MOLECULAR CHARACTERIZATION AND GENETIC DIAGNOSIS OF CANCER PREDISPOSITION SYNDROMES USING GENOME AND TRANSCRIPTOME SEQUENCING

by

Katherine Dixon

B.Sc., Biology, McGill University, 2014

M.Sc., Cellular and Molecular Medicine, University of Ottawa, 2016

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES

(Medical Genetics)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

August 2020

The following individuals certify that they have read, and recommend to the Faculty of Graduate and Postdoctoral Studies for acceptance, the dissertation entitled:

Molecular Characterization and Genetic Diagnosis of Cancer Predisposition Syndromes Using Genome and Transcriptome Sequencing

submitted by   Katherine Dixon   in partial fulfillment of the requirements for

the degree of   Doctor of Philosophy

in   Medical Genetics

Examining Committee:

Dr. Kasmintan A. Schrader, Medical Genetics

Co-supervisor

Dr. David G. Huntsman, Pathology and Laboratory Medicine

Co-supervisor

Dr. Jan M. Friedman, Medical Genetics

Supervisory Committee Member

Dr. Christopher Maxwell, Cell and Developmental Biology

University Examiner

Dr. William T. Gibson, Medical Genetics

University Examiner

Additional Supervisory Committee Members:

Dr. Angela Brooks-Wilson, Medical Genetics

Supervisory Committee Member

Dr. Wyeth Wasserman, Medical Genetics

Supervisory Committee Member

Abstract

Genetic variation makes important and often uncharacterized contributions to both rare syndromes and more common complex diseases. Around 8% of all cancers are caused by high- and moderate-penetrance deleterious germline variants that are present in an individual from birth and predispose to specific cancer types throughout life. Such cancer predisposition genes associated with moderate to high lifetime cancer risks, conventionally defined as greater than twofold and fivefold increases, respectively, are involved in various biological pathways required for regulating cellular proliferation, maintaining genome integrity, and mediating inter- and intracellular signaling. Clinical use of multigene next-generation sequencing panels has improved molecular diagnosis of cancer predisposition syndromes, demonstrating both genetic heterogeneity and phenotypic variability amongst carriers. However, many individuals with a strong personal or family cancer history receive uninformative results from clinical genetic testing that may lead to increased health anxiety and missed opportunities for increased cancer screening, cancer prevention or use of targeted therapies. Therefore, the main objective of my dissertation was to characterize the biological significance, functional impact, and heterogeneity of genetic variation underlying high-penetrance cancer predisposition syndromes to improve rates of genetic diagnosis. Using genome and transcriptome sequencing, I explored molecular characteristics associated with inactivation of high-penetrance cancer predisposition genes across advanced cancers and in an organoid model system. Tissue-specific molecular signatures provided insights into the aetiology of site-specific tumour development in carriers, allowing opportunities for carrier ascertainment and differential genetic diagnosis of suspected hereditary cancer families. Based on findings that structural variants account for 10% of causal variants, I

investigated the utility of long-read sequencing in the clinical interpretation of germline structural variants that were undetected or unresolved through next-generation sequencing. Short- and long-read genome sequencing improved genetic diagnosis in known and suspected carriers for autosomal dominant cancer syndromes, demonstrating incomplete penetrance and phenotypic heterogeneity in population-based and disease-specific cancer cohorts. The research presented here thus supports a broader understanding of the contributions of germline variation to cancer susceptibility and disease progression, ultimately informing guidelines for screening, variant interpretation, and clinical management.

Lay Summary

One in 400 individuals carries a harmful genetic variant in genes underlying inherited cancer predisposition, accounting for around 10% of all cancer cases. However, many of these individuals are not identified due to the absence of informative personal or family history. Even for individuals with suspected cancer predisposition syndromes, clinical genetic testing may not identify the causal genetic variant. Novel approaches to increasing the accessibility of genetic testing and to studying of the consequence of genetic variation may improve the identification and diagnosis of families who have a significant lifetime risk of cancer. These approaches include studying how different genes lead to specific types of cancer, integrating molecular tumour characteristics into genetic diagnosis, and describing the spectrum of clinical features in syndromes caused by rare genetic variants. Ultimately this research may allow opportunities for cancer prevention and early cancer detection in patients and their families.

