub.elsevier.com/retrieve/pii/S221112472030200X.

[155] Rosalyn W. Sayaman, Mohamad Saad, Vésteinn Thorsson, Donglei Hu, Wouter Hendrickx, Jessica Roelands, Eduard Porta-Pardo, Younes Mokrab, Farshad Farshidfar, Tomas Kirchhoff, Randy F. Sweis, Oliver F. Bathe, Carolina Heimann, Michael J. Campbell, Cynthia Stretch, Scott Huntsman, Rebecca E. Graff, Najeeb Syed, Laszlo Radvanyi, Simon Shelley, Denise Wolf, Francesco M. Marincola, Michele Ceccarelli, Jérôme Galon, Elad Ziv, and Davide Bedognetti. Germline genetic contribution to the immune landscape of cancer. *Immunity*, 54(2):367–386.e8, February 2021. ISSN 10747613. doi: 10.1016/j.immuni.2021.01.011. URL https://linkinghub.elsevier.com/retrieve/pii/S1074761321000340.

[156] William J. Astle, Heather Elding, Tao Jiang, Dave Allen, Dace Ruklisa, Alice L. Mann, Daniel Mead, Heleen Bouman, Fernando Riveros-Mckay, Myrto A. Kostadima, John J. Lambourne, Suthesh Sivapalaratnam, Kate Downes, Kousik Kundu, Lorenzo Bomba, Kim Berentsen, John R. Bradley, Louise C. Daugherty, Olivier Delaneau, Kathleen Freson, Stephen F. Garner, Luigi Grassi, Jose Guerrero, Matthias Haimel, Eva M. Janssen-Megens, Anita Kaan, Mihir Kamat, Bowon Kim, Amit Mandoli, Jonathan Marchini, Joost H.A. Martens, Stuart Meacham, Karyn Megy, Jared O'Connell, Romina Petersen, Nilofar Sharifi, Simon M. Sheard, James R. Staley, Salih Tuna, Martijn van der Ent, Klaudia Walter, Shuang-Yin Wang, Eleanor Wheeler, Steven P. Wilder, Valentina Iotchkova, Carmel Moore, Jennifer Sambrook, Hendrik G. Stunnenberg, Emanuele Di Angelantonio, Stephen Kaptoge, Taco W. Kuijpers, Enrique Carrillo-de Santa-Pau, David Juan, Daniel Rico, Alfonso Valencia, Lu Chen, Bing Ge, Louella Vasquez, Tony Kwan, Diego Garrido-Martín, Stephen Watt, Ying Yang, Roderic Guigo, Stephan Beck, Dirk S. Paul,

Tomi Pastinen, David Bujold, Guillaume Bourque, Mattia Frontini, John Danesh, David J. Roberts, Willem H. Ouwehand, Adam S. Butterworth, and Nicole Soranzo. The Allelic Landscape of Human Blood Cell Trait Variation and Links to Common Complex Disease. *Cell*, 167(5):1415–1429.e19, November 2016. ISSN 00928674. doi: 10.1016/j.cell.2016.10.042. URL https://linkinghub.elsevier.com/retrieve/pii/S0092867416314635.

[157] Andrew R. Marderstein, Manik Uppal, Akanksha Verma, Bhavneet Bhinder, Zakieh Tayyebi, Jason Mezey, Andrew G. Clark, and Olivier Elemento. Demographic and genetic factors influence the abundance of infiltrating immune cells in human tissues. *Nature Communications*, 11(1):2213, December 2020. ISSN 2041-1723. doi: 10.1038/s41467-020-16097-9. URL http://www.nature.com/articles/s41467-020-16097-9.

[158] Daniel Kogan, Alexander Grabner, Christopher Yanucil, Christian Faul, and Vijay Kumar Ulaganathan. STAT3-enhancing germline mutations contribute to tumor-extrinsic immune evasion. *Journal of Clinical Investigation*, 128(5):1867–1872, May 2018. ISSN 0021-9738, 1558-8238. doi: 10.1172/JCI96708. URL https://www.jci.org/articles/view/96708.

[159] Yi Zhang, Mohith Manjunath, Jialu Yan, Brittany A. Baur, Shilu Zhang, Sushmita Roy, and Jun S. Song. The Cancer-Associated Genetic Variant Rs3903072 Modulates Immune Cells in the Tumor Microenvironment. *Frontiers in Genetics*, 10:754, August 2019. ISSN 1664-8021. doi: 10.3389/fgene.2019.00754. URL https://www.frontiersin.org/article/10.3389/fgene.2019.00754/full.

[160] Joshua M Korn, Finny G Kuruvilla, Steven A McCarroll, Alec Wysoker, James Nemesh, Simon Cawley, Earl Hubbell, Jim Veitch, Patrick J Collins, Katayoon Darvishi, Charles Lee, Marcia M Nizzari, Stacey B Gabriel, Shaun Purcell, Mark J Daly, and David Altshuler. Integrated genotype calling and association analysis of SNPs, common copy number polymorphisms and rare CNVs. *Nature Genetics*, 40 (10):1253–1260, October 2008. ISSN 1061-4036, 1546-1718. doi: 10.1038/ng.237. URL http://www.nature.com/articles/ng.237.

[161] Sayantan Das, Lukas Forer, Sebastian Schönherr, Carlo Sidore, Adam E Locke, Alan Kwong, Scott I Vrieze, Emily Y Chew, Shawn Levy, Matt McGue, David Schlessinger, Dwight Stambolian, Po-Ru Loh, William G Iacono, Anand Swaroop, Laura J Scott, Francesco Cucca, Florian Kronenberg, Michael Boehnke, Gonçalo R Abecasis, and Christian Fuchsberger. Next-generation genotype imputation service and methods. *Nature Genetics*, 48(10):1284–1287, October 2016. ISSN 1061-4036, 1546-1718. doi: 10.1038/ng.3656. URL http://www.nature.com/articles/ng.3656.

[162] Matthew T Chang, Saurabh Asthana, Sizhi Paul Gao, Byron H Lee, Jocelyn S Chapman, Cyriac Kandoth, JianJiong Gao, Nicholas D Socci, David B Solit, Adam B Olshen, Nikolaus Schultz, and Barry S Taylor. Identifying recurrent mutations in cancer reveals widespread lineage diversity and mutational specificity. *Nature Biotechnology*, 34(2):155–163, February 2016. ISSN 1087-0156, 1546-1696. doi: 10.1038/nbt.3391. URL http://www.nature.com/articles/nbt.3391.

[163] Liacine Bouaoun, Dmitriy Sonkin, Maude Ardin, Monica Hollstein, Graham Byrnes, Jiri Zavadil, and Magali Olivier. *TP53* Variations in Human Cancers: New Lessons from the IARC TP53 Database and Genomics Data: Human Mutation. *Human Mutation*, 37(9):865–876, September 2016. ISSN 10597794. doi: 10.1002/humu.23035. URL `http://doi.wiley.com/10.1002/humu.23035`.

[164] Rachel S Kerr, Sharon Love, Eva Segelov, Elaine Johnstone, Beverly Falcon, Peter Hewett, Andrew Weaver, David Church, Claire Scudder, Sarah Pearson, Patrick Julier, Francesco Pezzella, Ian Tomlinson, Enric Domingo, and David J Kerr. Adjuvant capecitabine plus bevacizumab versus capecitabine alone in patients with colorectal cancer (QUASAR 2): an open-label, randomised phase 3 trial. *The Lancet Oncology*, 17(11):1543–1557, November 2016. ISSN 14702045. doi: 10.1016/S1470-2045(16)30172-3. URL `https://linkinghub.elsevier.com/retrieve/pii/S1470204516301723`.

[165] Rachel S. Midgley, Christopher C. McConkey, Elaine C. Johnstone, Janet A. Dunn, Justine L. Smith, Simon A. Grumett, Patrick Julier, Claire Iveson, Yoko Yanagisawa, Bryan Warren, Michael J. Langman, and David J. Kerr. Phase III randomized trial assessing rofecoxib in the adjuvant setting of colorectal cancer: final results of the VICTOR trial. *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology*, 28(30):4575–4580, October 2010. ISSN 1527-7755. doi: 10.1200/JCO.2010.29.6244.

[166] Richard A. Adams, Angela M. Meade, Matthew T. Seymour, Richard H. Wilson, Ayman Madi, David Fisher, Sarah L. Kenny, Edward Kay, Elizabeth Hodgkinson, Malcolm Pope, Penny Rogers, Harpreet Wasan, Stephen Falk, Simon Gollins, Tamas Hickish, Eric M. Bessell, David Propper, M. John Kennedy, Richard Kaplan, Timothy S. Maughan, and MRC COIN Trial Investigators. Intermittent versus continuous oxaliplatin and fluoropyrimidine combination chemotherapy for first-line treatment of advanced colorectal cancer: results of the randomised phase 3 MRC COIN trial. *The Lancet. Oncology*, 12(7):642–653, July 2011. ISSN 1474-5488. doi: 10.1016/S1470-2045(11)70102-4.

[167] Timothy S. Maughan, Richard A. Adams, Christopher G. Smith, Angela M. Meade, Matthew T. Seymour, Richard H. Wilson, Shelley Idziaszczyk, Rebecca Harris, David Fisher, Sarah L. Kenny, Edward Kay, Jenna K. Mitchell, Ayman Madi, Bharat Jasani, Michelle D. James, John Bridgewater, M. John Kennedy, Bart Claes, Diether Lambrechts, Richard Kaplan, Jeremy P. Cheadle, and MRC COIN Trial Investigators. Addition of cetuximab to oxaliplatin-based first-line combination chemotherapy for treatment of advanced colorectal cancer: results of the randomised phase 3 MRC COIN trial. *Lancet (London, England)*, 377(9783):2103–2114, June 2011. ISSN 1474-547X. doi: 10.1016/S0140-6736(11)60613-2.

[168] Alkes L Price, Nick J Patterson, Robert M Plenge, Michael E Weinblatt, Nancy A Shadick, and David Reich. Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genetics*, 38(8):904–909, August 2006. ISSN 1061-4036, 1546-1718. doi: 10.1038/ng1847. URL `http://www.nature.com/articles/ng1847`.

[169] Christopher C Chang, Carson C Chow, Laurent CAM Tellier, Shashaank Vattikuti, Shaun M Purcell, and James J Lee. Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience*, 4(1), December 2015. ISSN 2047-217X. doi: 10.1186/s13742-015-0047-8. URL https://academic.oup.com/gigascience/article-lookup/doi/10.1186/s13742-015-0047-8.

[170] Olivier Delaneau, Jonathan Marchini, and Jean-François Zagury. A linear complexity phasing method for thousands of genomes. *Nature Methods*, 9(2):179–181, February 2012. ISSN 1548-7091, 1548-7105. doi: 10.1038/nmeth.1785. URL http://www.nature.com/articles/nmeth.1785.

[171] Bryan N. Howie, Peter Donnelly, and Jonathan Marchini. A Flexible and Accurate Genotype Imputation Method for the Next Generation of Genome-Wide Association Studies. *PLoS Genetics*, 5(6):e1000529, June 2009. ISSN 1553-7404. doi: 10.1371/journal.pgen.1000529. URL http://dx.plos.org/10.1371/journal.pgen.1000529.

[172] The UK10K Consortium. The UK10K project identifies rare variants in health and disease. *Nature*, 526(7571):82–90, October 2015. ISSN 0028-0836, 1476-4687. doi: 10.1038/nature14962. URL http://www.nature.com/articles/nature14962.

[173] Mitchell J. Machiela and Stephen J. Chanock. LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. *Bioinformatics (Oxford, England)*, 31(21):3555–3557, November 2015. ISSN 1367-4811. doi: 10.1093/bioinformatics/btv402.

[174] Enric Domingo, Carme Camps, Pamela J. Kaisaki, Marie J. Parsons, Dmitri Mouradov, Melissa M. Pentony, Seiko Makino, Michelle Palmieri, Robyn L. Ward, Nicholas J. Hawkins, Peter Gibbs, Hanne Askautrud, Dahmane Oukrif, Haitao Wang, Joe Wood, Evie Tomlinson, Yasmine Bark, Kulvinder Kaur, Elaine C. Johnstone, Claire Palles, David N. Church, Marco Novelli, Havard E. Danielsen, Jon Sherlock, David Kerr, Rachel Kerr, Oliver Sieber, Jenny C. Taylor, and Ian Tomlinson. Mutation burden and other molecular markers of prognosis in colorectal cancer treated with curative intent: results from the QUASAR 2 clinical trial and an Australian community-based series. *The Lancet. Gastroenterology & Hepatology*, 3(9):635–643, 2018. ISSN 2468-1253. doi: 10.1016/S2468-1253(18)30117-1.

[175] Nada A. Al-Tassan, Nicola Whiffin, Fay J. Hosking, Claire Palles, Susan M. Farrington, Sara E. Dobbins, Rebecca Harris, Maggie Gorman, Albert Tenesa, Brian F. Meyer, Salma M. Wakil, Ben Kinnersley, Harry Campbell, Lynn Martin, Christopher G. Smith, Shelley Idziaszczyk, Ella Barclay, Timothy S. Maughan, Richard Kaplan, Rachel Kerr, David Kerr, Daniel D. Buchanan, Aung Ko Win, John Hopper, Mark Jenkins, Noralane M. Lindor, Polly A. Newcomb, Steve Gallinger, David Conti, Fred Schumacher, Graham Casey, Malcolm G. Dunlop, Ian P. Tomlinson, Jeremy P. Cheadle, and Richard S. Houlston. A new GWAS and meta-analysis with 1000Genomes imputation identifies novel risk variants for colorectal cancer. *Scientific Reports*, 5(1):10442, September 2015. ISSN 2045-2322. doi: 10.1038/srep10442. URL http://www.nature.com/articles/srep10442.

[176] Timothy J Iveson, Rachel S Kerr, Mark P Saunders, Jim Cassidy, Niels Henrik Hollander, Josep Tabernero, Andrew Haydon, Bengt Glimelius, Andrea Harkin,

Karen Allan, John McQueen, Claire Scudder, Kathleen Anne Boyd, Andrew Briggs, Ashita Waterston, Louise Medley, Charles Wilson, Richard Ellis, Sharadah Essapen, Amandeep S Dhadda, Mark Harrison, Stephen Falk, Sherif Raouf, Charlotte Rees, Rene K Olesen, David Propper, John Bridgewater, Ashraf Azzabi, David Farrugia, Andrew Webb, David Cunningham, Tamas Hickish, Andrew Weaver, Simon Gollins, Harpreet S Wasan, and James Paul. 3 versus 6 months of adjuvant oxaliplatin-fluoropyrimidine combination therapy for colorectal cancer (SCOT): an international, randomised, phase 3, non-inferiority trial. *The Lancet Oncology*, 19 (4):562–578, April 2018. ISSN 14702045. doi: 10.1016/S1470-2045(18)30093-7. URL `https://linkinghub.elsevier.com/retrieve/pii/S1470204518300937`.

[177] Xiaoming Jia, Buhm Han, Suna Onengut-Gumuscu, Wei-Min Chen, Patrick J. Concannon, Stephen S. Rich, Soumya Raychaudhuri, and Paul I.W. de Bakker. Imputing Amino Acid Polymorphisms in Human Leukocyte Antigens. *PLoS ONE*, 8(6): e64683, June 2013. ISSN 1932-6203. doi: 10.1371/journal.pone.0064683. URL `https://dx.plos.org/10.1371/journal.pone.0064683`.

[178] Sharon R. Browning and Brian L. Browning. Rapid and Accurate Haplotype Phasing and Missing-Data Inference for Whole-Genome Association Studies By Use of Localized Haplotype Clustering. *The American Journal of Human Genetics*, 81 (5):1084–1097, November 2007. ISSN 00029297. doi: 10.1086/521987. URL `https://linkinghub.elsevier.com/retrieve/pii/S0002929707638828`.

[179] A. Sette and J. Sidney. Nine major HLA class I supertypes account for the vast preponderance of HLA-A and -B polymorphism. *Immunogenetics*, 50(3-4):201–212, November 1999. ISSN 0093-7711, 1432-1211. doi: 10.1007/s002510050594. URL `http://link.springer.com/10.1007/s002510050594`.

[180] John Sidney, Bjoern Peters, Nicole Frahm, Christian Brander, and Alessandro Sette. HLA class I supertypes: a revised and updated classification. *BMC Immunology*, 9 (1):1, 2008. ISSN 1471-2172. doi: 10.1186/1471-2172-9-1. URL `http://bmcimmunol.biomedcentral.com/articles/10.1186/1471-2172-9-1`.

[181] Andrew O. Giacomelli, Xiaoping Yang, Robert E. Lintner, James M. McFarland, Marc Duby, Jaegil Kim, Thomas P. Howard, David Y. Takeda, Seav Huong Ly, Eejung Kim, Hugh S. Gannon, Brian Hurhula, Ted Sharpe, Amy Goodale, Briana Fritchman, Scott Steelman, Francisca Vazquez, Aviad Tsherniak, Andrew J. Aguirre, John G. Doench, Federica Piccioni, Charles W. M. Roberts, Matthew Meyerson, Gad Getz, Cory M. Johannessen, David E. Root, and William C. Hahn. Mutational processes shape the landscape of TP53 mutations in human cancer. *Nature Genetics*, 50(10):1381–1387, October 2018. ISSN 1061-4036, 1546-1718. doi: 10.1038/s41588-018-0204-y. URL `http://www.nature.com/articles/s41588-018-0204-y`.

[182] Rick Jansen, Jouke-Jan Hottenga, Michel G. Nivard, Abdel Abdellaoui, Bram Laport, Eco J. de Geus, Fred A. Wright, Brenda W.J.H. Penninx, and Dorret I. Boomsma. Conditional eQTL analysis reveals allelic heterogeneity of gene expression. *Human Molecular Genetics*, 26(8):1444–1451, April 2017. ISSN 0964-6906,

1460-2083. doi: 10.1093/hmg/ddx043. URL https://academic.oup.com/hmg/article/26/8/1444/2970473.

[183] K. Wang, M. Li, and H. Hakonarson. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Research*, 38(16): e164–e164, September 2010. ISSN 0305-1048, 1362-4962. doi: 10.1093/nar/gkq603. URL https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkq603.

[184] Graham R. S. Ritchie, Ian Dunham, Eleftheria Zeggini, and Paul Flicek. Functional annotation of noncoding sequence variants. *Nature Methods*, 11(3):294–296, March 2014. ISSN 1548-7105. doi: 10.1038/nmeth.2832.

[185] The ENCODE Project Consortium, Jill E. Moore, Michael J. Purcaro, Henry E. Pratt, Charles B. Epstein, Noam Shoresh, Jessika Adrian, Trupti Kawli, Carrie A. Davis, Alexander Dobin, Rajinder Kaul, Jessica Halow, Eric L. Van Nostrand, Peter Freese, David U. Gorkin, Yin Shen, Yupeng He, Mark Mackiewicz, Florencia Pauli-Behn, Brian A. Williams, Ali Mortazavi, Cheryl A. Keller, Xiao-Ou Zhang, Shaimae I. Elhajjajy, Jack Huey, Diane E. Dickel, Valentina Snetkova, Xintao Wei, Xiaofeng Wang, Juan Carlos Rivera-Mulia, Joel Rozowsky, Jing Zhang, Surya B. Chhetri, Jialing Zhang, Alec Victorsen, Kevin P. White, Axel Visel, Gene W. Yeo, Christopher B. Burge, Eric Lécuyer, David M. Gilbert, Job Dekker, John Rinn, Eric M. Mendenhall, Joseph R. Ecker, Manolis Kellis, Robert J. Klein, William S. Noble, Anshul Kundaje, Roderic Guigó, Peggy J. Farnham, J. Michael Cherry, Richard M. Myers, Bing Ren, Brenton R. Graveley, Mark B. Gerstein, Len A. Pennacchio, Michael P. Snyder, Bradley E. Bernstein, Barbara Wold, Ross C. Hardison, Thomas R. Gingeras, John A. Stamatoyannopoulos, and Zhiping Weng. Expanded encyclopaedias of DNA elements in the human and mouse genomes. *Nature*, 583 (7818):699–710, July 2020. ISSN 0028-0836, 1476-4687. doi: 10.1038/s41586-020-2493-4. URL http://www.nature.com/articles/s41586-020-2493-4.

[186] C. J. Willer, Y. Li, and G. R. Abecasis. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics*, 26(17):2190–2191, September 2010. ISSN 1367-4803, 1460-2059. doi: 10.1093/bioinformatics/btq340. URL https://academic.oup.com/bioinformatics/article-lookup/doi/10.1093/bioinformatics/btq340.

[187] M Fischer. Census and evaluation of p53 target genes. *Oncogene*, 36(28):3943–3956, July 2017. ISSN 0950-9232, 1476-5594. doi: 10.1038/onc.2016.502. URL http://www.nature.com/articles/onc2016502.

[188] Giovanni Stracquadanio, Xuting Wang, Marsha D. Wallace, Anna M. Grawenda, Ping Zhang, Juliet Hewitt, Jorge Zeron-Medina, Francesc Castro-Giner, Ian P. Tomlinson, Colin R. Goding, Kamil J. Cygan, William G. Fairbrother, Laurent F. Thomas, Pål Sætrom, Federica Gemignani, Stefano Landi, Benjamin Schuster-Böckler, Douglas A. Bell, and Gareth L. Bond. The importance of p53 pathway genetics in inherited and somatic cancer genomes. *Nature Reviews Cancer*, 16(4): 251–265, April 2016. ISSN 1474-175X, 1474-1768. doi: 10.1038/nrc.2016.15. URL http://www.nature.com/articles/nrc.2016.15.

[189] Jianfang Liu, Tara Lichtenberg, Katherine A. Hoadley, Laila M. Poisson, Alexander J. Lazar, Andrew D. Cherniack, Albert J. Kovatich, Christopher C. Benz, Douglas A. Levine, Adrian V. Lee, Larsson Omberg, Denise M. Wolf, Craig D. Shriver, Vesteinn Thorsson, Cancer Genome Atlas Research Network, and Hai Hu. An Integrated TCGA Pan-Cancer Clinical Data Resource to Drive High-Quality Survival Outcome Analytics. *Cell*, 173(2):400–416.e11, April 2018. ISSN 1097-4172. doi: 10.1016/j.cell.2018.02.052.

[190] Jonathan Marchini, Bryan Howie, Simon Myers, Gil McVean, and Peter Donnelly. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nature Genetics*, 39(7):906–913, July 2007. ISSN 1061-4036, 1546-1718. doi: 10.1038/ng2088. URL `http://www.nature.com/articles/ng2088`.

[191] S. Purcell, S. S. Cherny, and P. C. Sham. Genetic Power Calculator: design of linkage and association genetic mapping studies of complex traits. *Bioinformatics (Oxford, England)*, 19(1):149–150, January 2003. ISSN 1367-4803.

[192] Randall J. Pruim, Ryan P. Welch, Serena Sanna, Tanya M. Teslovich, Peter S. Chines, Terry P. Gliedt, Michael Boehnke, Gonçalo R. Abecasis, and Cristen J. Willer. LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics (Oxford, England)*, 26(18):2336–2337, September 2010. ISSN 1367-4811. doi: 10.1093/bioinformatics/btq419.

[193] Abbas A Rizvi, Ezgi Karaesmen, Martin Morgan, Leah Preus, Junke Wang, Michael Sovic, Theresa Hahn, and Lara E Sucheston-Campbell. gwasurvivr: an R package for genome-wide survival analysis. *Bioinformatics*, 35(11):1968–1970, June 2019. ISSN 1367-4803, 1460-2059. doi: 10.1093/bioinformatics/bty920. URL `https://academic.oup.com/bioinformatics/article/35/11/1968/5161085`.

[194] Mark A. Glaire, Enric Domingo, Anita Sveen, Jarle Bruun, Arild Nesbakken, George Nicholson, Marco Novelli, Kay Lawson, Dahmane Oukrif, Wanja Kildal, Havard E. Danielsen, Rachel Kerr, David Kerr, Ian Tomlinson, Ragnhild A. Lothe, and David N. Church. Tumour-infiltrating CD8+ lymphocytes and colorectal cancer recurrence by tumour and nodal stage. *British Journal of Cancer*, 121(6):474–482, September 2019. ISSN 0007-0920, 1532-1827. doi: 10.1038/s41416-019-0540-4. URL `http://www.nature.com/articles/s41416-019-0540-4`.

[195] Shing Wan Choi and Paul F O'Reilly. PRSice-2: Polygenic Risk Score software for biobank-scale data. *GigaScience*, 8(7):giz082, July 2019. ISSN 2047-217X. doi: 10.1093/gigascience/giz082. URL `https://academic.oup.com/gigascience/article/doi/10.1093/gigascience/giz082/5532407`.

[196] Stephen Burgess, Dylan S Small, and Simon G Thompson. A review of instrumental variable estimators for Mendelian randomization. *Statistical Methods in Medical Research*, 26(5):2333–2355, October 2017. ISSN 0962-2802, 1477-0334. doi: 10.1177/0962280215597579. URL `http://journals.sagepub.com/doi/10.1177/0962280215597579`.

[197] D. N. Reshef, Y. A. Reshef, H. K. Finucane, S. R. Grossman, G. McVean, P. J. Turnbaugh, E. S. Lander, M. Mitzenmacher, and P. C. Sabeti. Detecting Novel Associations in Large Data Sets. *Science*, 334(6062):1518–1524, December 2011.

ISSN 0036-8075, 1095-9203. doi: 10.1126/science.1205438. URL https://www.sciencemag.org/lookup/doi/10.1126/science.1205438.

[198] A. J. Levine. p53, the cellular gatekeeper for growth and division. *Cell*, 88(3): 323–331, February 1997. ISSN 0092-8674. doi: 10.1016/s0092-8674(00)81871-1.

[199] B. Vogelstein, D. Lane, and A. J. Levine. Surfing the p53 network. *Nature*, 408 (6810):307–310, November 2000. ISSN 0028-0836. doi: 10.1038/35042675.

[200] Edward R. Kastenhuber and Scott W. Lowe. Putting p53 in Context. *Cell*, 170(6): 1062–1078, September 2017. ISSN 00928674. doi: 10.1016/j.cell.2017.08.028. URL https://linkinghub.elsevier.com/retrieve/pii/S0092867417309534.

[201] J. E. Purvis, K. W. Karhohs, C. Mock, E. Batchelor, A. Loewer, and G. Lahav. p53 Dynamics Control Cell Fate. *Science*, 336(6087):1440–1444, June 2012. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.1218351. URL https://www.sciencemag.org/lookup/doi/10.1126/science.1218351.

[202] Zdenek Andrysik, Matthew D. Galbraith, Anna L. Guarnieri, Sara Zaccara, Kelly D. Sullivan, Ahwan Pandey, Morgan MacBeth, Alberto Inga, and Joaquín M. Espinosa. Identification of a core TP53 transcriptional program with highly distributed tumor suppressive activity. *Genome Research*, 27(10):1645–1657, October 2017. ISSN 1088-9051, 1549-5469. doi: 10.1101/gr.220533.117. URL http://genome.cshlp.org/lookup/doi/10.1101/gr.220533.117.

[203] Clemens A Schmitt, Jordan S Fridman, Meng Yang, Eugene Baranov, Robert M Hoffman, and Scott W Lowe. Dissecting p53 tumor suppressor functions in vivo. *Cancer Cell*, 1(3):289–298, April 2002. ISSN 15356108. doi: 10.1016/S1535-6108(02)00047-8. URL https://linkinghub.elsevier.com/retrieve/pii/S1535610802000478.

[204] H F Horn and K H Vousden. Coping with stress: multiple ways to activate p53. *Oncogene*, 26(9):1306–1316, February 2007. ISSN 0950-9232, 1476-5594. doi: 10.1038/sj.onc.1210263. URL http://www.nature.com/articles/1210263.

[205] M Hollstein, D Sidransky, B Vogelstein, and C. Harris. p53 mutations in human cancers. *Science*, 253(5015):49–53, July 1991. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.1905840. URL https://www.sciencemag.org/lookup/doi/10.1126/science.1905840.

[206] Lawrence A. Donehower, Thierry Soussi, Anil Korkut, Yuexin Liu, Andre Schultz, Maria Cardenas, Xubin Li, Ozgun Babur, Teng-Kuei Hsu, Olivier Lichtarge, John N. Weinstein, Rehan Akbani, and David A. Wheeler. Integrated Analysis of TP53 Gene and Pathway Alterations in The Cancer Genome Atlas. *Cell Reports*, 28(5):1370–1384.e5, July 2019. ISSN 22111247. doi: 10.1016/j.celrep.2019.07.001. URL https://linkinghub.elsevier.com/retrieve/pii/S221112471930885X.

[207] Shunsuke Kato, Shuang-Yin Han, Wen Liu, Kazunori Otsuka, Hiroyuki Shibata, Ryunosuke Kanamaru, and Chikashi Ishioka. Understanding the function-structure and function-mutation relationships of p53 tumor suppressor protein by high-resolution missense mutation analysis. *Proceedings of the National Academy of Sciences of the United States of America*, 100(14):8424–8429, July 2003. ISSN 0027-8424. doi: 10.1073/pnas.1431692100.

[208] Eran Kotler, Odem Shani, Guy Goldfeld, Maya Lotan-Pompan, Ohad Tarcic, Anat Gershoni, Thomas A. Hopf, Debora S. Marks, Moshe Oren, and Eran Segal. A Systematic p53 Mutation Library Links Differential Functional Impact to Cancer Mutation Pattern and Evolutionary Conservation. *Molecular Cell*, 71(1):178–190.e8, July 2018. ISSN 10972765. doi: 10.1016/j.molcel.2018.06.012. URL `https://linkinghub.elsevier.com/retrieve/pii/S1097276518304544`.