Preface

All data chapters in this thesis (Chapters 2-5) are presented in manuscript format as they are currently published (Chapter 2) or under submission (Chapters 2-5).

Chapter 1 is original unpublished work, which I wrote.

Portions of **Chapter 2** have been adapted from a previously published manuscript:

- Pleasance E*, Titmuss E*, Williamson L*, Kwan H, Culibrk L, Zhao EY, Dixon K, Fan K, Bowlby R, Jones MR, Shen Y, Grewal J, Ashkani J, Wee K, Grisdale C, Thibodeau ML, Bozoky Z, Pearson H, Majounie E, Vira T, Shenwai R, Mungall K, Chuah E, Davies J, Warren M, Reisle C, Bonakdar M, Taylor GA, Csizmok V, Chan S, Zong S, Bilobram S, Zadeh A, D'Souza D, Corbett R, MacMillan D, Carreira M, Choo C, Bleile D, Sadeghi S, Zhang W, Wong T, Cheng D, Moore R, Mungall A, Zhao Y, Nelson J, Fok A, Roscoe R, Ma Y, Lee M, Lavoie JM, Karasinska J, Deol B, Fisic A, Schaeffer D, Yip S, Schrader K, Regier D, Chia S, Gelmon K, Tinker A, Sun S, Lim H, Renouf D, Jones SJM, Laskin J, Marra MA. Pan-cancer analysis of advanced patient tumors reveals interactions between therapy and genomic landscapes. *Nature Cancer*. 2020.

This research was approved by the BC Cancer Research Ethics Board (H12-00137, H14-00681, H16-00291). I performed a retrospective germline analysis, relevant mutational signature analysis, clinical chart review, and relevant manuscript preparation with assistance ML. Thibodeau.

Chapter 2 is also based in part on original unpublished work. This research was approved by the UBC Clinical Research Ethics Board (H19-02571). I performed the clinical chart review and data analysis.

Chapter 3 is based on original unpublished work. This research was approved by the UBC Clinical Research Ethics Board (H19-02520). Organoid development was performed by T. Brew and P. Guilford at the University of Otago (Otago, New Zealand). Single-cell RNA sequencing was performed at the University of British Columbia Biomedical Research Centre, and immunohistochemistry was performed at the Vancouver General Hospital and Genome Pathology Evaluation Centre (Vancouver, Canada). I collected and analyzed the clinical data, performed the bioinformatic analysis and scored immunohistochemistry slides with assistance from M. Bui, D. Farnell and A. El Naggar.

A version of **Chapter 4** has been published:

- Thibodeau ML*, O'Neill K*, Dixon K*, Reisle C, Mungall KL, Krzywinski M, Shen Y, Lim HJ, Cheng D, Tse K, Wong T, Chuah E, Fok A, Sun S, Renouf D, Schaeffer DF, Cremin C, Chia S, Young S, Pandoh P, Pleasance S, Pleasance E, Mungall AJ, Moore R, Yip S, Karsan A, Laskin J, Marra MA, Schrader KA, Jones SJM. Improving structural variant interpretation for hereditary cancer susceptibility using long-read sequencing. *Genetics in Medicine*. 2020.

  *ML. Thibodeau, K. O'Neill and I made equal contributions to data analysis, data interpretation and manuscript preparation.

Chapter 4 is also based in part on original unpublished work. This research was approved by the UBC Clinical Research Ethics Board (H16-02124). Next-generation and long-read sequencing were performed at Canada's Michael Smith Genome Sciences Centre (Vancouver, Canada). I analyzed the data and performed validations, with assistance from Y. Shen, J. Senz and A. Lum.

A portion of **Chapter 5** has been accepted for publication:

- Dixon K, Senz J, Kaurah P, Huntsman DG, Schrader KA. Rare *APC* promoter 1B variants in gastric cancer kindreds unselected for fundic gland polyposis. *Gut.*

This research was approved by the UBC Clinical Research Ethics Board (H02-61149). P. Kaurah, K. Schrader and D. Huntsman designed the study, J. Senz performed DNA extraction and sequencing, and I collected and analyzed the clinical data and wrote the manuscript.

Chapter 5 is also based in part on original unpublished work. This research was approved by the UBC Clinical Research Ethics Board (H17-01449) in collaboration with K. Calzone, L. Foretova, M. Iwatsuki, and C. Kiraly-Borri. I participated in questionnaire design, participant and clinician ascertainment and data analysis.

Chapter 6 is original unpublished work, which I wrote.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

ACMG	American College of Medical Genetics and Genomics
AJ	Ashkenazi Jewish
AMP	Association of Molecular Pathology
BER	base excision repair
BMP	bone morphogenic protein
BWA	Burrows Wheeler alignment
COSMIC	Catalogue of Somatic Mutations in Cancer
CRC	colorectal cancer
DBS	doublet base substitutions
DC	dyskeratosis congenita
DGC	diffuse gastric cancer
DNA	deoxyribonucleic acid
DSBR	double-stranded break repair
EC	endometrial cancer
ER	estrogen receptor
ES	exome sequencing
FAMMM	familial atypical multiple mole melanoma
FAMMM-PC	familial atypical multiple mole melanoma-pancreatic cancer
FAP	familial adenomatous polyposis
FDR	false discovery rate
FFPE	formalin-fixed paraffin-embedded

FIGC	familial intestinal gastric cancer
FPC	familial pancreatic cancer
GAPPS	gastric adenocarcinoma and proximal polyposis of the stomach
GS	genome sequencing
GWAS	genome-wide association studies
HBOC	hereditary breast and ovarian cancer
HDGC	hereditary diffuse gastric cancer
HER2	human epidermal growth factor receptor
HMPS	hereditary mixed polyposis syndrome
HR	homologous recombination
ICGC	International Cancer Genome Consortium
IGCLC	International Gastric Cancer Linkage Consortium
IGV	Integrated Genomics Viewer
IHC	immunohistochemistry
indel	small insertion and deletion
LINE	long interspersed nuclear element
LOH	loss of heterozygosity
MAP	*MUTYH*-associated polyposis
MLPA	multiplex ligation-dependent probe amplification
MMR	mismatch repair
MSI	microsatellite instability
NCCN	National Comprehensive Cancer Network
NER	nucleotide excision repair

NGS	next-generation sequencing
NHEJ	non-homologous end joining
NMD	nonsense-mediated decay
NSAID	non-steroidal anti-inflammatory drug
PI3K	phosphoinositide 3 kinase
POG	Personalized OncoGenomics
PR	progesterone receptor
RNA-seq	RNA sequencing
scRNA-seq	single-cell RNA sequencing
SINE	short interspersed nuclear element
SNP	single nucleotide polymorphism
SNV	single nucleotide variant
SRC	signet ring cell
SV	structural variant
TCGA	The Cancer Genome Atlas
TSC	tuberous sclerosis complex
UTR	untranslated region
VUS	variant of uncertain significance

Acknowledgements

To my supervisors, Drs. Kasmintan Schrader and David Huntsman, thank you for your exceptional mentorship and kindness. Working with you has been wonderful and has provided both challenging and rewarding opportunities for both academic and personal growth. I would also like to thank members of my supervisory committee, Drs. Angela Brooks-Wilson, Jan Friedman and Wyeth Wasserman, and Cheryl Bishop for your ongoing support and guidance throughout my program.

This work would not have been possible without the guidance of Janine Senz and Pardeep Kaurah, who were deeply involved in my technical training and who provided endless support. Thank you to current and former members of the Huntsman lab, Hereditary Cancer Program and Personalized OncoGenomics program, with whom I had the pleasure of working and building lifetime friendships. I would especially like to thank Dr. Kieran O'Neill for bioinformatics support and Dr. My Linh Thibodeau for clinical expertise. Many thanks to our collaborators, Dr. Parry Guilford and Tom Brew, for including us in their important organoid work.