[209] Theo A. Knijnenburg, Linghua Wang, Michael T. Zimmermann, Nyasha Chambwe, Galen F. Gao, Andrew D. Cherniack, Huihui Fan, Hui Shen, Gregory P. Way, Casey S. Greene, Yuexin Liu, Rehan Akbani, Bin Feng, Lawrence A. Donehower, Chase Miller, Yang Shen, Mostafa Karimi, Haoran Chen, Pora Kim, Peilin Jia, Eve Shinbrot, Shaojun Zhang, Jianfang Liu, Hai Hu, Matthew H. Bailey, Christina Yau, Denise Wolf, Zhongming Zhao, John N. Weinstein, Lei Li, Li Ding, Gordon B. Mills, Peter W. Laird, David A. Wheeler, Ilya Shmulevich, Raymond J. Monnat, Yonghong Xiao, Chen Wang, Samantha J. Caesar-Johnson, John A. Demchok, Ina Felau, Melpomeni Kasapi, Martin L. Ferguson, Carolyn M. Hutter, Heidi J. Sofia, Roy Tarnuzzer, Zhining Wang, Liming Yang, Jean C. Zenklusen, Jiashan (Julia) Zhang, Sudha Chudamani, Jia Liu, Laxmi Lolla, Rashi Naresh, Todd Pihl, Qiang Sun, Yunhu Wan, Ye Wu, Juok Cho, Timothy DeFreitas, Scott Frazer, Nils Gehlenborg, Gad Getz, David I. Heiman, Jaegil Kim, Michael S. Lawrence, Pei Lin, Sam Meier, Michael S. Noble, Gordon Saksena, Doug Voet, Hailei Zhang, Brady Bernard, Nyasha Chambwe, Varsha Dhankani, Theo Knijnenburg, Roger Kramer, Kalle Leinonen, Yuexin Liu, Michael Miller, Sheila Reynolds, Ilya Shmulevich, Vesteinn Thorsson, Wei Zhang, Rehan Akbani, Bradley M. Broom, Apurva M. Hegde, Zhenlin Ju, Rupa S. Kanchi, Anil Korkut, Jun Li, Han Liang, Shiyun Ling, Wenbin Liu, Yiling Lu, Gordon B. Mills, Kwok-Shing Ng, Arvind Rao, Michael Ryan, Jing Wang, John N. Weinstein, Jiexin Zhang, Adam Abeshouse, Joshua Armenia, Debyani Chakravarty, Walid K. Chatila, Ino de Bruijn, Jianjiong Gao, Benjamin E. Gross, Zachary J. Heins, Ritika Kundra, Konnor La, Marc Ladanyi, Augustin Luna, Moriah G. Nissan, Angelica Ochoa, Sarah M. Phillips, Ed Reznik, Francisco Sanchez-Vega, Chris Sander, Nikolaus Schultz, Robert Sheridan, S. Onur Sumer, Yichao Sun, Barry S. Taylor, Jioajiao Wang, Hongxin Zhang, Pavana Anur, Myron Peto, Paul Spellman, Christopher Benz, Joshua M. Stuart, Christopher K. Wong, Christina Yau, D. Neil Hayes, Joel S. Parker, Matthew D. Wilkerson, Adrian Ally, Miruna Balasundaram, Reanne Bowlby, Denise Brooks, Rebecca Carlsen, Eric Chuah, Noreen Dhalla, Robert Holt, Steven J.M. Jones, Katayoon Kasaian, Darlene Lee, Yussanne Ma, Marco A. Marra, Michael Mayo, Richard A. Moore, Andrew J. Mungall, Karen Mungall, A. Gordon Robertson, Sara Sadeghi, Jacqueline E. Schein, Payal Sipahimalani, Angela Tam, Nina Thiessen, Kane Tse, Tina Wong, Ashton C. Berger, Rameen Beroukhim, Andrew D. Cherniack, Carrie Cibulskis, Stacey B. Gabriel, Galen F. Gao, Gavin Ha, Matthew Meyerson, Steven E. Schumacher, Juliann Shih, Melanie H. Kucherlapati, Raju S. Kucherlapati, Stephen Baylin, Leslie Cope, Ludmila Danilova, Moiz S. Bootwalla, Phillip H. Lai, Dennis T. Maglinte, David J. Van Den Berg, Daniel J. Weisenberger, J. Todd Au-

man, Saianand Balu, Tom Bodenheimer, Cheng Fan, Katherine A. Hoadley, Alan P. Hoyle, Stuart R. Jefferys, Corbin D. Jones, Shaowu Meng, Piotr A. Mieczkowski, Lisle E. Mose, Amy H. Perou, Charles M. Perou, Jeffrey Roach, Yan Shi, Janae V. Simons, Tara Skelly, Matthew G. Soloway, Donghui Tan, Umadevi Veluvolu, Huihui Fan, Toshinori Hinoue, Peter W. Laird, Hui Shen, Wanding Zhou, Michelle Bellair, Kyle Chang, Kyle Covington, Chad J. Creighton, Huyen Dinh, HarshaVardhan Doddapaneni, Lawrence A. Donehower, Jennifer Drummond, Richard A. Gibbs, Robert Glenn, Walker Hale, Yi Han, Jianhong Hu, Viktoriya Korchina, Sandra Lee, Lora Lewis, Wei Li, Xiuping Liu, Margaret Morgan, Donna Morton, Donna Muzny, Jireh Santibanez, Margi Sheth, Eve Shinbrot, Linghua Wang, Min Wang, David A. Wheeler, Liu Xi, Fengmei Zhao, Julian Hess, Elizabeth L. Appelbaum, Matthew Bailey, Matthew G. Cordes, Li Ding, Catrina C. Fronick, Lucinda A. Fulton, Robert S. Fulton, Cyriac Kandoth, Elaine R. Mardis, Michael D. McLellan, Christopher A. Miller, Heather K. Schmidt, Richard K. Wilson, Daniel Crain, Erin Curley, Johanna Gardner, Kevin Lau, David Mallery, Scott Morris, Joseph Paulauskis, Robert Penny, Candace Shelton, Troy Shelton, Mark Sherman, Eric Thompson, Peggy Yena, Jay Bowen, Julie M. Gastier-Foster, Mark Gerken, Kristen M. Leraas, Tara M. Lichtenberg, Nilsa C. Ramirez, Lisa Wise, Erik Zmuda, Niall Corcoran, Tony Costello, Christopher Hovens, Andre L. Carvalho, Ana C. de Carvalho, José H. Fregnani, Adhemar Longatto-Filho, Rui M. Reis, Cristovam Scapulatempo-Neto, Henrique C.S. Silveira, Daniel O. Vidal, Andrew Burnette, Jennifer Eschbacher, Beth Hermes, Ardene Noss, Rosy Singh, Matthew L. Anderson, Patricia D. Castro, Michael Ittmann, David Huntsman, Bernard Kohl, Xuan Le, Richard Thorp, Chris Andry, Elizabeth R. Duffy, Vladimir Lyadov, Oxana Paklina, Galiya Setdikova, Alexey Shabunin, Mikhail Tavobilov, Christopher McPherson, Ronald Warnick, Ross Berkowitz, Daniel Cramer, Colleen Feltmate, Neil Horowitz, Adam Kibel, Michael Muto, Chandrajit P. Raut, Andrei Malykh, Jill S. Barnholtz-Sloan, Wendi Barrett, Karen Devine, Jordonna Fulop, Quinn T. Ostrom, Kristen Shimmel, Yingli Wolinsky, Andrew E. Sloan, Agostino De Rose, Felice Giuliante, Marc Goodman, Beth Y. Karlan, Curt H. Hagedorn, John Eckman, Jodi Harr, Jerome Myers, Kelinda Tucker, Leigh Anne Zach, Brenda Deyarmin, Hai Hu, Leonid Kvecher, Caroline Larson, Richard J. Mural, Stella Somiari, Ales Vicha, Tomas Zelinka, Joseph Bennett, Mary Iacocca, Brenda Rabeno, Patricia Swanson, Mathieu Latour, Louis Lacombe, Bernard Têtu, Alain Bergeron, Mary McGraw, Susan M. Staugaitis, John Chabot, Hanina Hibshoosh, Antonia Sepulveda, Tao Su, Timothy Wang, Olga Potapova, Olga Voronina, Laurence Desjardins, Odette Mariani, Sergio Roman-Roman, Xavier Sastre, Marc-Henri Stern, Feixiong Cheng, Sabina Signoretti, Andrew Berchuck, Darell Bigner, Eric Lipp, Jeffrey Marks, Shannon McCall, Roger McLendon, Angeles Secord, Alexis Sharp, Madhusmita Behera, Daniel J. Brat, Amy Chen, Keith Delman, Seth Force, Fadlo Khuri, Kelly Magliocca, Shishir Maithel, Jeffrey J. Olson, Taofeek Owonikoko, Alan Pickens, Suresh Ramalingam, Dong M. Shin, Gabriel Sica, Erwin G. Van Meir, Hongzheng Zhang, Wil Eijckenboom, Ad Gillis, Esther Korpershoek, Leendert Looijenga, Wolter Ooster-

huis, Hans Stoop, Kim E. van Kessel, Ellen C. Zwarthoff, Chiara Calatozzolo, Lucia Cuppini, Stefania Cuzzubbo, Francesco DiMeco, Gaetano Finocchiaro, Luca Mattei, Alessandro Perin, Bianca Pollo, Chu Chen, John Houck, Pawadee Lohavanichbutr, Arndt Hartmann, Christine Stoehr, Robert Stoehr, Helge Taubert, Sven Wach, Bernd Wullich, Witold Kycler, Dawid Murawa, Maciej Wiznerowicz, Ki Chung, W. Jeffrey Edenfield, Julie Martin, Eric Baudin, Glenn Bubley, Raphael Bueno, Assunta De Rienzo, William G. Richards, Steven Kalkanis, Tom Mikkelsen, Houtan Noushmehr, Lisa Scarpace, Nicolas Girard, Marta Aymerich, Elias Campo, Eva Giné, Armando López Guillermo, Nguyen Van Bang, Phan Thi Hanh, Bui Duc Phu, Yufang Tang, Howard Colman, Kimberley Evason, Peter R. Dottino, John A. Martignetti, Hani Gabra, Hartmut Juhl, Teniola Akeredolu, Serghei Stepa, Dave Hoon, Keunsoo Ahn, Koo Jeong Kang, Felix Beuschlein, Anne Breggia, Michael Birrer, Debra Bell, Mitesh Borad, Alan H. Bryce, Erik Castle, Vishal Chandan, John Cheville, John A. Copland, Michael Farnell, Thomas Flotte, Nasra Giama, Thai Ho, Michael Kendrick, Jean-Pierre Kocher, Karla Kopp, Catherine Moser, David Nagorney, Daniel O'Brien, Brian Patrick O'Neill, Tushar Patel, Gloria Petersen, Florencia Que, Michael Rivera, Lewis Roberts, Robert Smallridge, Thomas Smyrk, Melissa Stanton, R. Houston Thompson, Michael Torbenson, Ju Dong Yang, Lizhi Zhang, Fadi Brimo, Jaffer A. Ajani, Ana Maria Angulo Gonzalez, Carmen Behrens, Jolanta Bondaruk, Russell Broaddus, Bogdan Czerniak, Bita Esmaeli, Junya Fujimoto, Jeffrey Gershenwald, Charles Guo, Alexander J. Lazar, Christopher Logothetis, Funda Meric-Bernstam, Cesar Moran, Lois Ramondetta, David Rice, Anil Sood, Pheroze Tamboli, Timothy Thompson, Patricia Troncoso, Anne Tsao, Ignacio Wistuba, Candace Carter, Lauren Haydu, Peter Hersey, Valerie Jakrot, Hojabr Kakavand, Richard Kefford, Kenneth Lee, Georgina Long, Graham Mann, Michael Quinn, Robyn Saw, Richard Scolyer, Kerwin Shannon, Andrew Spillane, Jonathan Stretch, Maria Synott, John Thompson, James Wilmott, Hikmat Al-Ahmadie, Timothy A. Chan, Ronald Ghossein, Anuradha Gopalan, Douglas A. Levine, Victor Reuter, Samuel Singer, Bhuvanesh Singh, Nguyen Viet Tien, Thomas Broudy, Cyrus Mirsaidi, Praveen Nair, Paul Drwiega, Judy Miller, Jennifer Smith, Howard Zaren, Joong-Won Park, Nguyen Phi Hung, Electron Kebebew, W. Marston Linehan, Adam R. Metwalli, Karel Pacak, Peter A. Pinto, Mark Schiffman, Laura S. Schmidt, Cathy D. Vocke, Nicolas Wentzensen, Robert Worrell, Hannah Yang, Marc Moncrieff, Chandra Goparaju, Jonathan Melamed, Harvey Pass, Natalia Botnariuc, Irina Caraman, Mircea Cernat, Inga Chemencedji, Adrian Clipca, Serghei Doruc, Ghenadie Gorincioi, Sergiu Mura, Maria Pirtac, Irina Stancul, Diana Tcaciuc, Monique Albert, Iakovina Alexopoulou, Angel Arnaout, John Bartlett, Jay Engel, Sebastien Gilbert, Jeremy Parfitt, Harman Sekhon, George Thomas, Doris M. Rassl, Robert C. Rintoul, Carlo Bifulco, Raina Tamakawa, Walter Urba, Nicholas Hayward, Henri Timmers, Anna Antenucci, Francesco Facciolo, Gianluca Grazi, Mirella Marino, Roberta Merola, Ronald de Krijger, Anne-Paule Gimenez-Roqueplo, Alain Piché, Simone Chevalier, Ginette McKercher, Kivanc Birsoy, Gene Barnett, Cathy Brewer, Carol Farver, Theresa Naska, Nathan A. Pennell, Daniel Raymond, Cathy Schilero, Kathy

Smolenski, Felicia Williams, Carl Morrison, Jeffrey A. Borgia, Michael J. Liptay, Mark Pool, Christopher W. Seder, Kerstin Junker, Larsson Omberg, Mikhail Dinkin, George Manikhas, Domenico Alvaro, Maria Consiglia Bragazzi, Vincenzo Cardinale, Guido Carpino, Eugenio Gaudio, David Chesla, Sandra Cottingham, Michael Dubina, Fedor Moiseenko, Renumathy Dhanasekaran, Karl-Friedrich Becker, Klaus-Peter Janssen, Julia Slotta-Huspenina, Mohamed H. Abdel-Rahman, Dina Aziz, Sue Bell, Colleen M. Cebulla, Amy Davis, Rebecca Duell, J. Bradley Elder, Joe Hilty, Bahavna Kumar, James Lang, Norman L. Lehman, Randy Mandt, Phuong Nguyen, Robert Pilarski, Karan Rai, Lynn Schoenfield, Kelly Senecal, Paul Wakely, Paul Hansen, Ronald Lechan, James Powers, Arthur Tischler, William E. Grizzle, Katherine C. Sexton, Alison Kastl, Joel Henderson, Sima Porten, Jens Waldmann, Martin Fassnacht, Sylvia L. Asa, Dirk Schadendorf, Marta Couce, Markus Graefen, Hartwig Huland, Guido Sauter, Thorsten Schlomm, Ronald Simon, Pierre Tennstedt, Oluwole Olabode, Mark Nelson, Oliver Bathe, Peter R. Carroll, June M. Chan, Philip Disaia, Pat Glenn, Robin K. Kelley, Charles N. Landen, Joanna Phillips, Michael Prados, Jeffry Simko, Karen Smith-McCune, Scott VandenBerg, Kevin Roggin, Ashley Fehrenbach, Ady Kendler, Suzanne Sifri, Ruth Steele, Antonio Jimeno, Francis Carey, Ian Forgie, Massimo Mannelli, Michael Carney, Brenda Hernandez, Benito Campos, Christel Herold-Mende, Christin Jungk, Andreas Unterberg, Andreas von Deimling, Aaron Bossler, Joseph Galbraith, Laura Jacobus, Michael Knudson, Tina Knutson, Deqin Ma, Mohammed Milhem, Rita Sigmund, Andrew K. Godwin, Rashna Madan, Howard G. Rosenthal, Clement Adebamowo, Sally N. Adebamowo, Alex Boussioutas, David Beer, Thomas Giordano, Anne-Marie Mes-Masson, Fred Saad, Therese Bocklage, Lisa Landrum, Robert Mannel, Kathleen Moore, Katherine Moxley, Russel Postier, Joan Walker, Rosemary Zuna, Michael Feldman, Federico Valdivieso, Rajiv Dhir, James Luketich, Edna M. Mora Pinero, Mario Quintero-Aguilo, Carlos Gilberto Carlotti, Jose Sebastião Dos Santos, Rafael Kemp, Ajith Sankarankuty, Daniela Tirapelli, James Catto, Kathy Agnew, Elizabeth Swisher, Jenette Creaney, Bruce Robinson, Carl Simon Shelley, Eryn M. Godwin, Sara Kendall, Cassaundra Shipman, Carol Bradford, Thomas Carey, Andrea Haddad, Jeffey Moyer, Lisa Peterson, Mark Prince, Laura Rozek, Gregory Wolf, Rayleen Bowman, Kwun M. Fong, Ian Yang, Robert Korst, W. Kimryn Rathmell, J. Leigh Fantacone-Campbell, Jeffrey A. Hooke, Albert J. Kovatich, Craig D. Shriver, John DiPersio, Bettina Drake, Ramaswamy Govindan, Sharon Heath, Timothy Ley, Brian Van Tine, Peter Westervelt, Mark A. Rubin, Jung Il Lee, Natália D. Aredes, and Armaz Mariamidze. Genomic and Molecular Landscape of DNA Damage Repair Deficiency across The Cancer Genome Atlas. *Cell Reports*, 23(1):239–254.e6, April 2018. ISSN 22111247. doi: 10.1016/j.celrep.2018.03.076. URL `https://linkinghub.elsevier.com/retrieve/pii/S2211124718304376`.

[210] Jonathan D. Oliner, Anne Y. Saiki, and Sean Caenepeel. The Role of MDM2 Amplification and Overexpression in Tumorigenesis. *Cold Spring Harbor Perspectives in Medicine*, 6(6), 2016. ISSN 2157-1422. doi: 10.1101/cshperspect.a026336.

[211] Thibaut Barnoud, Joshua L D Parris, and Maureen E Murphy. Common genetic

variants in the TP53 pathway and their impact on cancer. *Journal of Molecular Cell Biology*, 11(7):578–585, July 2019. ISSN 1759-4685. doi: 10.1093/jmcb/mjz052. URL https://academic.oup.com/jmcb/article/11/7/578/5509869.

[212] the CORGI Consortium, Ian Tomlinson, Emily Webb, Luis Carvajal-Carmona, Peter Broderick, Zoe Kemp, Sarah Spain, Steven Penegar, Ian Chandler, Maggie Gorman, Wendy Wood, Ella Barclay, Steven Lubbe, Lynn Martin, Gabrielle Sellick, Emma Jaeger, Richard Hubner, Ruth Wild, Andrew Rowan, Sarah Fielding, Kimberley Howarth, Andrew Silver, Wendy Atkin, Kenneth Muir, Richard Logan, David Kerr, Elaine Johnstone, Oliver Sieber, Richard Gray, Huw Thomas, Julian Peto, Jean-Baptiste Cazier, and Richard Houlston. A genome-wide association scan of tag SNPs identifies a susceptibility variant for colorectal cancer at 8q24.21. *Nature Genetics*, 39(8):984–988, August 2007. ISSN 1061-4036, 1546-1718. doi: 10.1038/ng2085. URL http://www.nature.com/articles/ng2085.

[213] Lambertus A Kiemeney, Steinunn Thorlacius, Patrick Sulem, Frank Geller, Katja K H Aben, Simon N Stacey, Julius Gudmundsson, Margret Jakobsdottir, Jon T Bergthorsson, Asgeir Sigurdsson, Thorarinn Blondal, J Alfred Witjes, Sita H Vermeulen, Christina A Hulsbergen-van de Kaa, Dorine W Swinkels, Martine Ploeg, Erik B Cornel, Henk Vergunst, Thorgeir E Thorgeirsson, Daniel Gudbjartsson, Sigurjon A Gudjonsson, Gudmar Thorleifsson, Kari T Kristinsson, Magali Mouy, Steinunn Snorradottir, Donatella Placidi, Marcello Campagna, Cecilia Arici, Kvetoslava Koppova, Eugene Gurzau, Peter Rudnai, Eliane Kellen, Silvia Polidoro, Simonetta Guarrera, Carlotta Sacerdote, Manuel Sanchez, Berta Saez, Gabriel Valdivia, Charlotta Ryk, Petra de Verdier, Annika Lindblom, Klaus Golka, D Timothy Bishop, Margaret A Knowles, Sigfus Nikulasson, Vigdis Petursdottir, Eirikur Jonsson, Gudmundur Geirsson, Baldvin Kristjansson, Jose I Mayordomo, Gunnar Steineck, Stefano Porru, Frank Buntinx, Maurice P Zeegers, Tony Fletcher, Rajiv Kumar, Giuseppe Matullo, Paolo Vineis, Anne E Kiltie, Jeffrey R Gulcher, Unnur Thorsteinsdottir, Augustine Kong, Thorunn Rafnar, and Kari Stefansson. Sequence variant on 8q24 confers susceptibility to urinary bladder cancer. *Nature Genetics*, 40 (11):1307–1312, November 2008. ISSN 1061-4036, 1546-1718. doi: 10.1038/ng.229. URL http://www.nature.com/articles/ng.229.

[214] The UK Genetic Prostate Cancer Study Collaborators/British Association of Urological Surgeons' Section of Oncology, The UK Prostate testing for cancer and Treatment study (ProtecT Study) Collaborators, Ali Amin Al Olama, Zsofia Kote-Jarai, Graham G Giles, Michelle Guy, Jonathan Morrison, Gianluca Severi, Daniel A Leongamornlert, Malgorzata Tymrakiewicz, Sameer Jhavar, Ed Saunders, John L Hopper, Melissa C Southey, Kenneth R Muir, Dallas R English, David P Dearnaley, Audrey T Ardern-Jones, Amanda L Hall, Lynne T O'Brien, Rosemary A Wilkinson, Emma Sawyer, Artitaya Lophatananon, Alan Horwich, Robert A Huddart, Vincent S Khoo, Christopher C Parker, Christopher J Woodhouse, Alan Thompson, Tim Christmas, Chris Ogden, Colin Cooper, Jenny L Donovan, Freddie C Hamdy, David E Neal, Rosalind A Eeles, and Douglas F Easton. Multiple loci on 8q24 associated with prostate cancer susceptibility. *Nature Genetics*, 41(10):

1058–1060, October 2009. ISSN 1061-4036, 1546-1718. doi: 10.1038/ng.452. URL http://www.nature.com/articles/ng.452.

[215] Dalemari Crowther-Swanepoel, Peter Broderick, Maria Chiara Di Bernardo, Sara E Dobbins, María Torres, Mahmoud Mansouri, Clara Ruiz-Ponte, Anna Enjuanes, Richard Rosenquist, Angel Carracedo, Jesper Jurlander, Elias Campo, Gunnar Juliusson, Emilio Montserrat, Karin E Smedby, Martin J S Dyer, Estella Matutes, Claire Dearden, Nicola J Sunter, Andrew G Hall, Tryfonia Mainou-Fowler, Graham H Jackson, Geoffrey Summerfield, Robert J Harris, Andrew R Pettitt, David J Allsup, James R Bailey, Guy Pratt, Chris Pepper, Chris Fegan, Anton Parker, David Oscier, James M Allan, Daniel Catovsky, and Richard S Houlston. Common variants at 2q37.3, 8q24.21, 15q21.3 and 16q24.1 influence chronic lymphocytic leukemia risk. *Nature Genetics*, 42(2):132–136, February 2010. ISSN 1061-4036, 1546-1718. doi: 10.1038/ng.510. URL http://www.nature.com/articles/ng.510.

[216] Sara R. Rashkin, Rebecca E. Graff, Linda Kachuri, Khanh K. Thai, Stacey E. Alexeeff, Maruta A. Blatchins, Taylor B. Cavazos, Douglas A. Corley, Nima C. Emami, Joshua D. Hoffman, Eric Jorgenson, Lawrence H. Kushi, Travis J. Meyers, Stephen K. Van Den Eeden, Elad Ziv, Laurel A. Habel, Thomas J. Hoffmann, Lori C. Sakoda, and John S. Witte. Pan-cancer study detects genetic risk variants and shared genetic basis in two large cohorts. *Nature Communications*, 11(1):4423, September 2020. ISSN 2041-1723. doi: 10.1038/s41467-020-18246-6.

[217] Daniel C. Koboldt, Robert S. Fulton, Michael D. McLellan, Heather Schmidt, Joelle Kalicki-Veizer, Joshua F. McMichael, Lucinda L. Fulton, David J. Dooling, Li Ding, Elaine R. Mardis, Richard K. Wilson, Adrian Ally, Miruna Balasundaram, Yaron S. N. Butterfield, Rebecca Carlsen, Candace Carter, Andy Chu, Eric Chuah, Hye-Jung E. Chun, Robin J. N. Coope, Noreen Dhalla, Ranabir Guin, Carrie Hirst, Martin Hirst, Robert A. Holt, Darlene Lee, Haiyan I. Li, Michael Mayo, Richard A. Moore, Andrew J. Mungall, Erin Pleasance, A. Gordon Robertson, Jacqueline E. Schein, Arash Shafiei, Payal Sipahimalani, Jared R. Slobodan, Dominik Stoll, Angela Tam, Nina Thiessen, Richard J. Varhol, Natasja Wye, Thomas Zeng, Yongjun Zhao, Inanc Birol, Steven J. M. Jones, Marco A. Marra, Andrew D. Cherniack, Gordon Saksena, Robert C. Onofrio, Nam H. Pho, Scott L. Carter, Steven E. Schumacher, Barbara Tabak, Bryan Hernandez, Jeff Gentry, Huy Nguyen, Andrew Crenshaw, Kristin Ardlie, Rameen Beroukhim, Wendy Winckler, Gad Getz, Stacey B. Gabriel, Matthew Meyerson, Lynda Chin, Peter J. Park, Raju Kucherlapati, Katherine A. Hoadley, J. Todd Auman, Cheng Fan, Yidi J. Turman, Yan Shi, Ling Li, Michael D. Topal, Xiaping He, Hann-Hsiang Chao, Aleix Prat, Grace O. Silva, Michael D. Iglesia, Wei Zhao, Jerry Usary, Jonathan S. Berg, Michael Adams, Jessica Booker, Junyuan Wu, Anisha Gulabani, Tom Bodenheimer, Alan P. Hoyle, Janae V. Simons, Matthew G. Soloway, Lisle E. Mose, Stuart R. Jefferys, Saianand Balu, Joel S. Parker, D. Neil Hayes, Charles M. Perou, Simeen Malik, Swapna Mahurkar, Hui Shen, Daniel J. Weisenberger, Timothy Triche Jr, Phillip H. Lai, Moiz S. Bootwalla, Dennis T. Maglinte, Benjamin P. Berman, David J. Van Den Berg, Stephen B. Baylin, Peter W. Laird, Chad J. Creighton, Lawrence A. Done-

hower, Gad Getz, Michael Noble, Doug Voet, Gordon Saksena, Nils Gehlenborg, Daniel DiCara, Juinhua Zhang, Hailei Zhang, Chang-Jiun Wu, Spring Yingchun Liu, Michael S. Lawrence, Lihua Zou, Andrey Sivachenko, Pei Lin, Petar Stojanov, Rui Jing, Juok Cho, Raktim Sinha, Richard W. Park, Marc-Danie Nazaire, Jim Robinson, Helga Thorvaldsdottir, Jill Mesirov, Peter J. Park, Lynda Chin, Sheila Reynolds, Richard B. Kreisberg, Brady Bernard, Ryan Bressler, Timo Erkkila, Jake Lin, Vesteinn Thorsson, Wei Zhang, Ilya Shmulevich, Giovanni Ciriello, Nils Weinhold, Nikolaus Schultz, Jianjiong Gao, Ethan Cerami, Benjamin Gross, Anders Jacobsen, Rileen Sinha, B. Arman Aksoy, Yevgeniy Antipin, Boris Reva, Ronglai Shen, Barry S. Taylor, Marc Ladanyi, Chris Sander, Pavana Anur, Paul T. Spellman, Yiling Lu, Wenbin Liu, Roel R. G. Verhaak, Gordon B. Mills, Rehan Akbani, Nianxiang Zhang, Bradley M. Broom, Tod D. Casasent, Chris Wakefield, Anna K. Unruh, Keith Baggerly, Kevin Coombes, John N. Weinstein, David Haussler, Christopher C. Benz, Joshua M. Stuart, Stephen C. Benz, Jingchun Zhu, Christopher C. Szeto, Gary K. Scott, Christina Yau, Evan O. Paull, Daniel Carlin, Christopher Wong, Artem Sokolov, Janita Thusberg, Sean Mooney, Sam Ng, Theodore C. Goldstein, Kyle Ellrott, Mia Grifford, Christopher Wilks, Singer Ma, Brian Craft, Chunhua Yan, Ying Hu, Daoud Meerzaman, Julie M. Gastier-Foster, Jay Bowen, Nilsa C. Ramirez, Aaron D. Black, Robert E. XPATH ERROR: unknown variable "tname"., Peter White, Erik J. Zmuda, Jessica Frick, Tara M. Lichtenberg, Robin Brookens, Myra M. George, Mark A. Gerken, Hollie A. Harper, Kristen M. Leraas, Lisa J. Wise, Teresa R. Tabler, Cynthia McAllister, Thomas Barr, Melissa Hart-Kothari, Katie Tarvin, Charles Saller, George Sandusky, Colleen Mitchell, Mary V. Iacocca, Jennifer Brown, Brenda Rabeno, Christine Czerwinski, Nicholas Petrelli, Oleg Dolzhansky, Mikhail Abramov, Olga Voronina, Olga Potapova, Jeffrey R. Marks, Wiktoria M. Suchorska, Dawid Murawa, Witold Kycler, Matthew Ibbs, Konstanty Korski, Arkadiusz Spychała, Paweł Murawa, Jacek J. Brzeziński, Hanna Perz, Radosław Łaźniak, Marek Teresiak, Honorata Tatka, Ewa Leporowska, Marta Bogusz-Czerniewicz, Julian Malicki, Andrzej Mackiewicz, Maciej Wiznerowicz, Xuan Van Le, Bernard Kohl, Nguyen Viet Tien, Richard Thorp, Nguyen Van Bang, Howard Sussman, Bui Duc Phu, Richard Hajek, Nguyen Phi Hung, Tran Viet The Phuong, Huynh Quyet Thang, Khurram Zaki Khan, Robert Penny, David Mallery, Erin Curley, Candace Shelton, Peggy Yena, James N. Ingle, Fergus J. Couch, Wilma L. Lingle, Tari A. King, Ana Maria Gonzalez-Angulo, Gordon B. Mills, Mary D. Dyer, Shuying Liu, Xiaolong Meng, Modesto Patangan, Frederic Waldman, Hubert Stöppler, W. Kimryn Rathmell, Leigh Thorne, Mei Huang, Lori Boice, Ashley Hill, Carl Morrison, Carmelo Gaudioso, Wiam Bshara, Kelly Daily, Sophie C. Egea, Mark D. Pegram, Carmen Gomez-Fernandez, Rajiv Dhir, Rohit Bhargava, Adam Brufsky, Craig D. Shriver, Jeffrey A. Hooke, Jamie Leigh Campbell, Richard J. Mural, Hai Hu, Stella Somiari, Caroline Larson, Brenda Deyarmin, Leonid Kvecher, Albert J. Kovatich, Matthew J. Ellis, Tari A. King, Hai Hu, Fergus J. Couch, Richard J. Mural, Thomas Stricker, Kevin White, Olufunmilayo Olopade, James N. Ingle, Chunqing Luo, Yaqin Chen, Jeffrey R. Marks, Fred-

eric Waldman, Maciej Wiznerowicz, Ron Bose, Li-Wei Chang, Andrew H. Beck, Ana Maria Gonzalez-Angulo, Todd Pihl, Mark Jensen, Robert Sfeir, Ari Kahn, Anna Chu, Prachi Kothiyal, Zhining Wang, Eric Snyder, Joan Pontius, Brenda Ayala, Mark Backus, Jessica Walton, Julien Baboud, Dominique Berton, Matthew Nicholls, Deepak Srinivasan, Rohini Raman, Stanley Girshik, Peter Kigonya, Shelley Alonso, Rashmi Sanbhadti, Sean Barletta, David Pot, Margi Sheth, John A. Demchok, Kenna R. Mills Shaw, Liming Yang, Greg Eley, Martin L. Ferguson, Roy W. Tarnuzzer, Jiashan Zhang, Laura A. L. Dillon, Kenneth Buetow, Peter Fielding, Bradley A. Ozenberger, Mark S. Guyer, Heidi J. Sofia, and Jacqueline D. Palchik. Comprehensive molecular portraits of human breast tumours. *Nature*, 490(7418): 61–70, September 2012. ISSN 0028-0836, 1476-4687. doi: 10.1038/nature11412. URL `http://www.nature.com/doifinder/10.1038/nature11412`.

[218] Alexander J. Cole, Trisha Dwight, Anthony J. Gill, Kristie-Ann Dickson, Ying Zhu, Adele Clarkson, Gregory B. Gard, Jayne Maidens, Susan Valmadre, Roderick Clifton-Bligh, and Deborah J. Marsh. Assessing mutant p53 in primary high-grade serous ovarian cancer using immunohistochemistry and massively parallel sequencing. *Scientific Reports*, 6(1):26191, September 2016. ISSN 2045-2322. doi: 10.1038/srep26191. URL `http://www.nature.com/articles/srep26191`.

[219] Kyriaki Michailidou, Sara Lindström, Joe Dennis, Jonathan Beesley, Shirley Hui, Siddhartha Kar, Audrey Lemaçon, Penny Soucy, Dylan Glubb, Asha Rostamianfar, Manjeet K. Bolla, Qin Wang, Jonathan Tyrer, Ed Dicks, Andrew Lee, Zhaoming Wang, Jamie Allen, Renske Keeman, Ursula Eilber, Juliet D. French, Xiao Qing Chen, Laura Fachal, Karen McCue, Amy E. McCart Reed, Maya Ghoussaini, Jason S. Carroll, Xia Jiang, Hilary Finucane, Marcia Adams, Muriel A. Adank, Habibul Ahsan, Kristiina Aittomäki, Hoda Anton-Culver, Natalia N. Antonenkova, Volker Arndt, Kristan J. Aronson, Banu Arun, Paul L. Auer, François Bacot, Myrto Barrdahl, Caroline Baynes, Matthias W. Beckmann, Sabine Behrens, Javier Benitez, Marina Bermisheva, Leslie Bernstein, Carl Blomqvist, Natalia V. Bogdanova, Stig E. Bojesen, Bernardo Bonanni, Anne-Lise Børresen-Dale, Judith S. Brand, Hiltrud Brauch, Paul Brennan, Hermann Brenner, Louise Brinton, Per Broberg, Ian W. Brock, Annegien Broeks, Angela Brooks-Wilson, Sara Y. Brucker, Thomas Brüning, Barbara Burwinkel, Katja Butterbach, Qiuyin Cai, Hui Cai, Trinidad Caldés, Federico Canzian, Angel Carracedo, Brian D. Carter, Jose E. Castelao, Tsun L. Chan, Ting-Yuan David Cheng, Kee Seng Chia, Ji-Yeob Choi, Hans Christiansen, Christine L. Clarke, Margriet Collée, Don M. Conroy, Emilie Cordina-Duverger, Sten Cornelissen, David G. Cox, Angela Cox, Simon S. Cross, Julie M. Cunningham, Kamila Czene, Mary B. Daly, Peter Devilee, Kimberly F. Doheny, Thilo Dörk, Isabel dos Santos-Silva, Martine Dumont, Lorraine Durcan, Miriam Dwek, Diana M. Eccles, Arif B. Ekici, A. Heather Eliassen, Carolina Ellberg, Mingajeva Elvira, Christoph Engel, Mikael Eriksson, Peter A. Fasching, Jonine Figueroa, Dieter Flesch-Janys, Olivia Fletcher, Henrik Flyger, Lin Fritschi, Valerie Gaborieau, Marike Gabrielson, Manuela Gago-Dominguez, Yu-Tang Gao, Susan M. Gapstur, José A. García-Sáenz, Mia M. Gaudet, Vassilios Georgoulias, Graham G. Giles, Gord Glendon,

Mark S. Goldberg, David E. Goldgar, Anna González-Neira, Grethe I. Grenaker Alnæs, Mervi Grip, Jacek Gronwald, Anne Grundy, Pascal Guénel, Lothar Haeberle, Eric Hahnen, Christopher A. Haiman, Niclas Håkansson, Ute Hamann, Nathalie Hamel, Susan Hankinson, Patricia Harrington, Steven N. Hart, Jaana M. Hartikainen, Mikael Hartman, Alexander Hein, Jane Heyworth, Belynda Hicks, Peter Hillemanns, Dona N. Ho, Antoinette Hollestelle, Maartje J. Hooning, Robert N. Hoover, John L. Hopper, Ming-Feng Hou, Chia-Ni Hsiung, Guanmengqian Huang, Keith Humphreys, Junko Ishiguro, Hidemi Ito, Motoki Iwasaki, Hiroji Iwata, Anna Jakubowska, Wolfgang Janni, Esther M. John, Nichola Johnson, Kristine Jones, Michael Jones, Arja Jukkola-Vuorinen, Rudolf Kaaks, Maria Kabisch, Katarzyna Kaczmarek, Daehee Kang, Yoshio Kasuga, Michael J. Kerin, Sofia Khan, Elza Khusnutdinova, Johanna I. Kiiski, Sung-Won Kim, Julia A. Knight, Veli-Matti Kosma, Vessela N. Kristensen, Ute Krüger, Ava Kwong, Diether Lambrechts, Loic Le Marchand, Eunjung Lee, Min Hyuk Lee, Jong Won Lee, Chuen Neng Lee, Flavio Lejbkowicz, Jingmei Li, Jenna Lilyquist, Annika Lindblom, Jolanta Lissowska, Wing-Yee Lo, Sibylle Loibl, Jirong Long, Artitaya Lophatananon, Jan Lubinski, Craig Luccarini, Michael P. Lux, Edmond S. K. Ma, Robert J. MacInnis, Tom Maishman, Enes Makalic, Kathleen E. Malone, Ivana Maleva Kostovska, Arto Mannermaa, Siranoush Manoukian, JoAnn E. Manson, Sara Margolin, Shivaani Mariapun, Maria Elena Martinez, Keitaro Matsuo, Dimitrios Mavroudis, James McKay, Catriona McLean, Hanne Meijers-Heijboer, Alfons Meindl, Primitiva Menéndez, Usha Menon, Jeffery Meyer, Hui Miao, Nicola Miller, Nur Aishah Mohd Taib, Kenneth Muir, Anna Marie Mulligan, Claire Mulot, Susan L. Neuhausen, Heli Nevanlinna, Patrick Neven, Sune F. Nielsen, Dong-Young Noh, Børge G. Nordestgaard, Aaron Norman, Olufunmilayo I. Olopade, Janet E. Olson, Håkan Olsson, Curtis Olswold, Nick Orr, V. Shane Pankratz, Sue K. Park, Tjoung-Won Park-Simon, Rachel Lloyd, Jose I. A. Perez, Paolo Peterlongo, Julian Peto, Kelly-Anne Phillips, Mila Pinchev, Dijana Plaseska-Karanfilska, Ross Prentice, Nadege Presneau, Darya Prokofyeva, Elizabeth Pugh, Katri Pylkäs, Brigitte Rack, Paolo Radice, Nazneen Rahman, Gadi Rennert, Hedy S. Rennert, Valerie Rhenius, Atocha Romero, Jane Romm, Kathryn J. Ruddy, Thomas Rüdiger, Anja Rudolph, Matthias Ruebner, Emiel J. T. Rutgers, Emmanouil Saloustros, Dale P. Sandler, Suleeporn Sangrajrang, Elinor J. Sawyer, Daniel F. Schmidt, Rita K. Schmutzler, Andreas Schneeweiss, Minouk J. Schoemaker, Fredrick Schumacher, Peter Schürmann, Rodney J. Scott, Christopher Scott, Sheila Seal, Caroline Seynaeve, Mitul Shah, Priyanka Sharma, Chen-Yang Shen, Grace Sheng, Mark E. Sherman, Martha J. Shrubsole, Xiao-Ou Shu, Ann Smeets, Christof Sohn, Melissa C. Southey, John J. Spinelli, Christa Stegmaier, Sarah Stewart-Brown, Jennifer Stone, Daniel O. Stram, Harald Surowy, Anthony Swerdlow, Rulla Tamimi, Jack A. Taylor, Maria Tengström, Soo H. Teo, Mary Beth Terry, Daniel C. Tessier, Somchai Thanasitthichai, Kathrin Thöne, Rob A. E. M. Tollenaar, Ian Tomlinson, Ling Tong, Diana Torres, Thérèse Truong, Chiu-Chen Tseng, Shoichiro Tsugane, Hans-Ulrich Ulmer, Giske Ursin, Michael Untch, Celine Vachon, Christi J. van Asperen, David Van Den Berg, Ans M. W. van den

Ouweland, Lizet van der Kolk, Rob B. van der Luijt, Daniel Vincent, Jason Vollenweider, Quinten Waisfisz, Shan Wang-Gohrke, Clarice R. Weinberg, Camilla Wendt, Alice S. Whittemore, Hans Wildiers, Walter Willett, Robert Winqvist, Alicja Wolk, Anna H. Wu, Lucy Xia, Taiki Yamaji, Xiaohong R. Yang, Cheng Har Yip, Keun-Young Yoo, Jyh-Cherng Yu, Wei Zheng, Ying Zheng, Bin Zhu, Argyrios Ziogas, Elad Ziv, Sunil R. Lakhani, Antonis C. Antoniou, Arnaud Droit, Irene L. Andrulis, Christopher I. Amos, Fergus J. Couch, Paul D. P. Pharoah, Jenny Chang-Claude, Per Hall, David J. Hunter, Roger L. Milne, Montserrat García-Closas, Marjanka K. Schmidt, Stephen J. Chanock, Alison M. Dunning, Stacey L. Edwards, Gary D. Bader, Georgia Chenevix-Trench, Jacques Simard, Peter Kraft, and Douglas F. Easton. Association analysis identifies 65 new breast cancer risk loci. *Nature*, 551(7678): 92–94, October 2017. ISSN 0028-0836, 1476-4687. doi: 10.1038/nature24284. URL `http://www.nature.com/doifinder/10.1038/nature24284`.

[220] Catherine M Phelan, Karoline B Kuchenbaecker, Jonathan P Tyrer, Siddhartha P Kar, Kate Lawrenson, Stacey J Winham, Joe Dennis, Ailith Pirie, Marjorie J Riggan, Ganna Chornokur, Madalene A Earp, Paulo C Lyra, Janet M Lee, Simon Coetzee, Jonathan Beesley, Lesley McGuffog, Penny Soucy, Ed Dicks, Andrew Lee, Daniel Barrowdale, Julie Lecarpentier, Goska Leslie, Cora M Aalfs, Katja K H Aben, Marcia Adams, Julian Adlard, Irene L Andrulis, Hoda Anton-Culver, Natalia Antonenkova, Gerasimos Aravantinos, Norbert Arnold, Banu K Arun, Brita Arver, Jacopo Azzollini, Judith Balmaña, Susana N Banerjee, Laure Barjhoux, Rosa B Barkardottir, Yukie Bean, Matthias W Beckmann, Alicia Beeghly-Fadiel, Javier Benitez, Marina Bermisheva, Marcus Q Bernardini, Michael J Birrer, Line Bjorge, Amanda Black, Kenneth Blankstein, Marinus J Blok, Clara Bodelon, Natalia Bogdanova, Anders Bojesen, Bernardo Bonanni, Åke Borg, Angela R Bradbury, James D Brenton, Carole Brewer, Louise Brinton, Per Broberg, Angela Brooks-Wilson, Fiona Bruinsma, Joan Brunet, Bruno Buecher, Ralf Butzow, Saundra S Buys, Trinidad Caldes, Maria A Caligo, Ian Campbell, Rikki Cannioto, Michael E Carney, Terence Cescon, Salina B Chan, Jenny Chang-Claude, Stephen Chanock, Xiao Qing Chen, Yoke-Eng Chiew, Jocelyne Chiquette, Wendy K Chung, Kathleen B M Claes, Thomas Conner, Linda S Cook, Jackie Cook, Daniel W Cramer, Julie M Cunningham, Aimee A D'Aloisio, Mary B Daly, Francesca Damiola, Sakaeva Dina Damirovna, Agnieszka Dansonka-Mieszkowska, Fanny Dao, Rosemarie Davidson, Anna DeFazio, Capucine Delnatte, Kimberly F Doheny, Orland Diez, Yuan Chun Ding, Jennifer Anne Doherty, Susan M Domchek, Cecilia M Dorfling, Thilo Dörk, Laure Dossus, Mercedes Duran, Matthias Dürst, Bernd Dworniczak, Diana Eccles, Todd Edwards, Ros Eeles, Ursula Eilber, Bent Ejlertsen, Arif B Ekici, Steve Ellis, Mingajeva Elvira, Kevin H Eng, Christoph Engel, D Gareth Evans, Peter A Fasching, Sarah Ferguson, Sandra Fert Ferrer, James M Flanagan, Zachary C Fogarty, Renée T Fortner, Florentia Fostira, William D Foulkes, George Fountzilas, Brooke L Fridley, Tara M Friebel, Eitan Friedman, Debra Frost, Patricia A Ganz, Judy Garber, María J García, Vanesa Garcia-Barberan, Andrea Gehrig, Aleksandra Gentry-Maharaj, Anne-Marie Gerdes, Graham G Giles, Rosalind Glasspool,

Gord Glendon, Andrew K Godwin, David E Goldgar, Teodora Goranova, Martin Gore, Mark H Greene, Jacek Gronwald, Stephen Gruber, Eric Hahnen, Christopher A Haiman, Niclas Håkansson, Ute Hamann, Thomas V O Hansen, Patricia A Harrington, Holly R Harris, Jan Hauke, Alexander Hein, Alex Henderson, Michelle A T Hildebrandt, Peter Hillemanns, Shirley Hodgson, Claus K Høgdall, Estrid Høgdall, Frans B L Hogervorst, Helene Holland, Maartje J Hooning, Karen Hosking, Ruea-Yea Huang, Peter J Hulick, Jillian Hung, David J Hunter, David G Huntsman, Tomasz Huzarski, Evgeny N Imyanitov, Claudine Isaacs, Edwin S Iversen, Louise Izatt, Angel Izquierdo, Anna Jakubowska, Paul James, Ramunas Janavicius, Mats Jernetz, Allan Jensen, Uffe Birk Jensen, Esther M John, Sharon Johnatty, Michael E Jones, Päivi Kannisto, Beth Y Karlan, Anthony Karnezis, Karin Kast, Catherine J Kennedy, Elza Khusnutdinova, Lambertus A Kiemeney, Johanna I Kiiski, Sung-Won Kim, Susanne K Kjaer, Martin Köbel, Reidun K Kopperud, Torben A Kruse, Jolanta Kupryjanczyk, Ava Kwong, Yael Laitman, Diether Lambrechts, Nerea Larrañaga, Melissa C Larson, Conxi Lazaro, Nhu D Le, Loic Le Marchand, Jong Won Lee, Shashikant B Lele, Arto Leminen, Dominique Leroux, Jenny Lester, Fabienne Lesueur, Douglas A Levine, Dong Liang, Clemens Liebrich, Jenna Lilyquist, Loren Lipworth, Jolanta Lissowska, Karen H Lu, Jan Lubiński, Craig Luccarini, Lene Lundvall, Phuong L Mai, Gustavo Mendoza-Fandiño, Siranoush Manoukian, Leon F A G Massuger, Taymaa May, Sylvie Mazoyer, Jessica N McAlpine, Valerie McGuire, John R McLaughlin, Iain McNeish, Hanne Meijers-Heijboer, Alfons Meindl, Usha Menon, Arjen R Mensenkamp, Melissa A Merritt, Roger L Milne, Gillian Mitchell, Francesmary Modugno, Joanna Moes-Sosnowska, Melissa Moffitt, Marco Montagna, Kirsten B Moysich, Anna Marie Mulligan, Jacob Musinsky, Katherine L Nathanson, Lotte Nedergaard, Roberta B Ness, Susan L Neuhausen, Heli Nevanlinna, Dieter Niederacher, Robert L Nussbaum, Kunle Odunsi, Edith Olah, Olufunmilayo I Olopade, Håkan Olsson, Curtis Olswold, David M O'Malley, Kai-ren Ong, N Charlotte Onland-Moret, Nicholas Orr, Sandra Orsulic, Ana Osorio, Domenico Palli, Laura Papi, Tjoung-Won Park-Simon, James Paul, Celeste L Pearce, Inge Søkilde Pedersen, Petra H M Peeters, Bernard Peissel, Ana Peixoto, Tanja Pejovic, Liisa M Pelttari, Jennifer B Permuth, Paolo Peterlongo, Lidia Pezzani, Georg Pfeiler, Kelly-Anne Phillips, Marion Piedmonte, Malcolm C Pike, Anna M Piskorz, Samantha R Poblete, Timea Pocza, Elizabeth M Poole, Bruce Poppe, Mary E Porteous, Fabienne Prieur, Darya Prokofyeva, Elizabeth Pugh, Miquel Angel Pujana, Pascal Pujol, Paolo Radice, Johanna Rantala, Christine Rappaport-Fuerhauser, Gad Rennert, Kerstin Rhiem, Patricia Rice, Andrea Richardson, Mark Robson, Gustavo C Rodriguez, Cristina Rodríguez-Antona, Jane Romm, Matti A Rookus, Mary Anne Rossing, Joseph H Rothstein, Anja Rudolph, Ingo B Runnebaum, Helga B Salvesen, Dale P Sandler, Minouk J Schoemaker, Leigha Senter, V Wendy Setiawan, Gianluca Severi, Priyanka Sharma, Tameka Shelford, Nadeem Siddiqui, Lucy E Side, Weiva Sieh, Christian F Singer, Hagay Sobol, Honglin Song, Melissa C Southey, Amanda B Spurdle, Zsofia Stadler, Doris Steinemann, Dominique Stoppa-Lyonnet, Lara E Sucheston-Campbell, Grze-

gorz Sukiennicki, Rebecca Sutphen, Christian Sutter, Anthony J Swerdlow, Csilla I Szabo, Lukasz Szafron, Yen Y Tan, Jack A Taylor, Muy-Kheng Tea, Manuel R Teixeira, Soo-Hwang Teo, Kathryn L Terry, Pamela J Thompson, Liv Cecilie Vestrheim Thomsen, Darcy L Thull, Laima Tihomirova, Anna V Tinker, Marc Tischkowitz, Silvia Tognazzo, Amanda Ewart Toland, Alicia Tone, Britton Trabert, Ruth C Travis, Antonia Trichopoulou, Nadine Tung, Shelley S Tworoger, Anne M van Altena, David Van Den Berg, Annemarie H van der Hout, Rob B van der Luijt, Mattias Van Heetvelde, Els Van Nieuwenhuysen, Elizabeth J van Rensburg, Adriaan Vanderstichele, Raymonda Varon-Mateeva, Ana Vega, Digna Velez Edwards, Ignace Vergote, Robert A Vierkant, Joseph Vijai, Athanassios Vratimos, Lisa Walker, Christine Walsh, Dorothea Wand, Shan Wang-Gohrke, Barbara Wappenschmidt, Penelope M Webb, Clarice R Weinberg, Jeffrey N Weitzel, Nicolas Wentzensen, Alice S Whittemore, Juul T Wijnen, Lynne R Wilkens, Alicja Wolk, Michelle Woo, Xifeng Wu, Anna H Wu, Hannah Yang, Drakoulis Yannoukakos, Argyrios Ziogas, Kristin K Zorn, Steven A Narod, Douglas F Easton, Christopher I Amos, Joellen M Schildkraut, Susan J Ramus, Laura Ottini, Marc T Goodman, Sue K Park, Linda E Kelemen, Harvey A Risch, Mads Thomassen, Kenneth Offit, Jacques Simard, Rita Katharina Schmutzler, Dennis Hazelett, Alvaro N Monteiro, Fergus J Couch, Andrew Berchuck, Georgia Chenevix-Trench, Ellen L Goode, Thomas A Sellers, Simon A Gayther, Antonis C Antoniou, and Paul D P Pharoah. Identification of 12 new susceptibility loci for different histotypes of epithelial ovarian cancer. *Nature Genetics*, 49(5):680–691, March 2017. ISSN 1061-4036, 1546-1718. doi: 10.1038/ng.3826. URL `http://www.nature.com/doifinder/10.1038/ng.3826`.

[221] Jing Gong, Shufang Mei, Chunjie Liu, Yu Xiang, Youqiong Ye, Zhao Zhang, Jing Feng, Renyan Liu, Lixia Diao, An-Yuan Guo, Xiaoping Miao, and Leng Han. PancanQTL: systematic identification of cis-eQTLs and trans-eQTLs in 33 cancer types. *Nucleic Acids Research*, 46(D1):D971–D976, January 2018. ISSN 0305-1048, 1362-4962. doi: 10.1093/nar/gkx861. URL `http://academic.oup.com/nar/article/46/D1/D971/4210944`.

[222] Yong Li, Michael W. Gordon, Zijun Y. Xu-Monette, Carlo Visco, Alexander Tzankov, Dehui Zou, Lugui Qiu, Santiago Montes-Moreno, Karen Dybkaer, Attilio Orazi, Youli Zu, Govind Bhagat, Kristy L. Richards, Eric D. Hsi, William W. L. Choi, J. Han van Krieken, Qin Huang, Weiyun Ai, Maurilio Ponzoni, Andrés J. M. Ferreri, Jane N. Winter, Ronald S. Go, Miguel A. Piris, Michael B. Møller, Lin Wu, Michael Wang, Kenneth S. Ramos, L. Jeffrey Medeiros, and Ken H. Young. Single nucleotide variation in the TP53 3' untranslated region in diffuse large B-cell lymphoma treated with rituximab-CHOP: a report from the International DLBCL Rituximab-CHOP Consortium Program. *Blood*, 121(22):4529–4540, May 2013. ISSN 1528-0020. doi: 10.1182/blood-2012-12-471722.

[223] Simon N. Stacey, Hannes Helgason, Sigurjon A. Gudjonsson, Gudmar Thorleifsson, Florian Zink, Asgeir Sigurdsson, Birte Kehr, Julius Gudmundsson, Patrick Sulem, Bardur Sigurgeirsson, Kristrun R. Benediktsdottir, Kristin Thorisdottir, Rafn Ragnarsson, Victoria Fuentelsaz, Cristina Corredera, Yolanda Gilaberte, Matilde Grasa,

Dolores Planelles, Onofre Sanmartin, Peter Rudnai, Eugene Gurzau, Kvetoslava Koppova, Bjørn A. Nexø, Anne Tjønneland, Kim Overvad, Jon G. Jonasson, Laufey Tryggvadottir, Hrefna Johannsdottir, Anna M. Kristinsdottir, Hreinn Stefansson, Gisli Masson, Olafur T. Magnusson, Bjarni V. Halldorsson, Augustine Kong, Thorunn Rafnar, Unnur Thorsteinsdottir, Ulla Vogel, Rajiv Kumar, Eduardo Nagore, José I. Mayordomo, Daniel F. Gudbjartsson, Jon H. Olafsson, and Kari Stefansson. New basal cell carcinoma susceptibility loci. *Nature Communications*, 6(1), December 2015. ISSN 2041-1723. doi: 10.1038/ncomms7825. URL http://www.nature.com/articles/ncomms7825.

[224] Subhasree Basu and Maureen E. Murphy. Genetic Modifiers of the p53 Pathway. *Cold Spring Harbor Perspectives in Medicine*, 6(4):a026302, April 2016. ISSN 2157-1422. doi: 10.1101/cshperspect.a026302. URL http://perspectivesinmedicine.cshlp.org/lookup/doi/10.1101/cshperspect.a026302.

[225] Gabriel S. Macedo, Igor Araujo Vieira, Ana Paula Brandalize, Juliana Giacomazzi, Edenir Inez Palmero, Sahlua Volc, Vanessa Rodrigues Paixão-Côrtes, Maira Caleffi, Michele Silva Alves, Maria Isabel Achatz, Pierre Hainaut, and Patricia Ashton-Prolla. Rare germline variant (rs78378222) in the TP53 3' UTR: Evidence for a new mechanism of cancer predisposition in Li-Fraumeni syndrome. *Cancer Genetics*, 209(3):97–106, March 2016. ISSN 22107762. doi: 10.1016/j.cancergen.2015.12.012. URL https://linkinghub.elsevier.com/retrieve/pii/S221077621600003X.

[226] Beatrice S Melin, Jill S Barnholtz-Sloan, Margaret R Wrensch, Christoffer Johansen, Dora Il'yasova, Ben Kinnersley, Quinn T Ostrom, Karim Labreche, Yanwen Chen, Georgina Armstrong, Yanhong Liu, Jeanette E Eckel-Passow, Paul A Decker, Marianne Labussière, Ahmed Idbaih, Khe Hoang-Xuan, Anna-Luisa Di Stefano, Karima Mokhtari, Jean-Yves Delattre, Peter Broderick, Pilar Galan, Konstantinos Gousias, Johannes Schramm, Minouk J Schoemaker, Sarah J Fleming, Stefan Herms, Stefanie Heilmann, Markus M Nöthen, Heinz-Erich Wichmann, Stefan Schreiber, Anthony Swerdlow, Mark Lathrop, Matthias Simon, Marc Sanson, Ulrika Andersson, Preetha Rajaraman, Stephen Chanock, Martha Linet, Zhaoming Wang, Meredith Yeager, John K Wiencke, Helen Hansen, Lucie McCoy, Terri Rice, Matthew L Kosel, Hugues Sicotte, Christopher I Amos, Jonine L Bernstein, Faith Davis, Dan Lachance, Ching Lau, Ryan T Merrell, Joellen Shildkraut, Francis Ali-Osman, Siegal Sadetzki, Michael Scheurer, Sanjay Shete, Rose K Lai, Elizabeth B Claus, Sara H Olson, Robert B Jenkins, Richard S Houlston, and Melissa L Bondy. Genome-wide association study of glioma subtypes identifies specific differences in genetic susceptibility to glioblastoma and non-glioblastoma tumors. *Nature Genetics*, 49(5): 789–794, March 2017. ISSN 1061-4036, 1546-1718. doi: 10.1038/ng.3823. URL http://www.nature.com/doifinder/10.1038/ng.3823.