Finally, I would like to extend my deepest appreciation and love to my partner, family and best friends. Your unconditional support continues to be motivating and grounding, and this experience would not have been possible without you.

Dedication

For William and Don.

Chapter 1: Introduction

1.1 Cancer as an inherently genetic disease

Cancer is a complex and multifactorial disease resulting from the accumulation of genetic mutations over the lifetime of a cell. Uncontrolled proliferation, loss of growth inhibition, altered metabolism, and evasion of the immune system describe some of the most common hallmarks of cancer (Hanahan and Weinberg 2000; Hanahan and Weinberg 2011). Although most cancers arise sporadically, resulting from tissue-specific somatic mutations that occur after birth, familial clustering occurs in 15-20%. Cancer predisposition syndromes caused by moderate- to high-penetrance germline variants conferring a two- to fivefold and more than fivefold increase in cancer risk, respectively, account for an additional 5-10% of cases (Nagy et al. 2004; Tung et al. 2016). Carriers of pathogenic variants have an increased lifetime risk for certain cancer types that may be associated with an earlier age of onset than in sporadic cases of the same cancer type and syndrome-related non-cancer phenotypes. Therefore, identifying individuals harbouring pathogenic germline variants in cancer predisposition genes may allow the opportunity for predictive genetic testing in at-risk relatives and cancer prevention or early cancer detection through prophylactic surgery, chemoprevention or regular cancer screening.

1.1.1 Historical perspectives of cancer syndromes

Before the genetic basis of cancer was understood, early observations of familial clustering of site-specific cancers suggested a role for undefined inherited factors in cancer susceptibility. For example, French physician Paul Broca made one of the earliest reports of hereditary breast and ovarian cancer (HBOC), describing the pedigree of his wife's maternal

family in which 10 individuals across five generations had succumbed to breast cancer (Broca 1866). American physicians Aldred Scott Warthin and Henry Lynch similarly described families with Lynch syndrome beginning in the early twentieth century (Warthin 1913; Lynch et al. 1966; Lynch and Krush 1967; Lynch and Krush 1971). Family G was first reported by Warthin in 1913, who noted endometrial and gastrointestinal cancers in 18 affected individuals across three generations. Additional families showing similar clusters of endometrial and gastrointestinal cancers, particularly those of the colon, led to the subsequent characterization of Lynch syndrome, the most common form of hereditary colorectal cancer (CRC) (Boland and Lynch 2013). Although Lynch syndrome is primarily associated with a significantly increased lifetime risk of CRC and endometrial cancer (EC), it has also been associated with an increased risk for several other malignancies, including stomach, ovarian, pancreatic, ureter and renal pelvis, biliary tract, brain, sebaceous gland adenoma and keratoacanthoma, and small bowel cancer (Kohlmann and Gruber 2004).

Historical accounts of familial cancers indicated that cancer predisposition syndromes shared common characteristics, including multiple generations affected with the same or related cancer types, a high prevalence of early onset, bilateral or multiple primary tumours, and sometimes associated with non-cancer phenotypes. While many inherited cancer syndromes show an autosomal dominant pattern of inheritance with incomplete cancer penetrance, several autosomal recessive and X-linked syndromes have also been described. Subsequent epidemiological, cohort-based and meta-analysis studies indicated that family history is one of the strongest risk factors for many cancer types, including breast, ovarian, colon, pancreatic, prostate, and gastric cancers (Dupont and Page 1985; Slattery and Kerber 1993; Pharoah et al.

1997). These observations thus supported a broad spectrum of cancer predisposition syndromes that implicated yet unknown hereditary factors in affected families.