[227] ABCTB Investigators, EMBRACE, GEMO Study Collaborators, HEBON, kConFab/AOCS Investigators, NBSC Collaborators, Roger L Milne, Karoline B Kuchenbaecker, Kyriaki Michailidou, Jonathan Beesley, Siddhartha Kar, Sara Lindström, Shirley Hui, Audrey Lemaçon, Penny Soucy, Joe Dennis, Xia Jiang, Asha Rostamianfar, Hilary Finucane, Manjeet K Bolla, Lesley McGuffog, Qin Wang, Cora M Aalfs,

Marcia Adams, Julian Adlard, Simona Agata, Shahana Ahmed, Habibul Ahsan, Kristiina Aittomäki, Fares Al-Ejeh, Jamie Allen, Christine B Ambrosone, Christopher I Amos, Irene L Andrulis, Hoda Anton-Culver, Natalia N Antonenkova, Volker Arndt, Norbert Arnold, Kristan J Aronson, Bernd Auber, Paul L Auer, Margreet G E M Ausems, Jacopo Azzollini, François Bacot, Judith Balmaña, Monica Barile, Laure Barjhoux, Rosa B Barkardottir, Myrto Barrdahl, Daniel Barnes, Daniel Barrowdale, Caroline Baynes, Matthias W Beckmann, Javier Benitez, Marina Bermisheva, Leslie Bernstein, Yves-Jean Bignon, Kathleen R Blazer, Marinus J Blok, Carl Blomqvist, William Blot, Kristie Bobolis, Bram Boeckx, Natalia V Bogdanova, Anders Bojesen, Stig E Bojesen, Bernardo Bonanni, Anne-Lise Børresen-Dale, Aniko Bozsik, Angela R Bradbury, Judith S Brand, Hiltrud Brauch, Hermann Brenner, Brigitte Bressac-de Paillerets, Carole Brewer, Louise Brinton, Per Broberg, Angela Brooks-Wilson, Joan Brunet, Thomas Brüning, Barbara Burwinkel, Saundra S Buys, Jinyoung Byun, Qiuyin Cai, Trinidad Caldés, Maria A Caligo, Ian Campbell, Federico Canzian, Olivier Caron, Angel Carracedo, Brian D Carter, J Esteban Castelao, Laurent Castera, Virginie Caux-Moncoutier, Salina B Chan, Jenny Chang-Claude, Stephen J Chanock, Xiaoqing Chen, Ting-Yuan David Cheng, Jocelyne Chiquette, Hans Christiansen, Kathleen B M Claes, Christine L Clarke, Thomas Conner, Don M Conroy, Jackie Cook, Emilie Cordina-Duverger, Sten Cornelissen, Isabelle Coupier, Angela Cox, David G Cox, Simon S Cross, Katarina Cuk, Julie M Cunningham, Kamila Czene, Mary B Daly, Francesca Damiola, Hatef Darabi, Rosemarie Davidson, Kim De Leeneer, Peter Devilee, Ed Dicks, Orland Diez, Yuan Chun Ding, Nina Ditsch, Kimberly F Doheny, Susan M Domchek, Cecilia M Dorfling, Thilo Dörk, Isabel dos Santos-Silva, Stéphane Dubois, Pierre-Antoine Dugué, Martine Dumont, Alison M Dunning, Lorraine Durcan, Miriam Dwek, Bernd Dworniczak, Diana Eccles, Ros Eeles, Hans Ehrencrona, Ursula Eilber, Bent Ejlertsen, Arif B Ekici, A Heather Eliassen, Christoph Engel, Mikael Eriksson, Laura Fachal, Laurence Faivre, Peter A Fasching, Ulrike Faust, Jonine Figueroa, Dieter Flesch-Janys, Olivia Fletcher, Henrik Flyger, William D Foulkes, Eitan Friedman, Lin Fritschi, Debra Frost, Marike Gabrielson, Pragna Gaddam, Marilie D Gammon, Patricia A Ganz, Susan M Gapstur, Judy Garber, Vanesa Garcia-Barberan, José A García-Sáenz, Mia M Gaudet, Marion Gauthier-Villars, Andrea Gehrig, Vassilios Georgoulias, Anne-Marie Gerdes, Graham G Giles, Gord Glendon, Andrew K Godwin, Mark S Goldberg, David E Goldgar, Anna González-Neira, Paul Goodfellow, Mark H Greene, Grethe I Grenaker Alnæs, Mervi Grip, Jacek Gronwald, Anne Grundy, Daphne Gschwantler-Kaulich, Pascal Guénel, Qi Guo, Lothar Haeberle, Eric Hahnen, Christopher A Haiman, Niclas Håkansson, Emily Hallberg, Ute Hamann, Nathalie Hamel, Susan Hankinson, Thomas V O Hansen, Patricia Harrington, Steven N Hart, Jaana M Hartikainen, Catherine S Healey, Alexander Hein, Sonja Helbig, Alex Henderson, Jane Heyworth, Belynda Hicks, Peter Hillemanns, Shirley Hodgson, Frans B Hogervorst, Antoinette Hollestelle, Maartje J Hooning, Bob Hoover, John L Hopper, Chunling Hu, Guanmengqian Huang, Peter J Hulick, Keith Humphreys, David J Hunter, Evgeny N Imyanitov, Claudine Isaacs, Motoki

Iwasaki, Louise Izatt, Anna Jakubowska, Paul James, Ramunas Janavicius, Wolfgang Janni, Uffe Birk Jensen, Esther M John, Nichola Johnson, Kristine Jones, Michael Jones, Arja Jukkola-Vuorinen, Rudolf Kaaks, Maria Kabisch, Katarzyna Kaczmarek, Daehee Kang, Karin Kast, Renske Keeman, Michael J Kerin, Carolien M Kets, Machteld Keupers, Sofia Khan, Elza Khusnutdinova, Johanna I Kiiski, Sung-Won Kim, Julia A Knight, Irene Konstantopoulou, Veli-Matti Kosma, Vessela N Kristensen, Torben A Kruse, Ava Kwong, Anne-Vibeke Lænkholm, Yael Laitman, Fiona Lalloo, Diether Lambrechts, Keren Landsman, Christine Lasset, Conxi Lazaro, Loic Le Marchand, Julie Lecarpentier, Andrew Lee, Eunjung Lee, Jong Won Lee, Min Hyuk Lee, Flavio Lejbkowicz, Fabienne Lesueur, Jingmei Li, Jenna Lilyquist, Anne Lincoln, Annika Lindblom, Jolanta Lissowska, Wing-Yee Lo, Sibylle Loibl, Jirong Long, Jennifer T Loud, Jan Lubinski, Craig Luccarini, Michael Lush, Robert J MacInnis, Tom Maishman, Enes Makalic, Ivana Maleva Kostovska, Kathleen E Malone, Siranoush Manoukian, JoAnn E Manson, Sara Margolin, John W M Martens, Maria Elena Martinez, Keitaro Matsuo, Dimitrios Mavroudis, Sylvie Mazoyer, Catriona McLean, Hanne Meijers-Heijboer, Primitiva Menéndez, Jeffery Meyer, Hui Miao, Austin Miller, Nicola Miller, Gillian Mitchell, Marco Montagna, Kenneth Muir, Anna Marie Mulligan, Claire Mulot, Sue Nadesan, Katherine L Nathanson, Susan L Neuhausen, Heli Nevanlinna, Ines Nevelsteen, Dieter Niederacher, Sune F Nielsen, Børge G Nordestgaard, Aaron Norman, Robert L Nussbaum, Edith Olah, Olufunmilayo I Olopade, Janet E Olson, Curtis Olswold, Kai-ren Ong, Jan C Oosterwijk, Nick Orr, Ana Osorio, V Shane Pankratz, Laura Papi, Tjoung-Won Park-Simon, Ylva Paulsson-Karlsson, Rachel Lloyd, Inge Søkilde Pedersen, Bernard Peissel, Ana Peixoto, Jose I A Perez, Paolo Peterlongo, Julian Peto, Georg Pfeiler, Catherine M Phelan, Mila Pinchev, Dijana Plaseska-Karanfilska, Bruce Poppe, Mary E Porteous, Ross Prentice, Nadege Presneau, Darya Prokofieva, Elizabeth Pugh, Miquel Angel Pujana, Katri Pylkäs, Brigitte Rack, Paolo Radice, Nazneen Rahman, Johanna Rantala, Christine Rappaport-Fuerhauser, Gad Rennert, Hedy S Rennert, Valerie Rhenius, Kerstin Rhiem, Andrea Richardson, Gustavo C Rodriguez, Atocha Romero, Jane Romm, Matti A Rookus, Anja Rudolph, Thomas Ruediger, Emmanouil Saloustros, Joyce Sanders, Dale P Sandler, Suleeporn Sangrajrang, Elinor J Sawyer, Daniel F Schmidt, Minouk J Schoemaker, Fredrick Schumacher, Peter Schürmann, Lukas Schwentner, Christopher Scott, Rodney J Scott, Sheila Seal, Leigha Senter, Caroline Seynaeve, Mitul Shah, Priyanka Sharma, Chen-Yang Shen, Xin Sheng, Hermela Shimelis, Martha J Shrubsole, Xiao-Ou Shu, Lucy E Side, Christian F Singer, Christof Sohn, Melissa C Southey, John J Spinelli, Amanda B Spurdle, Christa Stegmaier, Dominique Stoppa-Lyonnet, Grzegorz Sukiennicki, Harald Surowy, Christian Sutter, Anthony Swerdlow, Csilla I Szabo, Rulla M Tamimi, Yen Y Tan, Jack A Taylor, Maria-Isabel Tejada, Maria Tengström, Soo H Teo, Mary B Terry, Daniel C Tessier, Alex Teulé, Kathrin Thöne, Darcy L Thull, Maria Grazia Tibiletti, Laima Tihomirova, Marc Tischkowitz, Amanda E Toland, Rob A E M Tollenaar, Ian Tomlinson, Ling Tong, Diana Torres, Martine Tranchant, Thérèse Truong, Kathy Tucker, Nadine Tung, Jonathan Tyrer,

Hans-Ulrich Ulmer, Celine Vachon, Christi J van Asperen, David Van Den Berg, Ans M W van den Ouweland, Elizabeth J van Rensburg, Liliana Varesco, Raymonda Varon-Mateeva, Ana Vega, Alessandra Viel, Joseph Vijai, Daniel Vincent, Jason Vollenweider, Lisa Walker, Zhaoming Wang, Shan Wang-Gohrke, Barbara Wappenschmidt, Clarice R Weinberg, Jeffrey N Weitzel, Camilla Wendt, Jelle Wesseling, Alice S Whittemore, Juul T Wijnen, Walter Willett, Robert Winqvist, Alicja Wolk, Anna H Wu, Lucy Xia, Xiaohong R Yang, Drakoulis Yannoukakos, Daniela Zaffaroni, Wei Zheng, Bin Zhu, Argyrios Ziogas, Elad Ziv, Kristin K Zorn, Manuela Gago-Dominguez, Arto Mannermaa, Håkan Olsson, Manuel R Teixeira, Jennifer Stone, Kenneth Offit, Laura Ottini, Sue K Park, Mads Thomassen, Per Hall, Alfons Meindl, Rita K Schmutzler, Arnaud Droit, Gary D Bader, Paul D P Pharoah, Fergus J Couch, Douglas F Easton, Peter Kraft, Georgia Chenevix-Trench, Montserrat García-Closas, Marjanka K Schmidt, Antonis C Antoniou, and Jacques Simard. Identification of ten variants associated with risk of estrogen-receptor-negative breast cancer. *Nature Genetics*, 49(12):1767–1778, December 2017. ISSN 1061-4036, 1546-1718. doi: 10.1038/ng.3785. URL `http://www.nature.com/articles/ng.3785`.

[228] Thorunn Rafnar, Bjarni Gunnarsson, Olafur A. Stefansson, Patrick Sulem, Andres Ingason, Michael L. Frigge, Lilja Stefansdottir, Jon K. Sigurdsson, Vinicius Tragante, Valgerdur Steinthorsdottir, Unnur Styrkarsdottir, Simon N. Stacey, Julius Gudmundsson, Gudny A. Arnadottir, Asmundur Oddsson, Florian Zink, Gisli Halldorsson, Gardar Sveinbjornsson, Ragnar P. Kristjansson, Olafur B. Davidsson, Anna Salvarsdottir, Asgeir Thoroddsen, Elisabet A. Helgadottir, Katrin Kristjansdottir, Orri Ingthorsson, Valur Gudmundsson, Reynir T. Geirsson, Ragnheidur Arnadottir, Daniel F. Gudbjartsson, Gisli Masson, Folkert W. Asselbergs, Jon G. Jonasson, Karl Olafsson, Unnur Thorsteinsdottir, Bjarni V. Halldorsson, Gudmar Thorleifsson, and Kari Stefansson. Variants associating with uterine leiomyoma highlight genetic background shared by various cancers and hormone-related traits. *Nature Communications*, 9(1):3636, September 2018. ISSN 2041-1723. doi: 10.1038/s41467-018-05428-6.

[229] Qipan Deng, Hui Hu, Xinfang Yu, Shuanglin Liu, Lei Wang, Weiqun Chen, Chi Zhang, Zhaoyang Zeng, Ya Cao, Zijun Y. Xu-Monette, Ling Li, Mingzhi Zhang, Steven Rosenfeld, Shideng Bao, Eric Hsi, Ken H. Young, Zhongxin Lu, and Yong Li. Tissue-specific microRNA expression alters cancer susceptibility conferred by a TP53 noncoding variant. *Nature Communications*, 10(1):5061, December 2019. ISSN 2041-1723. doi: 10.1038/s41467-019-13002-x. URL `http://www.nature.com/articles/s41467-019-13002-x`.

[230] Maria Teresa Landi, D. Timothy Bishop, Stuart MacGregor, Mitchell J. Machiela, Alexander J. Stratigos, Paola Ghiorzo, Myriam Brossard, Donato Calista, Jiyeon Choi, Maria Concetta Fargnoli, Tongwu Zhang, Monica Rodolfo, Adam J. Trower, Chiara Menin, Jacobo Martinez, Andreas Hadjisavvas, Lei Song, Irene Stefanaki, Richard Scolyer, Rose Yang, Alisa M. Goldstein, Miriam Potrony, Katerina P. Kypreou, Lorenza Pastorino, Paola Queirolo, Cristina Pellegrini, Laura Cattaneo, Matthew Zawistowski, Pol Gimenez-Xavier, Arantxa Rodriguez, Lisa Elefanti, Sir-

anoush Manoukian, Licia Rivoltini, Blair H. Smith, Maria A. Loizidou, Laura Del Regno, Daniela Massi, Mario Mandala, Kiarash Khosrotehrani, Lars A. Akslen, Christopher I. Amos, Per A. Andresen, Marie-Françoise Avril, Esther Azizi, H. Peter Soyer, Veronique Bataille, Bruna Dalmasso, Lisa M. Bowdler, Kathryn P. Burdon, Wei V. Chen, Veryan Codd, Jamie E. Craig, Tadeusz Debniak, Mario Falchi, Shenying Fang, Eitan Friedman, Sarah Simi, Pilar Galan, Zaida Garcia-Casado, Elizabeth M. Gillanders, Scott Gordon, Adele Green, Nelleke A. Gruis, Johan Hansson, Mark Harland, Jessica Harris, Per Helsing, Anjali Henders, Marko Hočevar, Veronica Höiom, David Hunter, Christian Ingvar, Rajiv Kumar, Julie Lang, G. Mark Lathrop, Jeffrey E. Lee, Xin Li, Jan Lubiński, Rona M. Mackie, Maryrose Malt, Josep Malvehy, Kerrie McAloney, Hamida Mohamdi, Anders Molven, Eric K. Moses, Rachel E. Neale, Srdjan Novaković, Dale R. Nyholt, Håkan Olsson, Nicholas Orr, Lars G. Fritsche, Joan Anton Puig-Butille, Abrar A. Qureshi, Graham L. Radford-Smith, Juliette Randerson-Moor, Celia Requena, Casey Rowe, Nilesh J. Samani, Marianna Sanna, Dirk Schadendorf, Hans-Joachim Schulze, Lisa A. Simms, Mark Smithers, Fengju Song, Anthony J. Swerdlow, Nienke van der Stoep, Nicole A. Kukutsch, Alessia Visconti, Leanne Wallace, Sarah V. Ward, Lawrie Wheeler, Richard A. Sturm, Amy Hutchinson, Kristine Jones, Michael Malasky, Aurelie Vogt, Weiyin Zhou, Karen A. Pooley, David E. Elder, Jiali Han, Belynda Hicks, Nicholas K. Hayward, Peter A. Kanetsky, Chad Brummett, Grant W. Montgomery, Catherine M. Olsen, Caroline Hayward, Alison M. Dunning, Nicholas G. Martin, Evangelos Evangelou, Graham J. Mann, Georgina Long, Paul D. P. Pharoah, Douglas F. Easton, Jennifer H. Barrett, Anne E. Cust, Goncalo Abecasis, David L. Duffy, David C. Whiteman, Helen Gogas, Arcangela De Nicolo, Margaret A. Tucker, Julia A. Newton-Bishop, GenoMEL Consortium, Q-MEGA and QTWIN Investigators, ATHENS Melanoma Study Group, 23andMe, SDH Study Group, IBD Investigators, Essen-Heidelberg Investigators, AMFS Investigators, MelaNostrum Consortium, Ketty Peris, Stephen J. Chanock, Florence Demenais, Kevin M. Brown, Susana Puig, Eduardo Nagore, Jianxin Shi, Mark M. Iles, and Matthew H. Law. Genome-wide association meta-analyses combining multiple risk phenotypes provide insights into the genetic architecture of cutaneous melanoma susceptibility. *Nature Genetics*, 52(5):494–504, May 2020. ISSN 1546-1718. doi: 10.1038/s41588-020-0611-8.

[231] Roger L. Milne, Karoline B. Kuchenbaecker, Kyriaki Michailidou, Jonathan Beesley, Siddhartha Kar, Sara Lindström, Shirley Hui, Audrey Lemaçon, Penny Soucy, Joe Dennis, Xia Jiang, Asha Rostamianfar, Hilary Finucane, Manjeet K. Bolla, Lesley McGuffog, Qin Wang, Cora M. Aalfs, ABCTB Investigators, Marcia Adams, Julian Adlard, Simona Agata, Shahana Ahmed, Habibul Ahsan, Kristiina Aittomäki, Fares Al-Ejeh, Jamie Allen, Christine B. Ambrosone, Christopher I. Amos, Irene L. Andrulis, Hoda Anton-Culver, Natalia N. Antonenkova, Volker Arndt, Norbert Arnold, Kristan J. Aronson, Bernd Auber, Paul L. Auer, Margreet G. E. M. Ausems, Jacopo Azzollini, François Bacot, Judith Balmaña, Monica Barile, Laure Barjhoux, Rosa B. Barkardottir, Myrto Barrdahl, Daniel Barnes, Daniel Barrowdale, Caroline Baynes, Matthias W. Beckmann, Javier Benitez, Marina Bermisheva, Leslie

Bernstein, Yves-Jean Bignon, Kathleen R. Blazer, Marinus J. Blok, Carl Blomqvist, William Blot, Kristie Bobolis, Bram Boeckx, Natalia V. Bogdanova, Anders Bojesen, Stig E. Bojesen, Bernardo Bonanni, Anne-Lise Børresen-Dale, Aniko Bozsik, Angela R. Bradbury, Judith S. Brand, Hiltrud Brauch, Hermann Brenner, Brigitte Bressac-de Paillerets, Carole Brewer, Louise Brinton, Per Broberg, Angela Brooks-Wilson, Joan Brunet, Thomas Brüning, Barbara Burwinkel, Saundra S. Buys, Jinyoung Byun, Qiuyin Cai, Trinidad Caldés, Maria A. Caligo, Ian Campbell, Federico Canzian, Olivier Caron, Angel Carracedo, Brian D. Carter, J. Esteban Castelao, Laurent Castera, Virginie Caux-Moncoutier, Salina B. Chan, Jenny Chang-Claude, Stephen J. Chanock, Xiaoqing Chen, Ting-Yuan David Cheng, Jocelyne Chiquette, Hans Christiansen, Kathleen B. M. Claes, Christine L. Clarke, Thomas Conner, Don M. Conroy, Jackie Cook, Emilie Cordina-Duverger, Sten Cornelissen, Isabelle Coupier, Angela Cox, David G. Cox, Simon S. Cross, Katarina Cuk, Julie M. Cunningham, Kamila Czene, Mary B. Daly, Francesca Damiola, Hatef Darabi, Rosemarie Davidson, Kim De Leeneer, Peter Devilee, Ed Dicks, Orland Diez, Yuan Chun Ding, Nina Ditsch, Kimberly F. Doheny, Susan M. Domchek, Cecilia M. Dorfling, Thilo Dörk, Isabel Dos-Santos-Silva, Stéphane Dubois, Pierre-Antoine Dugué, Martine Dumont, Alison M. Dunning, Lorraine Durcan, Miriam Dwek, Bernd Dworniczak, Diana Eccles, Ros Eeles, Hans Ehrencrona, Ursula Eilber, Bent Ejlertsen, Arif B. Ekici, A. Heather Eliassen, EMBRACE, Christoph Engel, Mikael Eriksson, Laura Fachal, Laurence Faivre, Peter A. Fasching, Ulrike Faust, Jonine Figueroa, Dieter Flesch-Janys, Olivia Fletcher, Henrik Flyger, William D. Foulkes, Eitan Friedman, Lin Fritschi, Debra Frost, Marike Gabrielson, Pragna Gaddam, Marilie D. Gammon, Patricia A. Ganz, Susan M. Gapstur, Judy Garber, Vanesa Garcia-Barberan, José A. García-Sáenz, Mia M. Gaudet, Marion Gauthier-Villars, Andrea Gehrig, GEMO Study Collaborators, Vassilios Georgoulias, Anne-Marie Gerdes, Graham G. Giles, Gord Glendon, Andrew K. Godwin, Mark S. Goldberg, David E. Goldgar, Anna González-Neira, Paul Goodfellow, Mark H. Greene, Grethe I. Grenaker Alnæs, Mervi Grip, Jacek Gronwald, Anne Grundy, Daphne Gschwantler-Kaulich, Pascal Guénel, Qi Guo, Lothar Haeberle, Eric Hahnen, Christopher A. Haiman, Niclas Håkansson, Emily Hallberg, Ute Hamann, Nathalie Hamel, Susan Hankinson, Thomas V. O. Hansen, Patricia Harrington, Steven N. Hart, Jaana M. Hartikainen, Catherine S. Healey, HEBON, Alexander Hein, Sonja Helbig, Alex Henderson, Jane Heyworth, Belynda Hicks, Peter Hillemanns, Shirley Hodgson, Frans B. Hogervorst, Antoinette Hollestelle, Maartje J. Hooning, Bob Hoover, John L. Hopper, Chunling Hu, Guanmengqian Huang, Peter J. Hulick, Keith Humphreys, David J. Hunter, Evgeny N. Imyanitov, Claudine Isaacs, Motoki Iwasaki, Louise Izatt, Anna Jakubowska, Paul James, Ramunas Janavicius, Wolfgang Janni, Uffe Birk Jensen, Esther M. John, Nichola Johnson, Kristine Jones, Michael Jones, Arja Jukkola-Vuorinen, Rudolf Kaaks, Maria Kabisch, Katarzyna Kaczmarek, Daehee Kang, Karin Kast, kConFab/AOCS Investigators, Renske Keeman, Michael J. Kerin, Carolien M. Kets, Machteld Keupers, Sofia Khan, Elza Khusnutdinova, Johanna I. Kiiski, Sung-Won Kim, Julia A. Knight, Irene Konstantopoulou, Veli-Matti Kosma,

Vessela N. Kristensen, Torben A. Kruse, Ava Kwong, Anne-Vibeke Lænkholm, Yael Laitman, Fiona Lalloo, Diether Lambrechts, Keren Landsman, Christine Lasset, Conxi Lazaro, Loic Le Marchand, Julie Lecarpentier, Andrew Lee, Eunjung Lee, Jong Won Lee, Min Hyuk Lee, Flavio Lejbkowicz, Fabienne Lesueur, Jingmei Li, Jenna Lilyquist, Anne Lincoln, Annika Lindblom, Jolanta Lissowska, Wing-Yee Lo, Sibylle Loibl, Jirong Long, Jennifer T. Loud, Jan Lubinski, Craig Luccarini, Michael Lush, Robert J. MacInnis, Tom Maishman, Enes Makalic, Ivana Maleva Kostovska, Kathleen E. Malone, Siranoush Manoukian, JoAnn E. Manson, Sara Margolin, John W. M. Martens, Maria Elena Martinez, Keitaro Matsuo, Dimitrios Mavroudis, Sylvie Mazoyer, Catriona McLean, Hanne Meijers-Heijboer, Primitiva Menéndez, Jeffery Meyer, Hui Miao, Austin Miller, Nicola Miller, Gillian Mitchell, Marco Montagna, Kenneth Muir, Anna Marie Mulligan, Claire Mulot, Sue Nadesan, Katherine L. Nathanson, NBSC Collaborators, Susan L. Neuhausen, Heli Nevanlinna, Ines Nevelsteen, Dieter Niederacher, Sune F. Nielsen, Børge G. Nordestgaard, Aaron Norman, Robert L. Nussbaum, Edith Olah, Olufunmilayo I. Olopade, Janet E. Olson, Curtis Olswold, Kai-Ren Ong, Jan C. Oosterwijk, Nick Orr, Ana Osorio, V. Shane Pankratz, Laura Papi, Tjoung-Won Park-Simon, Ylva Paulsson-Karlsson, Rachel Lloyd, Inge Søkilde Pedersen, Bernard Peissel, Ana Peixoto, Jose I. A. Perez, Paolo Peterlongo, Julian Peto, Georg Pfeiler, Catherine M. Phelan, Mila Pinchev, Dijana Plaseska-Karanfilska, Bruce Poppe, Mary E. Porteous, Ross Prentice, Nadege Presneau, Darya Prokofieva, Elizabeth Pugh, Miquel Angel Pujana, Katri Pylkäs, Brigitte Rack, Paolo Radice, Nazneen Rahman, Johanna Rantala, Christine Rappaport-Fuerhauser, Gad Rennert, Hedy S. Rennert, Valerie Rhenius, Kerstin Rhiem, Andrea Richardson, Gustavo C. Rodriguez, Atocha Romero, Jane Romm, Matti A. Rookus, Anja Rudolph, Thomas Ruediger, Emmanouil Saloustros, Joyce Sanders, Dale P. Sandler, Suleeporn Sangrajrang, Elinor J. Sawyer, Daniel F. Schmidt, Minouk J. Schoemaker, Fredrick Schumacher, Peter Schürmann, Lukas Schwentner, Christopher Scott, Rodney J. Scott, Sheila Seal, Leigha Senter, Caroline Seynaeve, Mitul Shah, Priyanka Sharma, Chen-Yang Shen, Xin Sheng, Hermela Shimelis, Martha J. Shrubsole, Xiao-Ou Shu, Lucy E. Side, Christian F. Singer, Christof Sohn, Melissa C. Southey, John J. Spinelli, Amanda B. Spurdle, Christa Stegmaier, Dominique Stoppa-Lyonnet, Grzegorz Sukiennicki, Harald Surowy, Christian Sutter, Anthony Swerdlow, Csilla I. Szabo, Rulla M. Tamimi, Yen Y. Tan, Jack A. Taylor, Maria-Isabel Tejada, Maria Tengström, Soo H. Teo, Mary B. Terry, Daniel C. Tessier, Alex Teulé, Kathrin Thöne, Darcy L. Thull, Maria Grazia Tibiletti, Laima Tihomirova, Marc Tischkowitz, Amanda E. Toland, Rob A. E. M. Tollenaar, Ian Tomlinson, Ling Tong, Diana Torres, Martine Tranchant, Thérèse Truong, Kathy Tucker, Nadine Tung, Jonathan Tyrer, Hans-Ulrich Ulmer, Celine Vachon, Christi J. van Asperen, David Van Den Berg, Ans M. W. van den Ouweland, Elizabeth J. van Rensburg, Liliana Varesco, Raymonda Varon-Mateeva, Ana Vega, Alessandra Viel, Joseph Vijai, Daniel Vincent, Jason Vollenweider, Lisa Walker, Zhaoming Wang, Shan Wang-Gohrke, Barbara Wappenschmidt, Clarice R. Weinberg, Jeffrey N. Weitzel, Camilla Wendt, Jelle Wesseling,

Alice S. Whittemore, Juul T. Wijnen, Walter Willett, Robert Winqvist, Alicja Wolk, Anna H. Wu, Lucy Xia, Xiaohong R. Yang, Drakoulis Yannoukakos, Daniela Zaffaroni, Wei Zheng, Bin Zhu, Argyrios Ziogas, Elad Ziv, Kristin K. Zorn, Manuela Gago-Dominguez, Arto Mannermaa, Håkan Olsson, Manuel R. Teixeira, Jennifer Stone, Kenneth Offit, Laura Ottini, Sue K. Park, Mads Thomassen, Per Hall, Alfons Meindl, Rita K. Schmutzler, Arnaud Droit, Gary D. Bader, Paul D. P. Pharoah, Fergus J. Couch, Douglas F. Easton, Peter Kraft, Georgia Chenevix-Trench, Montserrat García-Closas, Marjanka K. Schmidt, Antonis C. Antoniou, and Jacques Simard. Identification of ten variants associated with risk of estrogen-receptor-negative breast cancer. *Nature Genetics*, 49(12):1767–1778, December 2017. ISSN 1546-1718. doi: 10.1038/ng.3785.

[232] Kristen S. Purrington, Susan Slager, Diana Eccles, Drakoulis Yannoukakos, Peter A. Fasching, Penelope Miron, Jane Carpenter, Jenny Chang-Claude, Nicholas G. Martin, Grant W. Montgomery, Vessela Kristensen, Hoda Anton-Culver, Paul Goodfellow, William J. Tapper, Sajjad Rafiq, Susan M. Gerty, Lorraine Durcan, Irene Konstantopoulou, Florentia Fostira, Athanassios Vratimos, Paraskevi Apostolou, Irene Konstanta, Vassiliki Kotoula, Sotiris Lakis, Meletios A. Dimopoulos, Dimosthenis Skarlos, Dimitrios Pectasides, George Fountzilas, Matthias W. Beckmann, Alexander Hein, Matthias Ruebner, Arif B. Ekici, Arndt Hartmann, Ruediger Schulz-Wendtland, Stefan P. Renner, Wolfgang Janni, Brigitte Rack, Christoph Scholz, Julia Neugebauer, Ulrich Andergassen, Michael P. Lux, Lothar Haeberle, Christine Clarke, Nirmala Pathmanathan, Anja Rudolph, Dieter Flesch-Janys, Stefan Nickels, Janet E. Olson, James N. Ingle, Curtis Olswold, Seth Slettedahl, Jeanette E. Eckel-Passow, S. Keith Anderson, Daniel W. Visscher, Victoria L. Cafourek, Hugues Sicotte, Naresh Prodduturi, Elisabete Weiderpass, Leslie Bernstein, Argyrios Ziogas, Jennifer Ivanovich, Graham G. Giles, Laura Baglietto, Melissa Southey, Veli-Matti Kosma, Hans-Peter Fischer, GENICA Network, Malcom W. R. Reed, Simon S. Cross, Sandra Deming-Halverson, Martha Shrubsole, Qiuyin Cai, Xiao-Ou Shu, Mary Daly, Joellen Weaver, Eric Ross, Jennifer Klemp, Priyanka Sharma, Diana Torres, Thomas Rüdiger, Heidrun Wölfing, Hans-Ulrich Ulmer, Asta Försti, Thaer Khoury, Shicha Kumar, Robert Pilarski, Charles L. Shapiro, Dario Greco, Päivi Heikkilä, Kristiina Aittomäki, Carl Blomqvist, Astrid Irwanto, Jianjun Liu, Vernon Shane Pankratz, Xianshu Wang, Gianluca Severi, Arto Mannermaa, Douglas Easton, Per Hall, Hiltrud Brauch, Angela Cox, Wei Zheng, Andrew K. Godwin, Ute Hamann, Christine Ambrosone, Amanda Ewart Toland, Heli Nevanlinna, Celine M. Vachon, and Fergus J. Couch. Genome-wide association study identifies 25 known breast cancer susceptibility loci as risk factors for triple-negative breast cancer. *Carcinogenesis*, 35(5):1012–1019, May 2014. ISSN 1460-2180. doi: 10.1093/carcin/bgt404.

[233] Fredrick R. Schumacher, Ali Amin Al Olama, Sonja I. Berndt, Sara Benlloch, Mahbubl Ahmed, Edward J. Saunders, Tokhir Dadaev, Daniel Leongamornlert, Ezequiel Anokian, Clara Cieza-Borrella, Chee Goh, Mark N. Brook, Xin Sheng, Laura Fachal, Joe Dennis, Jonathan Tyrer, Kenneth Muir, Artitaya Lophatananon, Victo-

ria L. Stevens, Susan M. Gapstur, Brian D. Carter, Catherine M. Tangen, Phyllis J. Goodman, Ian M. Thompson, Jyotsna Batra, Suzanne Chambers, Leire Moya, Judith Clements, Lisa Horvath, Wayne Tilley, Gail P. Risbridger, Henrik Gronberg, Markus Aly, Tobias Nordström, Paul Pharoah, Nora Pashayan, Johanna Schleutker, Teuvo L. J. Tammela, Csilla Sipeky, Anssi Auvinen, Demetrius Albanes, Stephanie Weinstein, Alicja Wolk, Niclas Håkansson, Catharine M. L. West, Alison M. Dunning, Neil Burnet, Lorelei A. Mucci, Edward Giovannucci, Gerald L. Andriole, Olivier Cussenot, Géraldine Cancel-Tassin, Stella Koutros, Laura E. Beane Freeman, Karina Dalsgaard Sorensen, Torben Falck Orntoft, Michael Borre, Lovise Maehle, Eli Marie Grindedal, David E. Neal, Jenny L. Donovan, Freddie C. Hamdy, Richard M. Martin, Ruth C. Travis, Tim J. Key, Robert J. Hamilton, Neil E. Fleshner, Antonio Finelli, Sue Ann Ingles, Mariana C. Stern, Barry S. Rosenstein, Sarah L. Kerns, Harry Ostrer, Yong-Jie Lu, Hong-Wei Zhang, Ninghan Feng, Xueying Mao, Xin Guo, Guomin Wang, Zan Sun, Graham G. Giles, Melissa C. Southey, Robert J. MacInnis, Liesel M. FitzGerald, Adam S. Kibel, Bettina F. Drake, Ana Vega, Antonio Gómez-Caamaño, Robert Szulkin, Martin Eklund, Manolis Kogevinas, Javier Llorca, Gemma Castaño-Vinyals, Kathryn L. Penney, Meir Stampfer, Jong Y. Park, Thomas A. Sellers, Hui-Yi Lin, Janet L. Stanford, Cezary Cybulski, Dominika Wokolorczyk, Jan Lubinski, Elaine A. Ostrander, Milan S. Geybels, Børge G. Nordestgaard, Sune F. Nielsen, Maren Weischer, Rasmus Bisbjerg, Martin Andreas Røder, Peter Iversen, Hermann Brenner, Katarina Cuk, Bernd Holleczek, Christiane Maier, Manuel Luedeke, Thomas Schnoeller, Jeri Kim, Christopher J. Logothetis, Esther M. John, Manuel R. Teixeira, Paula Paulo, Marta Cardoso, Susan L. Neuhausen, Linda Steele, Yuan Chun Ding, Kim De Ruyck, Gert De Meerleer, Piet Ost, Azad Razack, Jasmine Lim, Soo-Hwang Teo, Daniel W. Lin, Lisa F. Newcomb, Davor Lessel, Marija Gamulin, Tomislav Kulis, Radka Kaneva, Nawaid Usmani, Sandeep Singhal, Chavdar Slavov, Vanio Mitev, Matthew Parliament, Frank Claessens, Steven Joniau, Thomas Van den Broeck, Samantha Larkin, Paul A. Townsend, Claire Aukim-Hastie, Manuela Gago-Dominguez, Jose Esteban Castelao, Maria Elena Martinez, Monique J. Roobol, Guido Jenster, Ron H. N. van Schaik, Florence Menegaux, Thérèse Truong, Yves Akoli Koudou, Jianfeng Xu, Kay-Tee Khaw, Lisa Cannon-Albright, Hardev Pandha, Agnieszka Michael, Stephen N. Thibodeau, Shannon K. McDonnell, Daniel J. Schaid, Sara Lindstrom, Constance Turman, Jing Ma, David J. Hunter, Elio Riboli, Afshan Siddiq, Federico Canzian, Laurence N. Kolonel, Loic Le Marchand, Robert N. Hoover, Mitchell J. Machiela, Zuxi Cui, Peter Kraft, Christopher I. Amos, David V. Conti, Douglas F. Easton, Fredrik Wiklund, Stephen J. Chanock, Brian E. Henderson, Zsofia Kote-Jarai, Christopher A. Haiman, Rosalind A. Eeles, Profile Study, Australian Prostate Cancer BioResource (APCB), IMPACT Study, Canary PASS Investigators, Breast and Prostate Cancer Cohort Consortium (BPC3), PRACTICAL (Prostate Cancer Association Group to Investigate Cancer-Associated Alterations in the Genome) Consortium, Cancer of the Prostate in Sweden (CAPS), Prostate Cancer Genome-wide Association Study of Uncommon Susceptibility Loci (PEGASUS),

and Genetic Associations and Mechanisms in Oncology (GAME-ON)/Elucidating Loci Involved in Prostate Cancer Susceptibility (ELLIPSE) Consortium. Association analyses of more than 140,000 men identify 63 new prostate cancer susceptibility loci. *Nature Genetics*, 50(7):928–936, July 2018. ISSN 1546-1718. doi: 10.1038/s41588-018-0142-8.

[234] Swedish Low-risk Colorectal Cancer Study Group, Simon N Stacey, Patrick Sulem, Aslaug Jonasdottir, Gisli Masson, Julius Gudmundsson, Daniel F Gudbjartsson, Olafur T Magnusson, Sigurjon A Gudjonsson, Bardur Sigurgeirsson, Kristin Thorisdottir, Rafn Ragnarsson, Kristrun R Benediktsdottir, Bjørn A Nexø, Anne Tjønneland, Kim Overvad, Peter Rudnai, Eugene Gurzau, Kvetoslava Koppova, Kari Hemminki, Cristina Corredera, Victoria Fuentelsaz, Pilar Grasa, Sebastian Navarrete, Fernando Fuertes, Maria D García-Prats, Enrique Sanambrosio, Angeles Panadero, Ana De Juan, Almudena Garcia, Fernando Rivera, Dolores Planelles, Virtudes Soriano, Celia Requena, Katja K Aben, Michelle M van Rossum, Ruben G H M Cremers, Inge M van Oort, Dick-Johan van Spronsen, Jack A Schalken, Wilbert H M Peters, Brian T Helfand, Jenny L Donovan, Freddie C Hamdy, Daniel Badescu, Ovidiu Codreanu, Mariana Jinga, Irma E Csiki, Vali Constantinescu, Paula Badea, Ioan N Mates, Daniela E Dinu, Adrian Constantin, Dana Mates, Sjofn Kristjansdottir, Bjarni A Agnarsson, Eirikur Jonsson, Rosa B Barkardottir, Gudmundur V Einarsson, Fridbjorn Sigurdsson, Pall H Moller, Tryggvi Stefansson, Trausti Valdimarsson, Oskar T Johannsson, Helgi Sigurdsson, Thorvaldur Jonsson, Jon G Jonasson, Laufey Tryggvadottir, Terri Rice, Helen M Hansen, Yuanyuan Xiao, Daniel H Lachance, Brian Patrick ONeill, Matthew L Kosel, Paul A Decker, Gudmar Thorleifsson, Hrefna Johannsdottir, Hafdis T Helgadottir, Asgeir Sigurdsson, Valgerdur Steinthorsdottir, Annika Lindblom, Robert S Sandler, Temitope O Keku, Karina Banasik, Torben Jørgensen, Daniel R Witte, Torben Hansen, Oluf Pedersen, Viorel Jinga, David E Neal, William J Catalona, Margaret Wrensch, John Wiencke, Robert B Jenkins, Eduardo Nagore, Ulla Vogel, Lambertus A Kiemeney, Rajiv Kumar, José I Mayordomo, Jon H Olafsson, Augustine Kong, Unnur Thorsteinsdottir, Thorunn Rafnar, and Kari Stefansson. A germline variant in the TP53 polyadenylation signal confers cancer susceptibility. *Nature Genetics*, 43(11):1098–1103, November 2011. ISSN 1061-4036, 1546-1718. doi: 10.1038/ng.926. URL `http://www.nature.com/articles/ng.926`.

[235] Xiaosheng Wang and Qingrong Sun. TP53 mutations, expression and interaction networks in human cancers. *Oncotarget*, 8(1):624–643, January 2017. ISSN 1949-2553. doi: 10.18632/oncotarget.13483.

[236] Tilmann Bürckstümmer, Carina Banning, Philipp Hainzl, Richard Schobesberger, Claudia Kerzendorfer, Florian M Pauler, Doris Chen, Nicole Them, Fiorella Schischlik, Manuele Rebsamen, Michal Smida, Ferran Fece de la Cruz, Ana Lapao, Melissa Liszt, Benjamin Eizinger, Philipp M Guenzl, Vincent A Blomen, Tomasz Konopka, Bianca Gapp, Katja Parapatics, Barbara Maier, Johannes Stöckl, Wolfgang Fischl, Sejla Salic, M Rita Taba Casari, Sylvia Knapp, Keiryn L Bennett, Christoph Bock, Jacques Colinge, Robert Kralovics, Gustav Ammerer, Georg Casari, Thijn R Brum-

melkamp, Giulio Superti-Furga, and Sebastian M B Nijman. A reversible gene trap collection empowers haploid genetics in human cells. *Nature Methods*, 10(10):965–971, October 2013. ISSN 1548-7091, 1548-7105. doi: 10.1038/nmeth.2609. URL `http://www.nature.com/articles/nmeth.2609`.

[237] Andrei V. Gudkov and Elena A. Komarova. The role of p53 in determining sensitivity to radiotherapy. *Nature Reviews. Cancer*, 3(2):117–129, February 2003. ISSN 1474-175X. doi: 10.1038/nrc992.

[238] Xiang Zhou, Qian Hao, and Hua Lu. Mutant p53 in cancer therapy-the barrier or the path. *Journal of Molecular Cell Biology*, 11(4):293–305, April 2019. ISSN 1759-4685. doi: 10.1093/jmcb/mjy072.

[239] A. I. Robles and C. C. Harris. Clinical Outcomes and Correlates of TP53 Mutations and Cancer. *Cold Spring Harbor Perspectives in Biology*, 2(3):a001016–a001016, March 2010. ISSN 1943-0264. doi: 10.1101/cshperspect.a001016. URL `http://cshperspectives.cshlp.org/lookup/doi/10.1101/cshperspect.a001016`.

[240] Rebecca E. Graff, Sören Möller, Michael N. Passarelli, John S. Witte, Axel Skytthe, Kaare Christensen, Qihua Tan, Hans-Olov Adami, Kamila Czene, Jennifer R. Harris, Eero Pukkala, Jaakko Kaprio, Edward L. Giovannucci, Lorelei A. Mucci, and Jacob B. Hjelmborg. Familial Risk and Heritability of Colorectal Cancer in the Nordic Twin Study of Cancer. *Clinical Gastroenterology and Hepatology: The Official Clinical Practice Journal of the American Gastroenterological Association*, 15(8):1256–1264, August 2017. ISSN 1542-7714. doi: 10.1016/j.cgh.2016.12.041.

[241] Jeroen R. Huyghe, Stephanie A. Bien, Tabitha A. Harrison, Hyun Min Kang, Sai Chen, Stephanie L. Schmit, David V. Conti, Conghui Qu, Jihyoun Jeon, Christopher K. Edlund, Peyton Greenside, Michael Wainberg, Fredrick R. Schumacher, Joshua D. Smith, David M. Levine, Sarah C. Nelson, Nasa A. Sinnott-Armstrong, Demetrius Albanes, M. Henar Alonso, Kristin Anderson, Coral Arnau-Collell, Volker Arndt, Christina Bamia, Barbara L. Banbury, John A. Baron, Sonja I. Berndt, Stéphane Bézieau, D. Timothy Bishop, Juergen Boehm, Heiner Boeing, Hermann Brenner, Stefanie Brezina, Stephan Buch, Daniel D. Buchanan, Andrea Burnett-Hartman, Katja Butterbach, Bette J. Caan, Peter T. Campbell, Christopher S. Carlson, Sergi Castellví-Bel, Andrew T. Chan, Jenny Chang-Claude, Stephen J. Chanock, Maria-Dolores Chirlaque, Sang Hee Cho, Charles M. Connolly, Amanda J. Cross, Katarina Cuk, Keith R. Curtis, Albert de la Chapelle, Kimberly F. Doheny, David Duggan, Douglas F. Easton, Sjoerd G. Elias, Faye Elliott, Dallas R. English, Edith J. M. Feskens, Jane C. Figueiredo, Rocky Fischer, Liesel M. FitzGerald, David Forman, Manish Gala, Steven Gallinger, W. James Gauderman, Graham G. Giles, Elizabeth Gillanders, Jian Gong, Phyllis J. Goodman, William M. Grady, John S. Grove, Andrea Gsur, Marc J. Gunter, Robert W. Haile, Jochen Hampe, Heather Hampel, Sophia Harlid, Richard B. Hayes, Philipp Hofer, Michael Hoffmeister, John L. Hopper, Wan-Ling Hsu, Wen-Yi Huang, Thomas J. Hudson, David J. Hunter, Gemma Ibañez-Sanz, Gregory E. Idos, Roxann Ingersoll, Rebecca D. Jackson, Eric J. Jacobs, Mark A. Jenkins, Amit D. Joshi, Corinne E. Joshu, Temitope O. Keku, Timothy J. Key, Hyeong Rok Kim, Emiko Kobayashi, Laurence N. Kolonel,

Charles Kooperberg, Tilman Kühn, Sébastien Küry, Sun-Seog Kweon, Susanna C. Larsson, Cecelia A. Laurie, Loic Le Marchand, Suzanne M. Leal, Soo Chin Lee, Flavio Lejbkowicz, Mathieu Lemire, Christopher I. Li, Li Li, Wolfgang Lieb, Yi Lin, Annika Lindblom, Noralane M. Lindor, Hua Ling, Tin L. Louie, Satu Männistö, Sanford D. Markowitz, Vicente Martín, Giovanna Masala, Caroline E. McNeil, Marilena Melas, Roger L. Milne, Lorena Moreno, Neil Murphy, Robin Myte, Alessio Naccarati, Polly A. Newcomb, Kenneth Offit, Shuji Ogino, N. Charlotte Onland-Moret, Barbara Pardini, Patrick S. Parfrey, Rachel Pearlman, Vittorio Perduca, Paul D. P. Pharoah, Mila Pinchev, Elizabeth A. Platz, Ross L. Prentice, Elizabeth Pugh, Leon Raskin, Gad Rennert, Hedy S. Rennert, Elio Riboli, Miguel Rodríguez-Barranco, Jane Romm, Lori C. Sakoda, Clemens Schafmayer, Robert E. Schoen, Daniela Seminara, Mitul Shah, Tameka Shelford, Min-Ho Shin, Katerina Shulman, Sabina Sieri, Martha L. Slattery, Melissa C. Southey, Zsofia K. Stadler, Christa Stegmaier, Yu-Ru Su, Catherine M. Tangen, Stephen N. Thibodeau, Duncan C. Thomas, Sushma S. Thomas, Amanda E. Toland, Antonia Trichopoulou, Cornelia M. Ulrich, David J. Van Den Berg, Franzel J. B. van Duijnhoven, Bethany Van Guelpen, Henk van Kranen, Joseph Vijai, Kala Visvanathan, Pavel Vodicka, Ludmila Vodickova, Veronika Vymetalkova, Korbinian Weigl, Stephanie J. Weinstein, Emily White, Aung Ko Win, C. Roland Wolf, Alicja Wolk, Michael O. Woods, Anna H. Wu, Syed H. Zaidi, Brent W. Zanke, Qing Zhang, Wei Zheng, Peter C. Scacheri, John D. Potter, Michael C. Bassik, Anshul Kundaje, Graham Casey, Victor Moreno, Goncalo R. Abecasis, Deborah A. Nickerson, Stephen B. Gruber, Li Hsu, and Ulrike Peters. Discovery of common and rare genetic risk variants for colorectal cancer. *Nature Genetics*, 51(1):76–87, 2019. ISSN 1546-1718. doi: 10.1038/s41588-018-0286-6.

[242] Louis Vermeulen and Hugo J. Snippert. Stem cell dynamics in homeostasis and cancer of the intestine. *Nature Reviews Cancer*, 14(7):468–480, July 2014. ISSN 1474-175X, 1474-1768. doi: 10.1038/nrc3744. URL `http://www.nature.com/articles/nrc3744`.

[243] Maureen Spit, Bon-Kyoung Koo, and Madelon M. Maurice. Tales from the crypt: intestinal niche signals in tissue renewal, plasticity and cancer. *Open Biology*, 8(9): 180120, September 2018. ISSN 2046-2441, 2046-2441. doi: 10.1098/rsob.180120. URL `https://royalsocietypublishing.org/doi/10.1098/rsob.180120`.

[244] B. Vogelstein, E. R. Fearon, S. R. Hamilton, S. E. Kern, A. C. Preisinger, M. Leppert, Y. Nakamura, R. White, A. M. Smits, and J. L. Bos. Genetic alterations during colorectal-tumor development. *The New England Journal of Medicine*, 319(9):525–532, September 1988. ISSN 0028-4793. doi: 10.1056/NEJM198809013190901.

[245] E. R. Fearon and B. Vogelstein. A genetic model for colorectal tumorigenesis. *Cell*, 61(5):759–767, June 1990. ISSN 0092-8674. doi: 10.1016/0092-8674(90)90186-i.

[246] Theresa P. Pretlow and Thomas G. Pretlow. Mutant KRAS in aberrant crypt foci (ACF): initiation of colorectal cancer? *Biochimica Et Biophysica Acta*, 1756(2): 83–96, November 2005. ISSN 0006-3002. doi: 10.1016/j.bbcan.2005.06.002.

[247] S. Shirasawa, M. Furuse, N. Yokoyama, and T. Sasazuki. Altered growth of human

colon cancer cell lines disrupted at activated Ki-ras. *Science (New York, N.Y.)*, 260 (5104):85–88, April 1993. ISSN 0036-8075. doi: 10.1126/science.8465203.

[248] O. J. Sansom, V. Meniel, J. A. Wilkins, A. M. Cole, K. A. Oien, V. Marsh, T. J. Jamieson, C. Guerra, G. H. Ashton, M. Barbacid, and A. R. Clarke. Loss of Apc allows phenotypic manifestation of the transforming properties of an endogenous K-ras oncogene in vivo. *Proceedings of the National Academy of Sciences*, 103(38): 14122–14127, September 2006. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.060 4130103. URL http://www.pnas.org/cgi/doi/10.1073/pnas.0604130103.

[249] Kevin M. Haigis, Krystle R. Kendall, Yufang Wang, Ann Cheung, Marcia C. Haigis, Jonathan N. Glickman, Michiko Niwa-Kawakita, Alejandro Sweet-Cordero, Judith Sebolt-Leopold, Kevin M. Shannon, Jeffrey Settleman, Marco Giovannini, and Tyler Jacks. Differential effects of oncogenic K-Ras and N-Ras on proliferation, differentiation and tumor progression in the colon. *Nature Genetics*, 40(5):600–608, May 2008. ISSN 1546-1718. doi: 10.1038/ng.115.

[250] Yuliya Pylayeva-Gupta, Elda Grabocka, and Dafna Bar-Sagi. RAS oncogenes: weaving a tumorigenic web. *Nature Reviews Cancer*, 11(11):761–774, November 2011. ISSN 1474-175X, 1474-1768. doi: 10.1038/nrc3106. URL http://www.nature.com/articles/nrc3106.

[251] W. J. Langlois, T. Sasaoka, A. R. Saltiel, and J. M. Olefsky. Negative feedback regulation and desensitization of insulin- and epidermal growth factor-stimulated p21ras activation. *The Journal of Biological Chemistry*, 270(43):25320–25323, October 1995. ISSN 0021-9258. doi: 10.1074/jbc.270.43.25320.

[252] S. B. Waters, K. H. Holt, S. E. Ross, L. J. Syu, K. L. Guan, A. R. Saltiel, G. A. Koretzky, and J. E. Pessin. Desensitization of Ras activation by a feedback disassociation of the SOS-Grb2 complex. *The Journal of Biological Chemistry*, 270(36): 20883–20886, September 1995. ISSN 0021-9258. doi: 10.1074/jbc.270.36.20883.

[253] Rony Seger and Edwin G. Krebs. The MAPK signaling cascade. *The FASEB Journal*, 9(9):726–735, June 1995. ISSN 0892-6638, 1530-6860. doi: 10.1096/fasebj .9.9.7601337. URL https://onlinelibrary.wiley.com/doi/abs/10.1096/fasebj.9.9.7601337.

[254] Richard J Orton, Michiel E Adriaens, Amelie Gormand, Oliver E Sturm, Walter Kolch, and David R Gilbert. Computational modelling of cancerous mutations in the EGFR/ERK signalling pathway. *BMC Systems Biology*, 3(1):100, December 2009. ISSN 1752-0509. doi: 10.1186/1752-0509-3-100. URL https://bmcsystbiol.biomedcentral.com/articles/10.1186/1752-0509-3-100.

[255] I. R. Vetter and A. Wittinghofer. The guanine nucleotide-binding switch in three dimensions. *Science (New York, N.Y.)*, 294(5545):1299–1304, November 2001. ISSN 0036-8075. doi: 10.1126/science.1062023.

[256] J. C. Hunter, A. Manandhar, M. A. Carrasco, D. Gurbani, S. Gondi, and K. D. Westover. Biochemical and Structural Analysis of Common Cancer-Associated KRAS Mutations. *Molecular Cancer Research*, 13(9):1325–1335, September 2015. ISSN 1541-7786, 1557-3125. doi: 10.1158/1541-7786.MCR-15-0203. URL http://mcr.aacrjournals.org/cgi/doi/10.1158/1541-7786.MCR-15-0203.

[257] M. P. Patricelli, M. R. Janes, L.-S. Li, R. Hansen, U. Peters, L. V. Kessler, Y. Chen, J. M. Kucharski, J. Feng, T. Ely, J. H. Chen, S. J. Firdaus, A. Babbar, P. Ren, and Y. Liu. Selective Inhibition of Oncogenic KRAS Output with Small Molecules Targeting the Inactive State. *Cancer Discovery*, 6(3):316–329, March 2016. ISSN 2159-8274, 2159-8290. doi: 10.1158/2159-8290.CD-15-1105. URL `http://cancerdiscovery.aacrjournals.org/cgi/doi/10.1158/2159-8290.CD-15-1105`.

[258] Carmen Muñoz-Maldonado, Yitzhak Zimmer, and Michaela Medová. A Comparative Analysis of Individual RAS Mutations in Cancer Biology. *Frontiers in Oncology*, 9:1088, October 2019. ISSN 2234-943X. doi: 10.3389/fonc.2019.01088. URL `https://www.frontiersin.org/article/10.3389/fonc.2019.01088/full`.

[259] Shelly R. Calcagno, Shuhua Li, Migdalisel Colon, Pamela A. Kreinest, E. Aubrey Thompson, Alan P. Fields, and Nicole R. Murray. Oncogenic K-ras promotes early carcinogenesis in the mouse proximal colon. *International Journal of Cancer*, 122 (11):2462–2470, June 2008. ISSN 1097-0215. doi: 10.1002/ijc.23383.

[260] Feijun Luo, David G. Brooks, Hongtao Ye, Rifat Hamoudi, George Poulogiannis, Charles E. Patek, Douglas J. Winton, and Mark J. Arends. Mutated K-ras(Asp12) promotes tumourigenesis in Apc(Min) mice more in the large than the small intestines, with synergistic effects between K-ras and Wnt pathways. *International Journal of Experimental Pathology*, 90(5):558–574, October 2009. ISSN 1365-2613. doi: 10.1111/j.1365-2613.2009.00667.x.

[261] Harpreet Wasan, Angela M. Meade, Richard Adams, Richard Wilson, Cheryl Pugh, David Fisher, Benjamin Sydes, Ayman Madi, Bruce Sizer, Charles Lowdell, Gary Middleton, Rachel Butler, Richard Kaplan, Tim Maughan, and COIN-B investigators. Intermittent chemotherapy plus either intermittent or continuous cetuximab for first-line treatment of patients with KRAS wild-type advanced colorectal cancer (COIN-B): a randomised phase 2 trial. *The Lancet. Oncology*, 15(6):631–639, May 2014. ISSN 1474-5488. doi: 10.1016/S1470-2045(14)70106-8.

[262] Ian P. M. Tomlinson, Emily Webb, Luis Carvajal-Carmona, Peter Broderick, Kimberley Howarth, Alan M. Pittman, Sarah Spain, Steven Lubbe, Axel Walther, Kate Sullivan, Emma Jaeger, Sarah Fielding, Andrew Rowan, Jayaram Vijayakrishnan, Enric Domingo, Ian Chandler, Zoe Kemp, Mobshra Qureshi, Susan M. Farrington, Albert Tenesa, James G. D. Prendergast, Rebecca A. Barnetson, Steven Penegar, Ella Barclay, Wendy Wood, Lynn Martin, Maggie Gorman, Huw Thomas, Julian Peto, D. Timothy Bishop, Richard Gray, Eamonn R. Maher, Anneke Lucassen, David Kerr, D. Gareth R. Evans, CORGI Consortium, Clemens Schafmayer, Stephan Buch, Henry Völzke, Jochen Hampe, Stefan Schreiber, Ulrich John, Thibaud Koessler, Paul Pharoah, Tom van Wezel, Hans Morreau, Juul T. Wijnen, John L. Hopper, Melissa C. Southey, Graham G. Giles, Gianluca Severi, Sergi Castellví-Bel, Clara Ruiz-Ponte, Angel Carracedo, Antoni Castells, EPICOLON Consortium, Asta Försti, Kari Hemminki, Pavel Vodicka, Alessio Naccarati, Lara Lipton, Judy W. C. Ho, K. K. Cheng, Pak C. Sham, J. Luk, Jose A. G. Agúndez, Jose M. Ladero, Miguel de la Hoya, Trinidad Caldés, Iina Niittymäki, Sari Tuupanen, Auli Karhu, Lauri Aaltonen, Jean-Baptiste Cazier, Harry Campbell, Mal-

colm G. Dunlop, and Richard S. Houlston. A genome-wide association study identifies colorectal cancer susceptibility loci on chromosomes 10p14 and 8q23.3. *Nature Genetics*, 40(5):623–630, May 2008. ISSN 1546-1718. doi: 10.1038/ng.111.

[263] Nicola Whiffin, Fay J. Hosking, Susan M. Farrington, Claire Palles, Sara E. Dobbins, Lina Zgaga, Amy Lloyd, Ben Kinnersley, Maggie Gorman, Albert Tenesa, Peter Broderick, Yufei Wang, Ella Barclay, Caroline Hayward, Lynn Martin, Daniel D. Buchanan, Aung Ko Win, John Hopper, Mark Jenkins, Noralane M. Lindor, Polly A. Newcomb, Steve Gallinger, David Conti, Fred Schumacher, Graham Casey, Tao Liu, Swedish Low-Risk Colorectal Cancer Study Group, Harry Campbell, Annika Lindblom, Richard S. Houlston, Ian P. Tomlinson, and Malcolm G. Dunlop. Identification of susceptibility loci for colorectal cancer in a genome-wide meta-analysis. *Human Molecular Genetics*, 23(17):4729–4737, September 2014. ISSN 1460-2083. doi: 10.1093/hmg/ddu177.