Following the discovery of the DNA double helix by James Watson and Francis Crick in 1953, with notable contributions from Rosalind Franklin and Maurice Wilkins, geneticist Alfred Knudson famously described a two-hit hypothesis explaining the mechanism of cancer development in individuals with inherited cancer predisposition (Watson and Crick 1953; Knudson 1971). Knudson's hypothesis suggested that inactivation of two alleles of what subsequently was termed a tumour suppressor gene was required for tumour initiation and progression, leading to deregulated suppression of cellular growth and proliferation (Marshall 1991; Lee and Muller 2010). Conducting an epidemiological study of retinoblastoma, an early childhood form of retinal cancer, Knudson observed that children with bilateral tumours were more likely to have a family history of the disease and an earlier age of diagnosis than children with unilateral non-hereditary disease. These findings supported a model where inactivation of both alleles of a tumour suppressor gene was required for cancer to occur. In retinoblastoma, tumourigenesis was initiated by biallelic inactivation of *RB1* (Friend et al. 1986). While double somatic mutations acquired during the lifetime were required in non-hereditary cancers, hereditary disease resulted from an inherited damaged copy of the gene and a secondary somatic mutation in the other allele. Consistent with Knudson's initial observations, the majority of known cancer predisposition genes described to date are classical tumour suppressor genes associated with autosomal dominant cancer susceptibility (Rahman 2014). Across recent cancer genomics cohorts, almost 50% of patients harbouring pathogenic germline variants have secondary somatic genomic hits or low tumour mRNA expression in the respective gene (Lu et al. 2015; Huang et al. 2018). This may occur through small somatic mutations, somatic structural variants (SVs), loss of

heterozygosity (LOH) or epigenetic silencing, resulting in gene inactivation and cancer development.

1.2 Genetic variation in health and disease

1.2.1 Inherited genetic landscape

Since the first human genome was sequenced in 2001, resulting from parallel efforts by the Human Genome Project and private biotechnology company Celera Genomics, advancements in technology have allowed increasing resolution of genetic variation across human genomes (Venter et al. 2001; Lander et al. 2001). From shotgun sequencing of bacterial artificial chromosomes to next-generation sequencing (NGS), and recently third-generation or long-read sequencing, the complexity and diversity of sequence variation has been shown across thousands of individuals from various ethnic populations (Altshuler et al. 2012; Auton et al. 2015). The average human has 4-5 million variant sites, consisting mostly of single nucleotide variants (SNVs) and small insertions and deletions (indels). A small proportion of genetic variation is the result of SVs greater than 50 bp, including unbalanced rearrangements, such as deletions, duplications, and insertions, and balanced rearrangements, such as inversions and translocations. Although fewer in number, SVs account for a substantial proportion of sequence diversity and underlie both complex and Mendelian diseases (Feuk et al. 2006; Stankiewicz and Lupski 2010; Weischenfeldt et al. 2013; Merker et al. 2018). Phenotypic variability in health and disease is thus related in part to varying contributions from genetic alterations that impact both protein-coding and non-coding regions.

1.2.2 Variant classification for inherited cancer susceptibility

Given the extent of natural genetic variation, standard recommendations for the interpretation of potential disease-causing variants are necessary for effective clinical translation. Current guidelines for variant classification were published in 2015 by the American College of Medical Genetics and Genomics (ACMG) and Association of Molecular Pathology (AMP) (Richards et al. 2015). These described how information from population genomic databases, functional molecular assays, *in silico* computational functional predictors, and pedigree- and cohort-based genotype-phenotype studies should be integrated to guide reproducible variant interpretations with clinical validity and utility. These broad guidelines have since been adapted for specific gene and disease contexts to improve the accuracy of variant interpretation. However, lack of biological, functional and/or phenotypic information for variants in both known and novel cancer predisposition genes limit accurate variant assessment and genetic diagnosis in hereditary cancer families. Accordingly, variants of uncertain significance (VUS) are identified in 15-41% of individuals referred for clinical genetic testing using multigene panel-based NGS (LaDuca et al. 2014; Lincoln et al. 2015; Kamps et al. 2017). The application of NGS technologies in patient populations and otherwise healthy individuals also necessitates clinical standards for reporting secondary or incidental findings that may have implications for the individual and their family (Green et al. 2013; Kalia et al. 2017).