[264] Ben Zhang, Wei-Hua Jia, Koichi Matsuda, Sun-Seog Kweon, Keitaro Matsuo, Yong-Bing Xiang, Aesun Shin, Sun Ha Jee, Dong-Hyun Kim, Qiuyin Cai, Jirong Long, Jiajun Shi, Wanqing Wen, Gong Yang, Yanfeng Zhang, Chun Li, Bingshan Li, Yan Guo, Zefang Ren, Bu-Tian Ji, Zhi-Zhong Pan, Atsushi Takahashi, Min-Ho Shin, Fumihiko Matsuda, Yu-Tang Gao, Jae Hwan Oh, Soriul Kim, Yoon-Ok Ahn, Genetics and Epidemiology of Colorectal Cancer Consortium (GECCO), Andrew T. Chan, Jenny Chang-Claude, Martha L. Slattery, Colorectal Transdisciplinary (CORECT) Study, Stephen B. Gruber, Fredrick R. Schumacher, Stephanie L. Stenzel, Colon Cancer Family Registry (CCFR), Graham Casey, Hyeong-Rok Kim, Jin-Young Jeong, Ji Won Park, Hong-Lan Li, Satoyo Hosono, Sang-Hee Cho, Michiaki Kubo, Xiao-Ou Shu, Yi-Xin Zeng, and Wei Zheng. Large-scale genetic study in East Asians identifies six new loci associated with colorectal cancer risk. *Nature Genetics*, 46(6): 533–542, June 2014. ISSN 1546-1718. doi: 10.1038/ng.2985.

[265] Chizu Tanikawa, Yoichiro Kamatani, Atsushi Takahashi, Yukihide Momozawa, Karine Leveque, Satoshi Nagayama, Koshi Mimori, Masaki Mori, Hideshi Ishii, Johji Inazawa, Jun Yasuda, Akito Tsuboi, Atsushi Shimizu, Makoto Sasaki, Taiki Yamaji, Norie Sawada, Motoki Iwasaki, Shoichiro Tsugane, Mariko Naito, Kenji Wakai, Teruhide Koyama, Toshiro Takezaki, Koichiro Yuji, Yoshinori Murakami, Yusuke Nakamura, Michiaki Kubo, and Koichi Matsuda. GWAS identifies two novel colorectal cancer loci at 16q24.1 and 20q13.12. *Carcinogenesis*, 39(5):652–660, 2018. ISSN 1460-2180. doi: 10.1093/carcin/bgy026.

[266] Yingchang Lu, Sun-Seog Kweon, Chizu Tanikawa, Wei-Hua Jia, Yong-Bing Xiang, Qiuyin Cai, Chenjie Zeng, Stephanie L. Schmit, Aesun Shin, Keitaro Matsuo, Sun Ha Jee, Dong-Hyun Kim, Jeongseon Kim, Wanqing Wen, Jiajun Shi, Xingyi Guo, Bingshan Li, Nan Wang, Ben Zhang, Xinxiang Li, Min-Ho Shin, Hong-Lan Li, Zefang Ren, Jae Hwan Oh, Isao Oze, Yoon-Ok Ahn, Keum Ji Jung, David V. Conti, Fredrick R. Schumacher, Gad Rennert, Mark A. Jenkins, Peter T. Campbell, Michael Hoffmeister, Graham Casey, Stephen B. Gruber, Jing Gao, Yu-Tang Gao, Zhi-Zhong Pan, Yoichiro Kamatani, Yi-Xin Zeng, Xiao-Ou Shu, Jirong Long, Koichi Matsuda, and Wei Zheng. Large-Scale Genome-Wide Association Study of East Asians Iden-

tifies Loci Associated With Risk for Colorectal Cancer. *Gastroenterology*, 156(5): 1455–1466, April 2019. ISSN 1528-0012. doi: 10.1053/j.gastro.2018.11.066.

[267] Stephanie L. Schmit, Christopher K. Edlund, Fredrick R. Schumacher, Jian Gong, Tabitha A. Harrison, Jeroen R. Huyghe, Chenxu Qu, Marilena Melas, David J. Van Den Berg, Hansong Wang, Stephanie Tring, Sarah J. Plummer, Demetrius Albanes, M. Henar Alonso, Christopher I. Amos, Kristen Anton, Aaron K. Aragaki, Volker Arndt, Elizabeth L. Barry, Sonja I. Berndt, Stéphane Bezieau, Stephanie Bien, Amanda Bloomer, Juergen Boehm, Marie-Christine Boutron-Ruault, Hermann Brenner, Stefanie Brezina, Daniel D. Buchanan, Katja Butterbach, Bette J. Caan, Peter T. Campbell, Christopher S. Carlson, Jose E. Castelao, Andrew T. Chan, Jenny Chang-Claude, Stephen J. Chanock, Iona Cheng, Ya-Wen Cheng, Lee Soo Chin, James M. Church, Timothy Church, Gerhard A. Coetzee, Michelle Cotterchio, Marcia Cruz Correa, Keith R. Curtis, David Duggan, Douglas F. Easton, Dallas English, Edith J. M. Feskens, Rocky Fischer, Liesel M. FitzGerald, Barbara K. Fortini, Lars G. Fritsche, Charles S. Fuchs, Manuela Gago-Dominguez, Manish Gala, Steven J. Gallinger, W. James Gauderman, Graham G. Giles, Edward L. Giovannucci, Stephanie M. Gogarten, Clicerio Gonzalez-Villalpando, Elena M. Gonzalez-Villalpando, William M. Grady, Joel K. Greenson, Andrea Gsur, Marc Gunter, Christopher A. Haiman, Jochen Hampe, Sophia Harlid, John F. Harju, Richard B. Hayes, Philipp Hofer, Michael Hoffmeister, John L. Hopper, Shu-Chen Huang, Jose Maria Huerta, Thomas J. Hudson, David J. Hunter, Gregory E. Idos, Motoki Iwasaki, Rebecca D. Jackson, Eric J. Jacobs, Sun Ha Jee, Mark A. Jenkins, Wei-Hua Jia, Shuo Jiao, Amit D. Joshi, Laurence N. Kolonel, Suminori Kono, Charles Kooperberg, Vittorio Krogh, Tilman Kuehn, Sébastien Küry, Andrea LaCroix, Cecelia A. Laurie, Flavio Lejbkowicz, Mathieu Lemire, Heinz-Josef Lenz, David Levine, Christopher I. Li, Li Li, Wolfgang Lieb, Yi Lin, Noralane M. Lindor, Yun-Ru Liu, Fotios Loupakis, Yingchang Lu, Frank Luh, Jing Ma, Christoph Mancao, Frank J. Manion, Sanford D. Markowitz, Vicente Martin, Koichi Matsuda, Keitaro Matsuo, Kevin J. McDonnell, Caroline E. McNeil, Roger Milne, Antonio J. Molina, Bhramar Mukherjee, Neil Murphy, Polly A. Newcomb, Kenneth Offit, Hanane Omichessan, Domenico Palli, Jesus P. Paredes Cotoré, Julyann Pérez-Mayoral, Paul D. Pharoah, John D. Potter, Conghui Qu, Leon Raskin, Gad Rennert, Hedy S. Rennert, Bridget M. Riggs, Clemens Schafmayer, Robert E. Schoen, Thomas A. Sellers, Daniela Seminara, Gianluca Severi, Wei Shi, David Shibata, Xiao-Ou Shu, Erin M. Siegel, Martha L. Slattery, Melissa Southey, Zsofia K. Stadler, Mariana C. Stern, Sebastian Stintzing, Darin Taverna, Stephen N. Thibodeau, Duncan C. Thomas, Antonia Trichopoulou, Shoichiro Tsugane, Cornelia M. Ulrich, Franzel J. B. van Duijnhoven, Bethany van Guelpan, Joseph Vijai, Jarmo Virtamo, Stephanie J. Weinstein, Emily White, Aung Ko Win, Alicja Wolk, Michael Woods, Anna H. Wu, Kana Wu, Yong-Bing Xiang, Yun Yen, Brent W. Zanke, Yi-Xin Zeng, Ben Zhang, Niha Zubair, Sun-Seog Kweon, Jane C. Figueiredo, Wei Zheng, Loic Le Marchand, Annika Lindblom, Victor Moreno, Ulrike Peters, Graham Casey, Li Hsu, David V. Conti, and Stephen B. Gruber. Novel Common Genetic Susceptibility Loci for Colorectal Can-

cer. *Journal of the National Cancer Institute*, 111(2):146–157, 2019. ISSN 1460-2105. doi: 10.1093/jnci/djy099.

[268] Yingchang Lu, Sun-Seog Kweon, Qiuyin Cai, Chizu Tanikawa, Xiao-Ou Shu, Wei-Hua Jia, Yong-Bing Xiang, Jeroen R. Huyghe, Tabitha A. Harrison, Jeongseon Kim, Aesun Shin, Dong-Hyun Kim, Keitaro Matsuo, Sun Ha Jee, Xingyi Guo, Wanqing Wen, Jiajun Shi, Bingshan Li, Nan Wang, Min-Ho Shin, Hong-Lan Li, Zefang Ren, Jae Hwan Oh, Isao Oze, Yoon-Ok Ahn, Keum Ji Jung, Jing Gao, Yu-Tang Gao, Zhi-Zhong Pan, Yoichiro Kamatani, Andrew T. Chan, Andrea Gsur, Jochen Hampe, Loic Le Marchand, Li Li, Annika Lindblom, Victor Moreno, Polly A. Newcomb, Kenneth Offit, Paul D. P. Pharoah, Franzel J. B. van Duijnhoven, Bethany Van Guelpen, Pavel Vodicka, Stephanie J. Weinstein, Alicja Wolk, Anna H. Wu, Li Hsu, Yi-Xin Zeng, Jirong Long, Ulrike Peters, Koichi Matsuda, and Wei Zheng. Identification of Novel Loci and New Risk Variant in Known Loci for Colorectal Cancer Risk in East Asians. *Cancer Epidemiology, Biomarkers & Prevention: A Publication of the American Association for Cancer Research, Cosponsored by the American Society of Preventive Oncology*, 29(2):477–486, 2020. ISSN 1538-7755. doi: 10.1158/1055-9965.EPI-19-0755.

[269] S. von Holst, S. Picelli, D. Edler, C. Lenander, J. Dalén, F. Hjern, N. Lundqvist, U. Lindforss, L. Påhlman, K. Smedh, A. Törnqvist, J. Holm, M. Janson, M. Andersson, S. Ekelund, L. Olsson, S. Ghazi, N. Papadogiannakis, A. Tenesa, S. M. Farrington, H. Campbell, M. G. Dunlop, and A. Lindblom. Association studies on 11 published colorectal cancer risk loci. *British Journal of Cancer*, 103(4):575–580, August 2010. ISSN 1532-1827. doi: 10.1038/sj.bjc.6605774.

[270] Fang Xiong, Chen Wu, Xinyu Bi, Dianke Yu, Liming Huang, Jian Xu, Tongwen Zhang, Kan Zhai, Jiang Chang, Wen Tan, Jianqiang Cai, and Dongxin Lin. Risk of genome-wide association study-identified genetic variants for colorectal cancer in a Chinese population. *Cancer Epidemiology, Biomarkers & Prevention: A Publication of the American Association for Cancer Research, Cosponsored by the American Society of Preventive Oncology*, 19(7):1855–1861, July 2010. ISSN 1538-7755. doi: 10.1158/1055-9965.EPI-10-0210.

[271] Iina Niittymäki, Eevi Kaasinen, Sari Tuupanen, Auli Karhu, Heikki Järvinen, Jukka-Pekka Mecklin, Ian P. M. Tomlinson, Maria Chiara Di Bernardo, Richard S. Houlston, and Lauri A. Aaltonen. Low-penetrance susceptibility variants in familial colorectal cancer. *Cancer Epidemiology, Biomarkers & Prevention: A Publication of the American Association for Cancer Research, Cosponsored by the American Society of Preventive Oncology*, 19(6):1478–1483, June 2010. ISSN 1538-7755. doi: 10.1158/1055-9965.EPI-09-1320.

[272] J. W. Ho, S.-c Choi, Y.-f Lee, T. C. Hui, S. S. Cherny, M.-M. Garcia-Barceló, L. Carvajal-Carmona, R. Liu, S.-h To, T.-k Yau, C. C. Chung, C. C. Yau, S. M. Hui, P. Y. Lau, C.-h Yuen, Y.-w Wong, S. Ho, S. S. Fung, I. P. Tomlinson, R. S. Houlston, K. K. Cheng, and P. C. Sham. Replication study of SNP associations for colorectal cancer in Hong Kong Chinese. *British Journal of Cancer*, 104(2):369–375, January 2011. ISSN 1532-1827. doi: 10.1038/sj.bjc.6605977.

[273] María Dolores Giráldez, Adriana López-Dóriga, Luis Bujanda, Anna Abulí, Xavier Bessa, Ceres Fernández-Rozadilla, Jenifer Muñoz, Miriam Cuatrecasas, Rodrigo Jover, Rosa M. Xicola, Xavier Llor, Josep M. Piqué, Angel Carracedo, Clara Ruiz-Ponte, Angel Cosme, José María Enríquez-Navascués, Victor Moreno, Montserrat Andreu, Antoni Castells, Francesc Balaguer, Sergi Castellví-Bel, and Gastrointestinal Oncology Group of the Spanish Gastroenterological Association. Susceptibility genetic variants associated with early-onset colorectal cancer. *Carcinogenesis*, 33(3): 613–619, March 2012. ISSN 1460-2180. doi: 10.1093/carcin/bgs009.

[274] Qin Qin, Li Liu, Rong Zhong, Li Zou, Jieyun Yin, Beibei Zhu, Beibei Cao, Wei Chen, Jigui Chen, Xiaorong Li, Tingting Li, Xuzai Lu, Jiao Lou, Juntao Ke, Sheng Wei, Xiaoping Miao, and Shaofa Nie. The genetic variant on chromosome 10p14 is associated with risk of colorectal cancer: results from a case-control study and a meta-analysis. *PloS One*, 8(5):e64310, 2013. ISSN 1932-6203. doi: 10.1371/journal.pone.0064310.

[275] Luis G. Carvajal-Carmona, Ann G. Zauber, Angela M. Jones, Kimberley Howarth, Jiping Wang, Timothy Cheng, APC Trial Collaborators, APPROVe Trial Collaborators, CORGI Study Collaborators, Colon Cancer Family Registry Collaborators, CGEMS Collaborators, Robert Riddell, Angel Lanas, Dion Morton, Monica M. Bertagnolli, and Ian Tomlinson. Much of the genetic risk of colorectal cancer is likely to be mediated through susceptibility to adenomas. *Gastroenterology*, 144(1): 53–55, January 2013. ISSN 1528-0012. doi: 10.1053/j.gastro.2012.09.016.

[276] Andrea N. Burnett-Hartman, Polly A. Newcomb, Carolyn M. Hutter, Ulrike Peters, Michael N. Passarelli, Malaika R. Schwartz, Melissa P. Upton, Lee-Ching Zhu, John D. Potter, and Karen W. Makar. Variation in the association between colorectal cancer susceptibility loci and colorectal polyps by polyp type. *American Journal of Epidemiology*, 180(2):223–232, July 2014. ISSN 1476-6256. doi: 10.1093/aje/kwu114.

[277] Timothy H. T. Cheng, Maggie Gorman, Lynn Martin, Ella Barclay, Graham Casey, Colon Cancer Family Registry, CGEMS, Brian Saunders, Huw Thomas, Sue Clark, and Ian Tomlinson. Common colorectal cancer risk alleles contribute to the multiple colorectal adenoma phenotype, but do not influence colonic polyposis in FAP. *European journal of human genetics: EJHG*, 23(2):260–263, February 2015. ISSN 1476-5438. doi: 10.1038/ejhg.2014.74.

[278] Sung Noh Hong, Changho Park, Jong-Il Kim, Duk-Hwan Kim, Hee Cheol Kim, Dong Kyung Chang, Poong-Lyul Rhee, Jae J. Kim, Jong Chul Rhee, Hee Jung Son, and Young-Ho Kim. Colorectal cancer-susceptibility single-nucleotide polymorphisms in Korean population. *Journal of Gastroenterology and Hepatology*, 30 (5):849–857, May 2015. ISSN 1440-1746. doi: 10.1111/jgh.12331.

[279] Michael R. Burgess, Eugene Hwang, Rana Mroue, Craig M. Bielski, Anica M. Wandler, Benjamin J. Huang, Ari J. Firestone, Amy Young, Jennifer A. Lacap, Lisa Crocker, Saurabh Asthana, Elizabeth M. Davis, Jin Xu, Keiko Akagi, Michelle M. Le Beau, Qing Li, Benjamin Haley, David Stokoe, Deepak Sampath, Barry S. Taylor, Marie Evangelista, and Kevin Shannon. KRAS Allelic Imbalance Enhances

Fitness and Modulates MAP Kinase Dependence in Cancer. *Cell*, 168(5):817–829.e15, February 2017. ISSN 00928674. doi: 10.1016/j.cell.2017.01.020. URL https://linkinghub.elsevier.com/retrieve/pii/S0092867417301058.

[280] Lenora W. M. Loo, Iona Cheng, Maarit Tiirikainen, Annette Lum-Jones, Ann Seifried, Lucas M. Dunklee, James M. Church, Robert Gryfe, Daniel J. Weisenberger, Robert W. Haile, Steven Gallinger, David J. Duggan, Stephen N. Thibodeau, Graham Casey, and Loïc Le Marchand. cis-Expression QTL Analysis of Established Colorectal Cancer Risk Variants in Colon Tumors and Adjacent Normal Tissue. *PLoS ONE*, 7(2):e30477, February 2012. ISSN 1932-6203. doi: 10.1371/journal.pone.0030477. URL https://dx.plos.org/10.1371/journal.pone.0030477.

[281] O. Warburg, F. Wind, and E. Negelein. THE METABOLISM OF TUMORS IN THE BODY. *The Journal of General Physiology*, 8(6):519–530, March 1927. ISSN 0022-1295. doi: 10.1085/jgp.8.6.519.

[282] Arnold J. Levine and Anna M. Puzio-Kuter. The control of the metabolic switch in cancers by oncogenes and tumor suppressor genes. *Science (New York, N.Y.)*, 330 (6009):1340–1344, December 2010. ISSN 1095-9203. doi: 10.1126/science.1193494.

[283] José M. Cuezva, Maryla Krajewska, Miguel López de Heredia, Stanislaw Krajewski, Gema Santamaría, Hoguen Kim, Juan M. Zapata, Hiroyuki Marusawa, Margarita Chamorro, and John C. Reed. The bioenergetic signature of cancer: a marker of tumor progression. *Cancer Research*, 62(22):6674–6681, November 2002. ISSN 0008-5472.

[284] Imke M. Willers, Antonio Isidoro, Alvaro D. Ortega, Pedro L. Fernández, and José M. Cuezva. Selective inhibition of beta-F1-ATPase mRNA translation in human tumours. *The Biochemical Journal*, 426(3):319–326, February 2010. ISSN 1470-8728. doi: 10.1042/BJ20091570.

[285] Kenji Kawada, Kosuke Toda, and Yoshiharu Sakai. Targeting metabolic reprogramming in KRAS-driven cancers. *International Journal of Clinical Oncology*, 22(4): 651–659, August 2017. ISSN 1437-7772. doi: 10.1007/s10147-017-1156-4.

[286] Jesus Lascorz, Melanie Bevier, Witigo V Schönfels, Holger Kalthoff, Heiko Aselmann, Jan Beckmann, Jan Egberts, Stephan Buch, Thomas Becker, Stefan Schreiber, Jochen Hampe, Kari Hemminki, Asta Försti, and Clemens Schafmayer. Polymorphisms in the mitochondrial oxidative phosphorylation chain genes as prognostic markers for colorectal cancer. *BMC Medical Genetics*, 13(1):31, December 2012. ISSN 1471-2350. doi: 10.1186/1471-2350-13-31. URL http://bmcmedgenet.biomedcentral.com/articles/10.1186/1471-2350-13-31.

[287] Yu Imamura, Teppei Morikawa, Xiaoyun Liao, Paul Lochhead, Aya Kuchiba, Mai Yamauchi, Zhi Rong Qian, Reiko Nishihara, Jeffrey A. Meyerhardt, Kevin M. Haigis, Charles S. Fuchs, and Shuji Ogino. Specific mutations in KRAS codons 12 and 13, and patient prognosis in 1075 BRAF wild-type colorectal cancers. *Clinical Cancer Research: An Official Journal of the American Association for Cancer Research*, 18 (17):4753–4763, September 2012. ISSN 1078-0432. doi: 10.1158/1078-0432.CCR-11-3210.

[288] H. Blons, J. F. Emile, K. Le Malicot, C. Julié, A. Zaanan, J. Tabernero, E. Mini, G. Folprecht, J. L. Van Laethem, J. Thaler, J. Bridgewater, L. Nørgård-Petersen, E. Van Cutsem, C. Lepage, M. A. Zawadi, R. Salazar, P. Laurent-Puig, and J. Taieb. Prognostic value of KRAS mutations in stage III colon cancer: post hoc analysis of the PETACC8 phase III trial dataset. *Annals of Oncology: Official Journal of the European Society for Medical Oncology*, 25(12):2378–2385, December 2014. ISSN 1569-8041. doi: 10.1093/annonc/mdu464.

[289] Dae-Won Lee, Kyung Ju Kim, Sae-Won Han, Hyun Jung Lee, Ye Young Rhee, Jeong Mo Bae, Nam-Yun Cho, Kyung-Hun Lee, Tae-Yong Kim, Do-Youn Oh, Seock-Ah Im, Yung-Jue Bang, Seung-Yong Jeong, Kyu Joo Park, Jae-Gahb Park, Gyeong Hoon Kang, and Tae-You Kim. KRAS mutation is associated with worse prognosis in stage III or high-risk stage II colon cancer patients treated with adjuvant FOLFOX. *Annals of Surgical Oncology*, 22(1):187–194, January 2015. ISSN 1534-4681. doi: 10.1245/s10434-014-3826-z.

[290] Jeanne Tie, Lara Lipton, Jayesh Desai, Peter Gibbs, Robert N. Jorissen, Michael Christie, Katharine J. Drummond, Benjamin N. J. Thomson, Valery Usatoff, Peter M. Evans, Adrian W. Pick, Simon Knight, Peter W. G. Carne, Roger Berry, Adrian Polglase, Paul McMurrick, Qi Zhao, Dana Busam, Robert L. Strausberg, Enric Domingo, Ian P. M. Tomlinson, Rachel Midgley, David Kerr, and Oliver M. Sieber. KRAS mutation is associated with lung metastasis in patients with curatively resected colorectal cancer. *Clinical Cancer Research: An Official Journal of the American Association for Cancer Research*, 17(5):1122–1130, March 2011. ISSN 1078-0432. doi: 10.1158/1078-0432.CCR-10-1720.

[291] Jinliang Xing, Ronald E. Myers, Xianli He, Falin Qu, Feng Zhou, Xi Ma, Terry Hyslop, Guoqiang Bao, Shaogui Wan, Hushan Yang, and Zhinan Chen. GWAS-identified colorectal cancer susceptibility locus associates with disease prognosis. *European Journal of Cancer (Oxford, England: 1990)*, 47(11):1699–1707, July 2011. ISSN 1879-0852. doi: 10.1016/j.ejca.2011.02.004.

[292] Amanda I. Phipps, Polly A. Newcomb, Xabier Garcia-Albeniz, Carolyn M. Hutter, Emily White, Charles S. Fuchs, Aditi Hazra, Shuji Ogino, Hongmei Nan, Jing Ma, Peter T. Campbell, Jane C. Figueiredo, Ulrike Peters, and Andrew T. Chan. Association between colorectal cancer susceptibility loci and survival time after diagnosis with colorectal cancer. *Gastroenterology*, 143(1):51–54.e4, July 2012. ISSN 1528-0012. doi: 10.1053/j.gastro.2012.04.052.

[293] Anna Abulí, Juan José Lozano, María Rodríguez-Soler, Rodrigo Jover, Xavier Bessa, Jenifer Muñoz, Clara Esteban-Jurado, Ceres Fernández-Rozadilla, Angel Carracedo, Clara Ruiz-Ponte, Joaquín Cubiella, Francesc Balaguer, Luis Bujanda, Josep M. Reñé, Juan Clofent, Juan Diego Morillas, David Nicolás-Pérez, Rosa M. Xicola, Xavier Llor, Josep M. Piqué, Montserrat Andreu, Antoni Castells, Sergi Castellví-Bel, and Gastrointestinal Oncology Group of Spanish Gastroenterological Association. Genetic susceptibility variants associated with colorectal cancer prognosis. *Carcinogenesis*, 34(10):2286–2291, October 2013. ISSN 1460-2180. doi: 10.1093/carcin/bgt179.

[294] Hanna K. Sanoff, Lindsay A. Renfro, Pradeep Poonnen, Pratibha Ambadwar, Daniel J. Sargent, Richard M. Goldberg, and Howard McLeod. Germline variation in colorectal risk Loci does not influence treatment effect or survival in metastatic colorectal cancer. *PloS One*, 9(4):e94727, 2014. ISSN 1932-6203. doi: 10.1371/journal.pone.0094727.

[295] Chiara De Divitiis. Prognostic and predictive response factors in colorectal cancer patients: Between hope and reality. *World Journal of Gastroenterology*, 20(41): 15049, 2014. ISSN 1007-9327. doi: 10.3748/wjg.v20.i41.15049. URL `http://www.wjgnet.com/1007-9327/full/v20/i41/15049.htm`.

[296] Evelien Dekker, Pieter J Tanis, Jasper L A Vleugels, Pashtoon M Kasi, and Michael B Wallace. Colorectal cancer. *The Lancet*, 394(10207):1467–1480, October 2019. ISSN 01406736. doi: 10.1016/S0140-6736(19)32319-0. URL `https://linkinghub.elsevier.com/retrieve/pii/S0140673619323190`.

[297] Mark A. Hull, Colin J. Rees, Linda Sharp, and Sara Koo. A risk-stratified approach to colorectal cancer prevention and diagnosis. *Nature Reviews Gastroenterology & Hepatology*, 17(12):773–780, December 2020. ISSN 1759-5045, 1759-5053. doi: 10.1038/s41575-020-00368-3. URL `http://www.nature.com/articles/s41575-020-00368-3`.

[298] Linda S Lindström, Per Hall, Mikael Hartman, Fredrik Wiklund, Henrik Grönberg, and Kamila Czene. Familial concordance in cancer survival: a Swedish population-based study. *The Lancet Oncology*, 8(11):1001–1006, November 2007. ISSN 14702045. doi: 10.1016/S1470-2045(07)70282-6. URL `https://linkinghub.elsevier.com/retrieve/pii/S1470204507702826`.

[299] F. A. Castro, A. Försti, S. Buch, H. Kalthoff, C. Krauss, M. Bauer, J. Egberts, B. Schniewind, D. C. Broering, S. Schreiber, M. Schmitt, J. Hampe, K. Hemminki, and C. Schafmayer. TLR-3 polymorphism is an independent prognostic marker for stage II colorectal cancer. *European Journal of Cancer (Oxford, England: 1990)*, 47 (8):1203–1210, May 2011. ISSN 1879-0852. doi: 10.1016/j.ejca.2010.12.011.

[300] S. N. Klimosch, A. Forsti, J. Eckert, J. Knezevic, M. Bevier, W. von Schonfels, N. Heits, J. Walter, S. Hinz, J. Lascorz, J. Hampe, D. Hartl, J.-S. Frick, K. Hemminki, C. Schafmayer, and A. N. R. Weber. Functional TLR5 Genetic Variants Affect Human Colorectal Cancer Survival. *Cancer Research*, 73(24):7232–7242, December 2013. ISSN 0008-5472, 1538-7445. doi: 10.1158/0008-5472.CAN-13-1746. URL `http://cancerres.aacrjournals.org/cgi/doi/10.1158/0008-5472.CAN-13-1746`.

[301] Jan Dimberg, Levar Shamoun, Kalle Landerholm, Roland E. Andersson, Blanka Kolodziej, and Dick Wågsäter. Genetic Variants of the *IL2* Gene Related to Risk and Survival in Patients With Colorectal Cancer. *Anticancer Research*, 39(9):4933–4940, September 2019. ISSN 0250-7005, 1791-7530. doi: 10.21873/anticanres.13681. URL `http://ar.iiarjournals.org/lookup/doi/10.21873/anticanres.13681`.

[302] A. Tenesa, E. Theodoratou, F. V. N. Din, S. M. Farrington, R. Cetnarskyj, R. A. Barnetson, M. E. Porteous, H. Campbell, and M. G. Dunlop. Ten Common Genetic Variants Associated with Colorectal Cancer Risk Are Not Associated with Survival

after Diagnosis. *Clinical Cancer Research*, 16(14):3754–3759, July 2010. ISSN 1078-0432, 1557-3265. doi: 10.1158/1078-0432.CCR-10-0439. URL http://clincancerres.aacrjournals.org/cgi/doi/10.1158/1078-0432.CCR-10-0439.

[303] Wei Xu, Jingxiong Xu, Konstantin Shestopaloff, Elizabeth Dicks, Jane Green, Patrick Parfrey, Roger Green, and Sevtap Savas. A genome wide association study on Newfoundland colorectal cancer patients' survival outcomes. *Biomarker Research*, 3(1):6, December 2015. ISSN 2050-7771. doi: 10.1186/s40364-015-0031-6. URL https://biomarkerres.biomedcentral.com/articles/10.1186/s40364-015-0031-6.

[304] Amanda I. Phipps, Michael N. Passarelli, Andrew T. Chan, Tabitha A. Harrison, Jihyoun Jeon, Carolyn M. Hutter, Sonja I. Berndt, Hermann Brenner, Bette J. Caan, Peter T. Campbell, Jenny Chang-Claude, Stephen J. Chanock, Jeremy P. Cheadle, Keith R. Curtis, David Duggan, David Fisher, Charles S. Fuchs, Manish Gala, Edward L. Giovannucci, Richard B. Hayes, Michael Hoffmeister, Li Hsu, Eric J. Jacobs, Lina Jansen, Richard Kaplan, Elisabeth J. Kap, Timothy S. Maughan, John D. Potter, Robert E. Schoen, Daniela Seminara, Martha L. Slattery, Hannah West, Emily White, Ulrike Peters, and Polly A. Newcomb. Common genetic variation and survival after colorectal cancer diagnosis: a genome-wide analysis. *Carcinogenesis*, 37 (1):87–95, January 2016. ISSN 0143-3334, 1460-2180. doi: 10.1093/carcin/bgv161. URL https://academic.oup.com/carcin/article-lookup/doi/10.1093/carcin/bgv161.

[305] Evropi Theodoratou, Susan M Farrington, Maria Timofeeva, Farhat VN Din, Victoria Svinti, Albert Tenesa, Tao Liu, Annika Lindblom, Steven Gallinger, Harry Campbell, and Malcolm G Dunlop. Genome-wide scan of the effect of common nsSNPs on colorectal cancer survival outcome. *British Journal of Cancer*, 119(8):988–993, October 2018. ISSN 0007-0920, 1532-1827. doi: 10.1038/s41416-018-0117-7. URL http://www.nature.com/articles/s41416-018-0117-7.

[306] Victoria Gray, Sarah Briggs, Claire Palles, Emma Jaeger, Timothy Iveson, Rachel Kerr, Mark P Saunders, James Paul, Andrea Harkin, John McQueen, Matthew G Summers, Elaine Johnstone, Haitao Wang, Laura Gatcombe, Timothy S Maughan, Richard Kaplan, Valentina Escott-Price, Nada A Al-Tassan, Brian F Meyer, Salma M Wakil, Richard S Houlston, Jeremy P Cheadle, Ian Tomlinson, and David N Church. Pattern Recognition Receptor Polymorphisms as Predictors of Oxaliplatin Benefit in Colorectal Cancer. *JNCI: Journal of the National Cancer Institute*, 111 (8):828–836, August 2019. ISSN 0027-8874, 1460-2105. doi: 10.1093/jnci/djy215. URL https://academic.oup.com/jnci/article/111/8/828/5289445.

[307] Yazhou He, Evropi Theodoratou, Xue Li, Farhat V.N. Din, Peter Vaughan-Shaw, Victoria Svinti, Susan M. Farrington, Harry Campbell, Malcolm G. Dunlop, and Maria Timofeeva. Effects of common genetic variants associated with colorectal cancer risk on survival outcomes after diagnosis: A large population-based cohort study. *International Journal of Cancer*, 145(9):2427–2432, November 2019. ISSN 0020-7136, 1097-0215. doi: 10.1002/ijc.32550. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/ijc.32550.

[308] Yazhou He, Maria Timofeeva, Xue Li, Farhat V.N. Din, James P. Blackmur, Peter Vaughan-Shaw, Victoria Svinti, Susan M. Farrington, Harry Campbell, Malcolm G. Dunlop, and Evropi Theodoratou. A Comprehensive Study of the Effect on Colorectal Cancer Survival of Common Germline Genetic Variation Previously Linked with Cancer Prognosis. *Cancer Epidemiology Biomarkers & Prevention*, 28(11):1944–1946, November 2019. ISSN 1055-9965, 1538-7755. doi: 10.1158/1055-9965.EPI-19-0596. URL http://cebp.aacrjournals.org/lookup/doi/10.1158/1055-9965.EPI-19-0596.

[309] Kathryn L. Penney, Barbara L. Banbury, Stephanie Bien, Tabitha A. Harrison, Xinwei Hua, Amanda I. Phipps, Wei Sun, Mingyang Song, Amit D. Joshi, Steven R. Alberts, Carmen J. Allegra, James Atkins, Linda H. Colangelo, Thomas J. George, Richard M. Goldberg, Peter C. Lucas, Suresh G. Nair, Qian Shi, Frank A. Sinicrope, Norman Wolmark, Greg Yothers, Ulrike Peters, Polly A. Newcomb, and Andrew T. Chan. Genetic Variant Associated With Survival of Patients With Stage II-III Colon Cancer. *Clinical Gastroenterology and Hepatology: The Official Clinical Practice Journal of the American Gastroenterological Association*, 18(12):2717–2723.e3, November 2020. ISSN 1542-7714. doi: 10.1016/j.cgh.2019.11.046.

[310] Lisa M. Coussens and Zena Werb. Inflammation and cancer. *Nature*, 420(6917):860–867, December 2002. ISSN 0028-0836. doi: 10.1038/nature01322.

[311] Wolf Herman Fridman, Franck Pagès, Catherine Sautès-Fridman, and Jérôme Galon. The immune contexture in human tumours: impact on clinical outcome. *Nature Reviews. Cancer*, 12(4):298–306, March 2012. ISSN 1474-1768. doi: 10.1038/nrc3245.

[312] Hugo Gonzalez, Catharina Hagerling, and Zena Werb. Roles of the immune system in cancer: from tumor initiation to metastatic progression. *Genes & Development*, 32(19-20):1267–1284, October 2018. ISSN 0890-9369, 1549-5477. doi: 10.1101/gad.314617.118. URL http://genesdev.cshlp.org/lookup/doi/10.1101/gad.314617.118.

[313] Alex D. Waldman, Jill M. Fritz, and Michael J. Lenardo. A guide to cancer immunotherapy: from T cell basic science to clinical practice. *Nature Reviews Immunology*, 20(11):651–668, November 2020. ISSN 1474-1733, 1474-1741. doi: 10.1038/s41577-020-0306-5. URL http://www.nature.com/articles/s41577-020-0306-5.

[314] E F Leitch, M Chakrabarti, J E M Crozier, R F McKee, J H Anderson, P G Horgan, and D C McMillan. Comparison of the prognostic value of selected markers of the systemic inflammatory response in patients with colorectal cancer. *British Journal of Cancer*, 97(9):1266–1270, November 2007. ISSN 0007-0920, 1532-1827. doi: 10.1038/sj.bjc.6604027. URL http://www.nature.com/articles/6604027.

[315] J. Maher and E. T. Davies. Targeting cytotoxic T lymphocytes for cancer immunotherapy. *British Journal of Cancer*, 91(5):817–821, August 2004. ISSN 0007-0920. doi: 10.1038/sj.bjc.6602022.

[316] Gregory E. Idos, Janet Kwok, Nirupama Bonthala, Lynn Kysh, Stephen B. Gruber, and Chenxu Qu. The Prognostic Implications of Tumor Infiltrating Lymphocytes in

Colorectal Cancer: A Systematic Review and Meta-Analysis. *Scientific Reports*, 10 (1):3360, December 2020. ISSN 2045-2322. doi: 10.1038/s41598-020-60255-4. URL http://www.nature.com/articles/s41598-020-60255-4.

[317] Y. Naito, K. Saito, K. Shiiba, A. Ohuchi, K. Saigenji, H. Nagura, and H. Ohtani. CD8+ T cells infiltrated within cancer cell nests as a prognostic factor in human colorectal cancer. *Cancer Research*, 58(16):3491–3494, August 1998. ISSN 0008-5472.

[318] M. Guidoboni, R. Gafà, A. Viel, C. Doglioni, A. Russo, A. Santini, L. Del Tin, E. Macrì, G. Lanza, M. Boiocchi, and R. Dolcetti. Microsatellite instability and high content of activated cytotoxic lymphocytes identify colon cancer patients with a favorable prognosis. *The American Journal of Pathology*, 159(1):297–304, July 2001. ISSN 0002-9440. doi: 10.1016/S0002-9440(10)61695-1.

[319] T. Chiba, H. Ohtani, T. Mizoi, Y. Naito, E. Sato, H. Nagura, A. Ohuchi, K. Ohuchi, K. Shiiba, Y. Kurokawa, and S. Satomi. Intraepithelial CD8+ T-cell-count becomes a prognostic factor after a longer follow-up period in human colorectal carcinoma: possible association with suppression of micrometastasis. *British Journal of Cancer*, 91(9):1711–1717, November 2004. ISSN 0007-0920. doi: 10.1038/sj.bjc.6602201.

[320] Friedrich Prall, Thomas Dührkop, Volker Weirich, Christiane Ostwald, Peter Lenz, Horst Nizze, and Malte Barten. Prognostic role of CD8+ tumor-infiltrating lymphocytes in stage III colorectal cancer with and without microsatellite instability. *Human Pathology*, 35(7):808–816, July 2004. ISSN 0046-8177. doi: 10.1016/j.humpath.2004.01.022.

[321] Vanessa Deschoolmeester, Marc Baay, Eric Van Marck, Joost Weyler, Peter Vermeulen, Filip Lardon, and Jan B. Vermorken. Tumor infiltrating lymphocytes: an intriguing player in the survival of colorectal cancer patients. *BMC immunology*, 11: 19, April 2010. ISSN 1471-2172. doi: 10.1186/1471-2172-11-19.

[322] Katsuhiko Nosho, Yoshifumi Baba, Noriko Tanaka, Kaori Shima, Marika Hayashi, Jeffrey A. Meyerhardt, Edward Giovannucci, Glenn Dranoff, Charles S. Fuchs, and Shuji Ogino. Tumour-infiltrating T-cell subsets, molecular changes in colorectal cancer, and prognosis: cohort study and literature review. *The Journal of Pathology*, 222(4):350–366, December 2010. ISSN 1096-9896. doi: 10.1002/path.2774.

[323] C. H. Richards, K. M. Flegg, C. S. D. Roxburgh, J. J. Going, Z. Mohammed, P. G. Horgan, and D. C. McMillan. The relationships between cellular components of the peritumoural inflammatory response, clinicopathological characteristics and survival in patients with primary operable colorectal cancer. *British Journal of Cancer*, 106 (12):2010–2015, June 2012. ISSN 1532-1827. doi: 10.1038/bjc.2012.211.

[324] Harry H. Yoon, Jared M. Orrock, Nathan R. Foster, Daniel J. Sargent, Thomas C. Smyrk, and Frank A. Sinicrope. Prognostic impact of FoxP3+ regulatory T cells in relation to CD8+ T lymphocyte density in human colon carcinomas. *PloS One*, 7 (8):e42274, 2012. ISSN 1932-6203. doi: 10.1371/journal.pone.0042274.

[325] A. Ling, S. Edin, M. L. Wikberg, Å Öberg, and R. Palmqvist. The intratumoural subsite and relation of CD8(+) and FOXP3(+) T lymphocytes in colorectal cancer

provide important prognostic clues. *British Journal of Cancer*, 110(10):2551–2559, May 2014. ISSN 1532-1827. doi: 10.1038/bjc.2014.161.

[326] Colin H. Richards, Campbell S. D. Roxburgh, Arfon G. Powell, Alan K. Foulis, Paul G. Horgan, and Donald C. McMillan. The clinical utility of the local inflammatory response in colorectal cancer. *European Journal of Cancer (Oxford, England: 1990)*, 50(2):309–319, January 2014. ISSN 1879-0852. doi: 10.1016/j.ejca.2013.09.008.

[327] Koichiro Mori, Yuji Toiyama, Susumu Saigusa, Hiroyuki Fujikawa, Junichiro Hiro, Minako Kobayashi, Masaki Ohi, Toshimitsu Araki, Yasuhiro Inoue, Koji Tanaka, Yasuhiko Mohri, and Masato Kusunoki. Systemic Analysis of Predictive Biomarkers for Recurrence in Colorectal Cancer Patients Treated with Curative Surgery. *Digestive Diseases and Sciences*, 60(8):2477–2487, August 2015. ISSN 1573-2568. doi: 10.1007/s10620-015-3648-2.

[328] Christoph Reissfelder, Slava Stamova, Christina Gossmann, Marion Braun, Andreas Bonertz, Ute Walliczek, Mario Grimm, Nuh N. Rahbari, Moritz Koch, Maral Saadati, Axel Benner, Markus W. Büchler, Dirk Jäger, Niels Halama, Khashayarsha Khazaie, Jürgen Weitz, and Philipp Beckhove. Tumor-specific cytotoxic T lymphocyte activity determines colorectal cancer patient prognosis. *The Journal of Clinical Investigation*, 125(2):739–751, February 2015. ISSN 1558-8238. doi: 10.1172/JCI74894.

[329] Yufeng Chen, Ruixue Yuan, Xianrui Wu, Xiaosheng He, Yang Zeng, Xinjuan Fan, Lei Wang, Jianping Wang, Ping Lan, and Xiaojian Wu. A Novel Immune Marker Model Predicts Oncological Outcomes of Patients with Colorectal Cancer. *Annals of Surgical Oncology*, 23(3):826–832, March 2016. ISSN 1534-4681. doi: 10.1245/s10434-015-4889-1.

[330] Jonna Berntsson, Maria C Svensson, Karin Leandersson, Björn Nodin, Patrick Micke, Anna H Larsson, Jakob Eberhard, and Karin Jirström. The clinical impact of tumour-infiltrating lymphocytes in colorectal cancer differs by anatomical subsite: A cohort study: The clinical impact of tumour-infiltrating lymphocytes. *International Journal of Cancer*, 141(8):1654–1666, October 2017. ISSN 00207136. doi: 10.1002/ijc.30869. URL `http://doi.wiley.com/10.1002/ijc.30869`.

[331] Ann C. Eriksen, Flemming B. Sørensen, Jan Lindebjerg, Henrik Hager, René dePont Christensen, Sanne Kjær-Frifeldt, and Torben F. Hansen. The Prognostic Value of Tumor-Infiltrating lymphocytes in Stage II Colon Cancer. A Nationwide Population-Based Study. *Translational Oncology*, 11(4):979–987, August 2018. ISSN 19365233. doi: 10.1016/j.tranon.2018.03.008. URL `https://linkinghub.elsevier.com/retrieve/pii/S1936523318300640`.

[332] Michael S. Rooney, Sachet A. Shukla, Catherine J. Wu, Gad Getz, and Nir Hacohen. Molecular and Genetic Properties of Tumors Associated with Local Immune Cytolytic Activity. *Cell*, 160(1-2):48–61, January 2015. ISSN 00928674. doi: 10.1016/j.cell.2014.12.033. URL `https://linkinghub.elsevier.com/retrieve/pii/S0092867414016390`.

[333] Vésteinn Thorsson, David L. Gibbs, Scott D. Brown, Denise Wolf, Dante S. Bortone, Tai-Hsien Ou Yang, Eduard Porta-Pardo, Galen F. Gao, Christopher L. Plaisier, James A. Eddy, Elad Ziv, Aedin C. Culhane, Evan O. Paull, I.K. Ashok Sivakumar, Andrew J. Gentles, Raunaq Malhotra, Farshad Farshidfar, Antonio Colaprico, Joel S. Parker, Lisle E. Mose, Nam Sy Vo, Jianfang Liu, Yuexin Liu, Janet Rader, Varsha Dhankani, Sheila M. Reynolds, Reanne Bowlby, Andrea Califano, Andrew D. Cherniack, Dimitris Anastassiou, Davide Bedognetti, Younes Mokrab, Aaron M. Newman, Arvind Rao, Ken Chen, Alexander Krasnitz, Hai Hu, Tathiane M. Malta, Houtan Noushmehr, Chandra Sekhar Pedamallu, Susan Bullman, Akinyemi I. Ojesina, Andrew Lamb, Wanding Zhou, Hui Shen, Toni K. Choueiri, John N. Weinstein, Justin Guinney, Joel Saltz, Robert A. Holt, Charles S. Rabkin, Alexander J. Lazar, Jonathan S. Serody, Elizabeth G. Demicco, Mary L. Disis, Benjamin G. Vincent, Ilya Shmulevich, Samantha J. Caesar-Johnson, John A. Demchok, Ina Felau, Melpomeni Kasapi, Martin L. Ferguson, Carolyn M. Hutter, Heidi J. Sofia, Roy Tarnuzzer, Zhining Wang, Liming Yang, Jean C. Zenklusen, Jiashan (Julia) Zhang, Sudha Chudamani, Jia Liu, Laxmi Lolla, Rashi Naresh, Todd Pihl, Qiang Sun, Yunhu Wan, Ye Wu, Juok Cho, Timothy DeFreitas, Scott Frazer, Nils Gehlenborg, Gad Getz, David I. Heiman, Jaegil Kim, Michael S. Lawrence, Pei Lin, Sam Meier, Michael S. Noble, Gordon Saksena, Doug Voet, Hailei Zhang, Brady Bernard, Nyasha Chambwe, Varsha Dhankani, Theo Knijnenburg, Roger Kramer, Kalle Leinonen, Yuexin Liu, Michael Miller, Sheila Reynolds, Ilya Shmulevich, Vesteinn Thorsson, Wei Zhang, Rehan Akbani, Bradley M. Broom, Apurva M. Hegde, Zhenlin Ju, Rupa S. Kanchi, Anil Korkut, Jun Li, Han Liang, Shiyun Ling, Wenbin Liu, Yiling Lu, Gordon B. Mills, Kwok-Shing Ng, Arvind Rao, Michael Ryan, Jing Wang, John N. Weinstein, Jiexin Zhang, Adam Abeshouse, Joshua Armenia, Debyani Chakravarty, Walid K. Chatila, Ino de Bruijn, Jianjiong Gao, Benjamin E. Gross, Zachary J. Heins, Ritika Kundra, Konnor La, Marc Ladanyi, Augustin Luna, Moriah G. Nissan, Angelica Ochoa, Sarah M. Phillips, Ed Reznik, Francisco Sanchez-Vega, Chris Sander, Nikolaus Schultz, Robert Sheridan, S. Onur Sumer, Yichao Sun, Barry S. Taylor, Jioajiao Wang, Hongxin Zhang, Pavana Anur, Myron Peto, Paul Spellman, Christopher Benz, Joshua M. Stuart, Christopher K. Wong, Christina Yau, D. Neil Hayes, Joel S. Parker, Matthew D. Wilkerson, Adrian Ally, Miruna Balasundaram, Reanne Bowlby, Denise Brooks, Rebecca Carlsen, Eric Chuah, Noreen Dhalla, Robert Holt, Steven J.M. Jones, Katayoon Kasaian, Darlene Lee, Yussanne Ma, Marco A. Marra, Michael Mayo, Richard A. Moore, Andrew J. Mungall, Karen Mungall, A. Gordon Robertson, Sara Sadeghi, Jacqueline E. Schein, Payal Sipahimalani, Angela Tam, Nina Thiessen, Kane Tse, Tina Wong, Ashton C. Berger, Rameen Beroukhim, Andrew D. Cherniack, Carrie Cibulskis, Stacey B. Gabriel, Galen F. Gao, Gavin Ha, Matthew Meyerson, Steven E. Schumacher, Juliann Shih, Melanie H. Kucherlapati, Raju S. Kucherlapati, Stephen Baylin, Leslie Cope, Ludmila Danilova, Moiz S. Bootwalla, Phillip H. Lai, Dennis T. Maglinte, David J. Van Den Berg, Daniel J. Weisenberger, J. Todd Auman, Saianand Balu, Tom Bodenheimer, Cheng Fan, Katherine A. Hoadley, Alan P.

Hoyle, Stuart R. Jefferys, Corbin D. Jones, Shaowu Meng, Piotr A. Mieczkowski, Lisle E. Mose, Amy H. Perou, Charles M. Perou, Jeffrey Roach, Yan Shi, Janae V. Simons, Tara Skelly, Matthew G. Soloway, Donghui Tan, Umadevi Veluvolu, Huihui Fan, Toshinori Hinoue, Peter W. Laird, Hui Shen, Wanding Zhou, Michelle Bellair, Kyle Chang, Kyle Covington, Chad J. Creighton, Huyen Dinh, HarshaVardhan Doddapaneni, Lawrence A. Donehower, Jennifer Drummond, Richard A. Gibbs, Robert Glenn, Walker Hale, Yi Han, Jianhong Hu, Viktoriya Korchina, Sandra Lee, Lora Lewis, Wei Li, Xiuping Liu, Margaret Morgan, Donna Morton, Donna Muzny, Jireh Santibanez, Margi Sheth, Eve Shinbrot, Linghua Wang, Min Wang, David A. Wheeler, Liu Xi, Fengmei Zhao, Julian Hess, Elizabeth L. Appelbaum, Matthew Bailey, Matthew G. Cordes, Li Ding, Catrina C. Fronick, Lucinda A. Fulton, Robert S. Fulton, Cyriac Kandoth, Elaine R. Mardis, Michael D. McLellan, Christopher A. Miller, Heather K. Schmidt, Richard K. Wilson, Daniel Crain, Erin Curley, Johanna Gardner, Kevin Lau, David Mallery, Scott Morris, Joseph Paulauskis, Robert Penny, Candace Shelton, Troy Shelton, Mark Sherman, Eric Thompson, Peggy Yena, Jay Bowen, Julie M. Gastier-Foster, Mark Gerken, Kristen M. Leraas, Tara M. Lichtenberg, Nilsa C. Ramirez, Lisa Wise, Erik Zmuda, Niall Corcoran, Tony Costello, Christopher Hovens, Andre L. Carvalho, Ana C. de Carvalho, José H. Fregnani, Adhemar Longatto-Filho, Rui M. Reis, Cristovam Scapulatempo-Neto, Henrique C.S. Silveira, Daniel O. Vidal, Andrew Burnette, Jennifer Eschbacher, Beth Hermes, Ardene Noss, Rosy Singh, Matthew L. Anderson, Patricia D. Castro, Michael Ittmann, David Huntsman, Bernard Kohl, Xuan Le, Richard Thorp, Chris Andry, Elizabeth R. Duffy, Vladimir Lyadov, Oxana Paklina, Galiya Setdikova, Alexey Shabunin, Mikhail Tavobilov, Christopher McPherson, Ronald Warnick, Ross Berkowitz, Daniel Cramer, Colleen Feltmate, Neil Horowitz, Adam Kibel, Michael Muto, Chandrajit P. Raut, Andrei Malykh, Jill S. Barnholtz-Sloan, Wendi Barrett, Karen Devine, Jordonna Fulop, Quinn T. Ostrom, Kristen Shimmel, Yingli Wolinsky, Andrew E. Sloan, Agostino De Rose, Felice Giuliante, Marc Goodman, Beth Y. Karlan, Curt H. Hagedorn, John Eckman, Jodi Harr, Jerome Myers, Kelinda Tucker, Leigh Anne Zach, Brenda Deyarmin, Hai Hu, Leonid Kvecher, Caroline Larson, Richard J. Mural, Stella Somiari, Ales Vicha, Tomas Zelinka, Joseph Bennett, Mary Iacocca, Brenda Rabeno, Patricia Swanson, Mathieu Latour, Louis Lacombe, Bernard Têtu, Alain Bergeron, Mary McGraw, Susan M. Staugaitis, John Chabot, Hanina Hibshoosh, Antonia Sepulveda, Tao Su, Timothy Wang, Olga Potapova, Olga Voronina, Laurence Desjardins, Odette Mariani, Sergio Roman-Roman, Xavier Sastre, Marc-Henri Stern, Feixiong Cheng, Sabina Signoretti, Andrew Berchuck, Darell Bigner, Eric Lipp, Jeffrey Marks, Shannon McCall, Roger McLendon, Angeles Secord, Alexis Sharp, Madhusmita Behera, Daniel J. Brat, Amy Chen, Keith Delman, Seth Force, Fadlo Khuri, Kelly Magliocca, Shishir Maithel, Jeffrey J. Olson, Taofeek Owonikoko, Alan Pickens, Suresh Ramalingam, Dong M. Shin, Gabriel Sica, Erwin G. Van Meir, Hongzheng Zhang, Wil Eijckenboom, Ad Gillis, Esther Korpershoek, Leendert Looijenga, Wolter Oosterhuis, Hans Stoop, Kim E. van Kessel, Ellen C. Zwarthoff, Chiara Calatozzolo, Lucia

Cuppini, Stefania Cuzzubbo, Francesco DiMeco, Gaetano Finocchiaro, Luca Mattei, Alessandro Perin, Bianca Pollo, Chu Chen, John Houck, Pawadee Lohavanichbutr, Arndt Hartmann, Christine Stoehr, Robert Stoehr, Helge Taubert, Sven Wach, Bernd Wullich, Witold Kycler, Dawid Murawa, Maciej Wiznerowicz, Ki Chung, W. Jeffrey Edenfield, Julie Martin, Eric Baudin, Glenn Bubley, Raphael Bueno, Assunta De Rienzo, William G. Richards, Steven Kalkanis, Tom Mikkelsen, Houtan Noushmehr, Lisa Scarpace, Nicolas Girard, Marta Aymerich, Elias Campo, Eva Giné, Armando López Guillermo, Nguyen Van Bang, Phan Thi Hanh, Bui Duc Phu, Yufang Tang, Howard Colman, Kimberley Evason, Peter R. Dottino, John A. Martignetti, Hani Gabra, Hartmut Juhl, Teniola Akeredolu, Serghei Stepa, Dave Hoon, Keunsoo Ahn, Koo Jeong Kang, Felix Beuschlein, Anne Breggia, Michael Birrer, Debra Bell, Mitesh Borad, Alan H. Bryce, Erik Castle, Vishal Chandan, John Cheville, John A. Copland, Michael Farnell, Thomas Flotte, Nasra Giama, Thai Ho, Michael Kendrick, Jean-Pierre Kocher, Karla Kopp, Catherine Moser, David Nagorney, Daniel O'Brien, Brian Patrick O'Neill, Tushar Patel, Gloria Petersen, Florencia Que, Michael Rivera, Lewis Roberts, Robert Smallridge, Thomas Smyrk, Melissa Stanton, R. Houston Thompson, Michael Torbenson, Ju Dong Yang, Lizhi Zhang, Fadi Brimo, Jaffer A. Ajani, Ana Maria Angulo Gonzalez, Carmen Behrens, Jolanta Bondaruk, Russell Broaddus, Bogdan Czerniak, Bita Esmaeli, Junya Fujimoto, Jeffrey Gershenwald, Charles Guo, Alexander J. Lazar, Christopher Logothetis, Funda Meric-Bernstam, Cesar Moran, Lois Ramondetta, David Rice, Anil Sood, Pheroze Tamboli, Timothy Thompson, Patricia Troncoso, Anne Tsao, Ignacio Wistuba, Candace Carter, Lauren Haydu, Peter Hersey, Valerie Jakrot, Hojabr Kakavand, Richard Kefford, Kenneth Lee, Georgina Long, Graham Mann, Michael Quinn, Robyn Saw, Richard Scolyer, Kerwin Shannon, Andrew Spillane, onathan Stretch, Maria Synott, John Thompson, James Wilmott, Hikmat Al-Ahmadie, Timothy A. Chan, Ronald Ghossein, Anuradha Gopalan, Douglas A. Levine, Victor Reuter, Samuel Singer, Bhuvanesh Singh, Nguyen Viet Tien, Thomas Broudy, Cyrus Mirsaidi, Praveen Nair, Paul Drwiega, Judy Miller, Jennifer Smith, Howard Zaren, Joong-Won Park, Nguyen Phi Hung, Electron Kebebew, W. Marston Linehan, Adam R. Metwalli, Karel Pacak, Peter A. Pinto, Mark Schiffman, Laura S. Schmidt, Cathy D. Vocke, Nicolas Wentzensen, Robert Worrell, Hannah Yang, Marc Moncrieff, Chandra Goparaju, Jonathan Melamed, Harvey Pass, Natalia Botnariuc, Irina Caraman, Mircea Cernat, Inga Chemencedji, Adrian Clipca, Serghei Doruc, Ghenadie Gorincioi, Sergiu Mura, Maria Pirtac, Irina Stancul, Diana Tcaciuc, Monique Albert, Iakovina Alexopoulou, Angel Arnaout, John Bartlett, Jay Engel, Sebastien Gilbert, Jeremy Parfitt, Harman Sekhon, George Thomas, Doris M. Rassl, Robert C. Rintoul, Carlo Bifulco, Raina Tamakawa, Walter Urba, Nicholas Hayward, Henri Timmers, Anna Antenucci, Francesco Facciolo, Gianluca Grazi, Mirella Marino, Roberta Merola, Ronald de Krijger, Anne-Paule Gimenez-Roqueplo, Alain Piché, Simone Chevalier, Ginette McKercher, Kivanc Birsoy, Gene Barnett, Cathy Brewer, Carol Farver, Theresa Naska, Nathan A. Pennell, Daniel Raymond, Cathy Schilero, Kathy Smolenski, Felicia Williams, Carl Morrison, Jeffrey A. Borgia, Michael J. Liptay,

Mark Pool, Christopher W. Seder, Kerstin Junker, Larsson Omberg, Mikhail Dinkin, George Manikhas, Domenico Alvaro, Maria Consiglia Bragazzi, Vincenzo Cardinale, Guido Carpino, Eugenio Gaudio, David Chesla, Sandra Cottingham, Michael Dubina, Fedor Moiseenko, Renumathy Dhanasekaran, Karl-Friedrich Becker, Klaus-Peter Janssen, Julia Slotta-Huspenina, Mohamed H. Abdel-Rahman, Dina Aziz, Sue Bell, Colleen M. Cebulla, Amy Davis, Rebecca Duell, J. Bradley Elder, Joe Hilty, Bahavna Kumar, James Lang, Norman L. Lehman, Randy Mandt, Phuong Nguyen, Robert Pilarski, Karan Rai, Lynn Schoenfield, Kelly Senecal, Paul Wakely, Paul Hansen, Ronald Lechan, James Powers, Arthur Tischler, William E. Grizzle, Katherine C. Sexton, Alison Kastl, Joel Henderson, Sima Porten, Jens Waldmann, Martin Fassnacht, Sylvia L. Asa, Dirk Schadendorf, Marta Couce, Markus Graefen, Hartwig Huland, Guido Sauter, Thorsten Schlomm, Ronald Simon, Pierre Tennstedt, Oluwole Olabode, Mark Nelson, Oliver Bathe, Peter R. Carroll, June M. Chan, Philip Disaia, Pat Glenn, Robin K. Kelley, Charles N. Landen, Joanna Phillips, Michael Prados, Jeffry Simko, Karen Smith-McCune, Scott VandenBerg, Kevin Roggin, Ashley Fehrenbach, Ady Kendler, Suzanne Sifri, Ruth Steele, Antonio Jimeno, Francis Carey, Ian Forgie, Massimo Mannelli, Michael Carney, Brenda Hernandez, Benito Campos, Christel Herold-Mende, Christin Jungk, Andreas Unterberg, Andreas von Deimling, Aaron Bossler, Joseph Galbraith, Laura Jacobus, Michael Knudson, Tina Knutson, Deqin Ma, Mohammed Milhem, Rita Sigmund, Andrew K. Godwin, Rashna Madan, Howard G. Rosenthal, Clement Adebamowo, Sally N. Adebamowo, Alex Boussioutas, David Beer, Thomas Giordano, Anne-Marie Mes-Masson, Fred Saad, Therese Bocklage, Lisa Landrum, Robert Mannel, Kathleen Moore, Katherine Moxley, Russel Postier, Joan Walker, Rosemary Zuna, Michael Feldman, Federico Valdivieso, Rajiv Dhir, James Luketich, Edna M. Mora Pinero, Mario Quintero-Aguilo, Carlos Gilberto Carlotti, Jose Sebastião Dos Santos, Rafael Kemp, Ajith Sankarankuty, Daniela Tirapelli, James Catto, Kathy Agnew, Elizabeth Swisher, Jenette Creaney, Bruce Robinson, Carl Simon Shelley, Eryn M. Godwin, Sara Kendall, Cassaundra Shipman, Carol Bradford, Thomas Carey, Andrea Haddad, Jeffey Moyer, Lisa Peterson, Mark Prince, Laura Rozek, Gregory Wolf, Rayleen Bowman, Kwun M. Fong, Ian Yang, Robert Korst, W. Kimryn Rathmell, J. Leigh Fantacone-Campbell, Jeffrey A. Hooke, Albert J. Kovatich, Craig D. Shriver, John DiPersio, Bettina Drake, Ramaswamy Govindan, Sharon Heath, Timothy Ley, Brian Van Tine, Peter Westervelt, Mark A. Rubin, Jung Il Lee, Natália D. Aredes, and Armaz Mariamidze. The Immune Landscape of Cancer. *Immunity*, 48 (4):812–830.e14, April 2018. ISSN 10747613. doi: 10.1016/j.immuni.2018.03.023. URL `https://linkinghub.elsevier.com/retrieve/pii/S1074761318301213`.

[334] Markus W. Löffler, Daniel J. Kowalewski, Linus Backert, Jörg Bernhardt, Patrick Adam, Heiko Schuster, Florian Dengler, Daniel Backes, Hans-Georg Kopp, Stefan Beckert, Silvia Wagner, Ingmar Königsrainer, Oliver Kohlbacher, Lothar Kanz, Alfred Königsrainer, Hans-Georg Rammensee, Stefan Stevanović, and Sebastian P. Haen. Mapping the HLA Ligandome of Colorectal Cancer Reveals an Imprint of Malignant Cell Transformation. *Cancer Research*, 78(16):4627–4641, August 2018.

ISSN 0008-5472, 1538-7445. doi: 10.1158/0008-5472.CAN-17-1745. URL http://cancerres.aacrjournals.org/lookup/doi/10.1158/0008-5472.CAN-17-1745.

[335] Bianca Heemskerk, Pia Kvistborg, and Ton N. M. Schumacher. The cancer antigenome. *The EMBO journal*, 32(2):194–203, January 2013. ISSN 1460-2075. doi: 10.1038/emboj.2012.333.

[336] Ton N. Schumacher and Robert D. Schreiber. Neoantigens in cancer immunotherapy. *Science*, 348(6230):69–74, April 2015. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.aaa4971. URL https://www.sciencemag.org/lookup/doi/10.1126/science.aaa4971.

[337] Yong-Chen Lu and Paul F. Robbins. Cancer immunotherapy targeting neoantigens. *Seminars in Immunology*, 28(1):22–27, February 2016. ISSN 10445323. doi: 10.1016/j.smim.2015.11.002. URL https://linkinghub.elsevier.com/retrieve/pii/S1044532315000731.

[338] A. E. Kennedy, U. Ozbek, and M. T. Dorak. What has GWAS done for HLA and disease associations? *International Journal of Immunogenetics*, 44(5):195–211, October 2017. ISSN 17443121. doi: 10.1111/iji.12332. URL http://doi.wiley.com/10.1111/iji.12332.

[339] Jacques Neefjes, Marlieke L. M. Jongsma, Petra Paul, and Oddmund Bakke. Towards a systems understanding of MHC class I and MHC class II antigen presentation. *Nature Reviews Immunology*, 11(12):823–836, December 2011. ISSN 1474-1733, 1474-1741. doi: 10.1038/nri3084. URL http://www.nature.com/articles/nri3084.

[340] J. A. Traherne. Human MHC architecture and evolution: implications for disease association studies. *International Journal of Immunogenetics*, 35(3):179–192, June 2008. ISSN 1744-3121, 1744-313X. doi: 10.1111/j.1744-313X.2008.00765.x. URL http://doi.wiley.com/10.1111/j.1744-313X.2008.00765.x.

[341] Takashi Shiina, Kazuyoshi Hosomichi, Hidetoshi Inoko, and Jerzy K Kulski. The HLA genomic loci map: expression, interaction, diversity and disease. *Journal of Human Genetics*, 54(1):15–39, January 2009. ISSN 1434-5161, 1435-232X. doi: 10.1038/jhg.2008.5. URL http://www.nature.com/articles/jhg20085.

[342] Sebastian Boegel, Martin Löwer, Thomas Bukur, Patrick Sorn, John C. Castle, and Ugur Sahin. HLA and proteasome expression body map. *BMC Medical Genomics*, 11(1):36, December 2018. ISSN 1755-8794. doi: 10.1186/s12920-018-0354-x. URL https://bmcmedgenomics.biomedcentral.com/articles/10.1186/s12920-018-0354-x.

[343] Christian Kurts, Bruce W. S. Robinson, and Percy A. Knolle. Cross-priming in health and disease. *Nature Reviews. Immunology*, 10(6):403–414, June 2010. ISSN 1474-1741. doi: 10.1038/nri2780.

[344] Victoria L. Crotzer and Janice S. Blum. Autophagy and adaptive immunity. *Immunology*, 131(1):9–17, September 2010. ISSN 1365-2567. doi: 10.1111/j.1365-2567.2010.03321.x.

[345] John Trowsdale. HLA genomics in the third millennium. *Current Opinion in Immunology*, 17(5):498–504, October 2005. ISSN 0952-7915. doi: 10.1016/j.coi.2005.07.015.

[346] James Robinson, Dominic J Barker, Xenia Georgiou, Michael A Cooper, Paul Flicek, and Steven G E Marsh. IPD-IMGT/HLA Database. *Nucleic Acids Research*, page gkz950, October 2019. ISSN 0305-1048, 1362-4962. doi: 10.1093/nar/gkz950. URL https://academic.oup.com/nar/advance-article/doi/10.1093/nar/gkz950/5610347.

[347] P. I. W. de Bakker and S. Raychaudhuri. Interrogating the major histocompatibility complex with high-throughput genomics. *Human Molecular Genetics*, 21(R1):R29–R36, October 2012. ISSN 0964-6906, 1460-2083. doi: 10.1093/hmg/dds384. URL https://academic.oup.com/hmg/article-lookup/doi/10.1093/hmg/dds384.

[348] Dengbo Ji, Haizhao Yi, Dakui Zhang, Tiancheng Zhan, Zhaowei Li, Ming Li, Jinying Jia, Meng Qiao, Jinhong Xia, Zhiwei Zhai, Can Song, and Jin Gu. Somatic Mutations and Immune Alternation in Rectal Cancer Following Neoadjuvant Chemoradiotherapy. *Cancer Immunology Research*, 6(11):1401–1416, November 2018. ISSN 2326-6074. doi: 10.1158/2326-6066.CIR-17-0630.

[349] Sachet A Shukla, Michael S Rooney, Mohini Rajasagi, Grace Tiao, Philip M Dixon, Michael S Lawrence, Jonathan Stevens, William J Lane, Jamie L Dellagatta, Scott Steelman, Carrie Sougnez, Kristian Cibulskis, Adam Kiezun, Nir Hacohen, Vladimir Brusic, Catherine J Wu, and Gad Getz. Comprehensive analysis of cancer-associated somatic mutations in class I HLA genes. *Nature Biotechnology*, 33(11):1152–1158, November 2015. ISSN 1087-0156, 1546-1696. doi: 10.1038/nbt.3344. URL http://www.nature.com/articles/nbt.3344.

[350] Marios Giannakis, Xinmeng Jasmine Mu, Sachet A. Shukla, Zhi Rong Qian, Ofir Cohen, Reiko Nishihara, Samira Bahl, Yin Cao, Ali Amin-Mansour, Mai Yamauchi, Yasutaka Sukawa, Chip Stewart, Mara Rosenberg, Kosuke Mima, Kentaro Inamura, Katsuhiko Nosho, Jonathan A. Nowak, Michael S. Lawrence, Edward L. Giovannucci, Andrew T. Chan, Kimmie Ng, Jeffrey A. Meyerhardt, Eliezer M. Van Allen, Gad Getz, Stacey B. Gabriel, Eric S. Lander, Catherine J. Wu, Charles S. Fuchs, Shuji Ogino, and Levi A. Garraway. Genomic Correlates of Immune-Cell Infiltrates in Colorectal Carcinoma. *Cell Reports*, 15(4):857–865, April 2016. ISSN 2211-1247. doi: 10.1016/j.celrep.2016.03.075.

[351] Nicholas McGranahan, Rachel Rosenthal, Crispin T. Hiley, Andrew J. Rowan, Thomas B.K. Watkins, Gareth A. Wilson, Nicolai J. Birkbak, Selvaraju Veeriah, Peter Van Loo, Javier Herrero, Charles Swanton, Charles Swanton, Mariam Jamal-Hanjani, Selvaraju Veeriah, Seema Shafi, Justyna Czyzewska-Khan, Diana Johnson, Joanne Laycock, Leticia Bosshard-Carter, Rachel Rosenthal, Pat Gorman, Robert E. Hynds, Gareth Wilson, Nicolai J. Birkbak, Thomas B.K. Watkins, Nicholas McGranahan, Stuart Horswell, Richard Mitter, Mickael Escudero, Aengus Stewart, Peter Van Loo, Andrew Rowan, Hang Xu, Samra Turajlic, Crispin Hiley, Christopher Abbosh, Jacki Goldman, Richard Kevin Stone, Tamara Denner, Nik Matthews, Greg Elgar, Sophia Ward, Marta Costa, Sharmin Begum, Ben Phillimore, Tim Chambers,

Emma Nye, Sofia Graca, Maise Al Bakir, Kroopa Joshi, Andrew Furness, Assma Ben Aissa, Yien Ning Sophia Wong, Andy Georgiou, Sergio Quezada, John A. Hartley, Helen L. Lowe, Javier Herrero, David Lawrence, Martin Hayward, Nikolaos Panagiotopoulos, Shyam Kolvekar, Mary Falzon, Elaine Borg, Teresa Marafioti, Celia Simeon, Gemma Hector, Amy Smith, Marie Aranda, Marco Novelli, Dahmane Oukrif, Sam M. Janes, Ricky Thakrar, Martin Forster, Tanya Ahmad, Siow Ming Lee, Dionysis Papadatos-Pastos, Dawn Carnell, Ruheena Mendes, Jeremy George, Neal Navani, Asia Ahmed, Magali Taylor, Junaid Choudhary, Yvonne Summers, Raffaele Califano, Paul Taylor, Rajesh Shah, Piotr Krysiak, Kendadai Rammohan, Eustace Fontaine, Richard Booton, Matthew Evison, Phil Crosbie, Stuart Moss, Faiza Idries, Leena Joseph, Paul Bishop, Anshuman Chaturved, Anne Marie Quinn, Helen Doran, Angela Leek, Phil Harrison, Katrina Moore, Rachael Waddington, Juliette Novasio, Fiona Blackhall, Jane Rogan, Elaine Smith, Caroline Dive, Jonathan Tugwood, Ged Brady, Dominic G. Rothwell, Francesca Chemi, Jackie Pierce, Sakshi Gulati, Babu Naidu, Gerald Langman, Simon Trotter, Mary Bellamy, Hollie Bancroft, Amy Kerr, Salma Kadiri, Joanne Webb, Gary Middleton, Madava Djearaman, Dean Fennell, Jacqui A. Shaw, John Le Quesne, David Moore, Apostolos Nakas, Sridhar Rathinam, William Monteiro, Hilary Marshall, Louise Nelson, Jonathan Bennett, Joan Riley, Lindsay Primrose, Luke Martinson, Girija Anand, Sajid Khan, Anita Amadi, Marianne Nicolson, Keith Kerr, Shirley Palmer, Hardy Remmen, Joy Miller, Keith Buchan, Mahendran Chetty, Lesley Gomersall, Jason Lester, Alison Edwards, Fiona Morgan, Haydn Adams, Helen Davies, Malgorzata Kornaszewska, Richard Attanoos, Sara Lock, Azmina Verjee, Mairead MacKenzie, Maggie Wilcox, Harriet Bell, Allan Hackshaw, Yenting Ngai, Sean Smith, Nicole Gower, Christian Ottensmeier, Serena Chee, Benjamin Johnson, Aiman Alzetani, Emily Shaw, Eric Lim, Paulo De Sousa, Monica Tavares Barbosa, Alex Bowman, Simon Jordan, Alexandra Rice, Hilgardt Raubenheimer, Chiara Proli, Maria Elena Cufari, John Carlo Ronquillo, Angela Kwayie, Harshil Bhayani, Morag Hamilton, Yusura Bakar, Natalie Mensah, Lyn Ambrose, Anand Devaraj, Silviu Buderi, Jonathan Finch, Leire Azcarate, Hema Chavan, Sophie Green, Hillaria Mashinga, Andrew G. Nicholson, Kelvin Lau, Michael Sheaff, Peter Schmid, John Conibear, Veni Ezhil, Babikir Ismail, Melanie Irvin-sellers, Vineet Prakash, Peter Russell, Teresa Light, Tracey Horey, Sarah Danson, Jonathan Bury, John Edwards, Jennifer Hill, Sue Matthews, Yota Kitsanta, Kim Suvarna, Patricia Fisher, Allah Dino Keerio, Michael Shackcloth, John Gosney, Pieter Postmus, Sarah Feeney, Julius Asante-Siaw, Hugo J.W.L. Aerts, Stefan Dentro, and Christophe Dessimoz. Allele-Specific HLA Loss and Immune Escape in Lung Cancer Evolution. *Cell*, 171(6): 1259–1271.e11, November 2017. ISSN 00928674. doi: 10.1016/j.cell.2017.10.001. URL `https://linkinghub.elsevier.com/retrieve/pii/S0092867417311856`.

[352] Diego Chowell, Luc G. T. Morris, Claud M. Grigg, Jeffrey K. Weber, Robert M. Samstein, Vladimir Makarov, Fengshen Kuo, Sviatoslav M. Kendall, David Requena, Nadeem Riaz, Benjamin Greenbaum, James Carroll, Edward Garon, David M. Hyman, Ahmet Zehir, David Solit, Michael Berger, Ruhong Zhou, Naiyer A.

Rizvi, and Timothy A. Chan. Patient HLA class I genotype influences cancer response to checkpoint blockade immunotherapy. *Science*, 359(6375):582–587, February 2018. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.aao4572. URL https://www.sciencemag.org/lookup/doi/10.1126/science.aao4572.

[353] Valeria Orrù, Maristella Steri, Gabriella Sole, Carlo Sidore, Francesca Virdis, Mariano Dei, Sandra Lai, Magdalena Zoledziewska, Fabio Busonero, Antonella Mulas, Matteo Floris, Wieslawa I. Mentzen, Silvana A.M. Urru, Stefania Olla, Michele Marongiu, Maria G. Piras, Monia Lobina, Andrea Maschio, Maristella Pitzalis, Maria F. Urru, Marco Marcelli, Roberto Cusano, Francesca Deidda, Valentina Serra, Manuela Oppo, Rosella Pilu, Frederic Reinier, Riccardo Berutti, Luca Pireddu, Ilenia Zara, Eleonora Porcu, Alan Kwong, Christine Brennan, Brendan Tarrier, Robert Lyons, Hyun M. Kang, Sergio Uzzau, Rossano Atzeni, Maria Valentini, Davide Firinu, Lidia Leoni, Gianluca Rotta, Silvia Naitza, Andrea Angius, Mauro Congia, Michael B. Whalen, Chris M. Jones, David Schlessinger, Gonçalo R. Abecasis, Edoardo Fiorillo, Serena Sanna, and Francesco Cucca. Genetic Variants Regulating Immune Cell Levels in Health and Disease. *Cell*, 155(1):242–256, September 2013. ISSN 00928674. doi: 10.1016/j.cell.2013.08.041. URL https://linkinghub.elsevier.com/retrieve/pii/S0092867413010726.

[354] Miles Parkes, Adrian Cortes, David A. van Heel, and Matthew A. Brown. Genetic insights into common pathways and complex relationships among immune-mediated diseases. *Nature Reviews Genetics*, 14(9):661–673, September 2013. ISSN 1471-0056, 1471-0064. doi: 10.1038/nrg3502. URL http://www.nature.com/articles/nrg3502.

[355] Mario Roederer, Lydia Quaye, Massimo Mangino, Margaret H. Beddall, Yolanda Mahnke, Pratip Chattopadhyay, Isabella Tosi, Luca Napolitano, Manuela Terranova Barberio, Cristina Menni, Federica Villanova, Paola Di Meglio, Tim D. Spector, and Frank O. Nestle. The Genetic Architecture of the Human Immune System: A Bioresource for Autoimmunity and Disease Pathogenesis. *Cell*, 161(2):387–403, April 2015. ISSN 00928674. doi: 10.1016/j.cell.2015.02.046. URL https://linkinghub.elsevier.com/retrieve/pii/S0092867415002470.

[356] M. Vogelsang, C. N. Martinez, J. Rendleman, A. Bapodra, K. Malecek, A. Romanchuk, E. Kazlow, R. L. Shapiro, R. S. Berman, M. Krogsgaard, I. Osman, and T. Kirchhoff. The Expression Quantitative Trait Loci in Immune Pathways and their Effect on Cutaneous Melanoma Prognosis. *Clinical Cancer Research*, 22(13):3268–3280, July 2016. ISSN 1078-0432, 1557-3265. doi: 10.1158/1078-0432.CCR-15-2066. URL http://clincancerres.aacrjournals.org/cgi/doi/10.1158/1078-0432.CCR-15-2066.

[357] Z. Fewell, G. Davey Smith, and J. A. C. Sterne. The Impact of Residual and Unmeasured Confounding in Epidemiologic Studies: A Simulation Study. *American Journal of Epidemiology*, 166(6):646–655, June 2007. ISSN 0002-9262, 1476-6256. doi: 10.1093/aje/kwm165. URL https://academic.oup.com/aje/article-lookup/doi/10.1093/aje/kwm165.

[358] D. L. Weed and L. S. Gorelic. The practice of causal inference in cancer epidemiology. *Cancer Epidemiology, Biomarkers & Prevention: A Publication of the American Association for Cancer Research, Cosponsored by the American Society of Preventive Oncology*, 5(4):303–311, April 1996. ISSN 1055-9965.

[359] James Yarmolinsky, Kaitlin H. Wade, Rebecca C. Richmond, Ryan J. Langdon, Caroline J. Bull, Kate M. Tilling, Caroline L. Relton, Sarah J. Lewis, George Davey Smith, and Richard M. Martin. Causal Inference in Cancer Epidemiology: What Is the Role of Mendelian Randomization? *Cancer Epidemiology, Biomarkers & Prevention: A Publication of the American Association for Cancer Research, Cosponsored by the American Society of Preventive Oncology*, 27(9): 995–1010, September 2018. ISSN 1538-7755. doi: 10.1158/1055-9965.EPI-17-1177.

[360] Nicholas J. Timpson, Kaitlin H. Wade, and George Davey Smith. Mendelian randomization: application to cardiovascular disease. *Current Hypertension Reports*, 14(1):29–37, February 2012. ISSN 1534-3111. doi: 10.1007/s11906-011-0242-7.

[361] Jakob Runge, Sebastian Bathiany, Erik Bollt, Gustau Camps-Valls, Dim Coumou, Ethan Deyle, Clark Glymour, Marlene Kretschmer, Miguel D. Mahecha, Jordi Muñoz-Marí, Egbert H. van Nes, Jonas Peters, Rick Quax, Markus Reichstein, Marten Scheffer, Bernhard Schölkopf, Peter Spirtes, George Sugihara, Jie Sun, Kun Zhang, and Jakob Zscheischler. Inferring causation from time series in Earth system sciences. *Nature Communications*, 10(1):2553, December 2019. ISSN 2041-1723. doi: 10.1038/s41467-019-10105-3. URL `http://www.nature.com/articles/s41467-019-10105-3`.

[362] Neil M Davies, Michael V Holmes, and George Davey Smith. Reading Mendelian randomisation studies: a guide, glossary, and checklist for clinicians. *BMJ*, page k601, July 2018. ISSN 0959-8138, 1756-1833. doi: 10.1136/bmj.k601. URL `http://www.bmj.com/lookup/doi/10.1136/bmj.k601`.

[363] Tom M Palmer, Debbie A Lawlor, Roger M Harbord, Nuala A Sheehan, Jon H Tobias, Nicholas J Timpson, George Davey Smith, and Jonathan AC Sterne. Using multiple genetic variants as instrumental variables for modifiable risk factors. *Statistical Methods in Medical Research*, 21(3):223–242, June 2012. ISSN 0962-2802, 1477-0334. doi: 10.1177/0962280210394459. URL `http://journals.sagepub.com/doi/10.1177/0962280210394459`.

[364] Stephen Burgess, Adam Butterworth, and Simon G. Thompson. Mendelian Randomization Analysis With Multiple Genetic Variants Using Summarized Data: Mendelian Randomization Using Summarized Data. *Genetic Epidemiology*, 37 (7):658–665, November 2013. ISSN 07410395. doi: 10.1002/gepi.21758. URL `http://doi.wiley.com/10.1002/gepi.21758`.

[365] Tomasa Barrientos, Derk Frank, Koichiro Kuwahara, Svetlana Bezprozvannaya, G. C. Teg Pipes, Rhonda Bassel-Duby, James A. Richardson, Hugo A. Katus, Eric N. Olson, and Norbert Frey. Two Novel Members of the ABLIM Protein Family, ABLIM-2 and -3, Associate with STARS and Directly Bind F-actin. *Journal of Biological Chemistry*, 282(11):8393–8403, March 2007. ISSN 00219258. doi:

10.1074/jbc.M607549200. URL https://linkinghub.elsevier.com/retrieve/pii/S0021925820638492.

[366] Jianbo Tian, Yimin Cai, Yue Li, Zequn Lu, Jinyu Huang, Yao Deng, Nan Yang, Xiaoyang Wang, Pingting Ying, Shanshan Zhang, Ying Zhu, Huilan Zhang, Rong Zhong, Jiang Chang, and Xiaoping Miao. CancerImmunityQTL: a database to systematically evaluate the impact of genetic variants on immune infiltration in human cancer. *Nucleic Acids Research*, 49(D1):D1065–D1073, January 2021. ISSN 0305-1048, 1362-4962. doi: 10.1093/nar/gkaa805. URL https://academic.oup.com/nar/article/49/D1/D1065/5917653.

[367] Wessel N. van Wieringen and Aad W. van der Vaart. Statistical analysis of the cancer cell's molecular entropy using high-throughput data. *Bioinformatics (Oxford, England)*, 27(4):556–563, February 2011. ISSN 1367-4811. doi: 10.1093/bioinformatics/btq704.

[368] Juan Zhao, Yiwei Zhou, Xiujun Zhang, and Luonan Chen. Part mutual information for quantifying direct associations in networks. *Proceedings of the National Academy of Sciences*, 113(18):5130–5135, May 2016. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.1522586113. URL http://www.pnas.org/lookup/doi/10.1073/pnas.1522586113.

[369] M. Kanehisa and S. Goto. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Research*, 28(1):27–30, January 2000. ISSN 0305-1048. doi: 10.1093/nar/28.1.27.

[370] Sung Rye Park, Sim Namkoong, Leon Friesen, Chun-Seok Cho, Zac Zezhi Zhang, Yu-Chih Chen, Euisik Yoon, Chang H. Kim, Hojoong Kwak, Hyun Min Kang, and Jun Hee Lee. Single-Cell Transcriptome Analysis of Colon Cancer Cell Response to 5-Fluorouracil-Induced DNA Damage. *Cell Reports*, 32(8):108077, August 2020. ISSN 22111247. doi: 10.1016/j.celrep.2020.108077. URL https://linkinghub.elsevier.com/retrieve/pii/S2211124720310627.

[371] Neeraj Lal, Brian S. White, Ghaleb Goussous, Oliver Pickles, Mike J. Mason, Andrew D. Beggs, Philippe Taniere, Benjamin E. Willcox, Justin Guinney, and Gary W. Middleton. KRAS Mutation and Consensus Molecular Subtypes 2 and 3 Are Independently Associated with Reduced Immune Infiltration and Reactivity in Colorectal Cancer. *Clinical Cancer Research: An Official Journal of the American Association for Cancer Research*, 24(1):224–233, January 2018. ISSN 1557-3265. doi: 10.1158/1078-0432.CCR-17-1090.

[372] Erkki-Ville Wirta, Toni Seppälä, Marjukka Friman, Juha Väyrynen, Maarit Ahtiainen, Hannu Kautiainen, Teijo Kuopio, Ilmo Kellokumpu, Jukka-Pekka Mecklin, and Jan Böhm. Immunoscore in mismatch repair-proficient and -deficient colon cancer: Immunoscore in colon cancer. *The Journal of Pathology: Clinical Research*, 3(3):203–213, July 2017. ISSN 20564538. doi: 10.1002/cjp2.71. URL http://doi.wiley.com/10.1002/cjp2.71.

[373] Michael J. Overman, Sara Lonardi, Ka Yeung Mark Wong, Heinz-Josef Lenz, Fabio Gelsomino, Massimo Aglietta, Michael A. Morse, Eric Van Cutsem, Ray McDermott,

Andrew Hill, Michael B. Sawyer, Alain Hendlisz, Bart Neyns, Magali Svrcek, Rebecca A. Moss, Jean-Marie Ledeine, Z. Alexander Cao, Shital Kamble, Scott Kopetz, and Thierry André. Durable Clinical Benefit With Nivolumab Plus Ipilimumab in DNA Mismatch Repair-Deficient/Microsatellite Instability-High Metastatic Colorectal Cancer. *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology*, 36(8):773–779, March 2018. ISSN 1527-7755. doi: 10.1200/JCO.2017.76.9901.

[374] Catherine S. Grasso, Marios Giannakis, Daniel K. Wells, Tsuyoshi Hamada, Xinmeng Jasmine Mu, Michael Quist, Jonathan A. Nowak, Reiko Nishihara, Zhi Rong Qian, Kentaro Inamura, Teppei Morikawa, Katsuhiko Nosho, Gabriel Abril-Rodriguez, Charles Connolly, Helena Escuin-Ordinas, Milan S. Geybels, William M. Grady, Li Hsu, Siwen Hu-Lieskovan, Jeroen R. Huyghe, Yeon Joo Kim, Paige Krystofinski, Mark D.M. Leiserson, Dennis J. Montoya, Brian B. Nadel, Matteo Pellegrini, Colin C. Pritchard, Cristina Puig-Saus, Elleanor H. Quist, Ben J. Raphael, Stephen J. Salipante, Daniel Sanghoon Shin, Eve Shinbrot, Brian Shirts, Sachet Shukla, Janet L. Stanford, Wei Sun, Jennifer Tsoi, Alexander Upfill-Brown, David A. Wheeler, Catherine J. Wu, Ming Yu, Syed H. Zaidi, Jesse M. Zaretsky, Stacey B. Gabriel, Eric S. Lander, Levi A. Garraway, Thomas J. Hudson, Charles S. Fuchs, Antoni Ribas, Shuji Ogino, and Ulrike Peters. Genetic Mechanisms of Immune Evasion in Colorectal Cancer. *Cancer Discovery*, 8(6):730–749, June 2018. ISSN 2159-8274, 2159-8290. doi: 10.1158/2159-8290.CD-17-1327. URL `http://cancerdiscovery.aacrjournals.org/lookup/doi/10.1158/2159-8290.CD-17-1327`.