VUS often consist of single base substitutions that are not predicted to result in a truncated protein (i.e. frameshift, nonsense, or canonical splice site variants) and where functional evidence supporting variant pathogenicity is unavailable or inconclusive. Compared to protein-truncating variants, missense and synonymous variants in high-penetrance genes reflect a small number of all pathogenic and likely pathogenic germline variants due to their uncertain biological and clinical

significance. As a consequence, rare missense and synonymous variants often remain classified as VUS. For example, a systematic review of missense variants reported in *BRCA1* and *BRCA2* revealed that only 43 and 28 variants, respectively, could be classified as deleterious or probably deleterious according to past guidelines from the International Agency for Research on Cancer (Plon et al. 2008; Corso et al. 2018). Nevertheless, hypomorphic missense variants in these genes have been associated with moderately higher breast cancer risk, suggesting that reduced protein function is sufficient to confer clinically-relevant cancer risk (Shimelis et al. 2017).

Alternative mechanisms underlying variant pathogenicity have also been characterized in dyskeratosis congenita (DC), an inherited cancer syndrome characterized by myelodysplastic syndrome, acute myeloid leukemia and squamous cell carcinoma of the head and neck or anogenital region (Savage and Alter 2009). Missense variants in *TERT* that show reduced telomerase activity *in vitro* have been reported in affected carriers and in families with DC, demonstrating that hypomorphic variants increase risk of developing acute myeloid leukemia (Calado et al. 2009). Furthermore, haploinsufficiency of telomerase associated with the K902N substitution leads to anticipation in DC caused by progressive shortening of telomeres across generations (Armanios et al. 2005). Haploinsufficiency has also been hypothesized to lead to early malignant lesions in von Hippel-Lindau syndrome and tuberous sclerosis complex (TSC), predominantly caused by pathogenic germline variants in *VHL* and *TSC1* or *TSC2*, respectively (Peri et al. 2017). Despite the potential functional impact of missense variants, the functional assessment for many cancer predisposition genes is limited by the uncertain clinical validity of computational predictors and *in vitro* assays (Duzkale et al. 2013; Starita et al. 2017). These present an ongoing challenge to the interpretation of VUS and clinical management of families with suspected inherited cancer syndromes.

1.2.3 Prevalence of pathogenic germline variants

The complete landscape of genetic factors underlying cancer susceptibility, including variable contributions from low-, moderate- and high-penetrance variants and complex interactions between genes and environment, is not well understood (Dempfle et al. 2008). Linkage analysis in families with multiple affected generations identified shared haplotypes, combinations of alleles that tend to be inherited together due to their physical proximity on a chromosome, that segregated with disease (Hall et al. 1990; Lenoir et al. 1991). This powerful approach helped identify several high-penetrance cancer predisposition genes, including *BRCA1* and *BRCA2* in HBOC, *MLH1*, *MSH2*, *MSH6* and *PMS2* in Lynch syndrome, and *CDH1* in hereditary diffuse gastric cancer (HDGC) (Hall et al. 1990; Fishel et al. 1993; Lindblom et al. 1993; Peltomäki et al. 1993; Bronner et al. 1994; Wooster et al. 1994; Guilford et al. 1998). Clinical implementation of multigene NGS panels has since allowed for broader characterization of cancer susceptibility genes underlying disease in phenotypically heterogeneous cancer populations. Among index cases meeting clinical criteria defined by the National Comprehensive Cancer Network (NCCN) for HBOC, 6-12% carry germline variants in *BRCA1* or *BRCA2* (Beck et al. 2020; Yoo et al. 2020). *BRCA1/2*-negative HBOC families may alternatively harbour germline variants in moderate-penetrance breast cancer susceptibility genes, including *ATM*, *CHEK2*, or *PALB2* (Couch et al. 2015; Schroeder et al. 2015; Tung et al. 2016).

Recent large-scale genome sequencing in patient cohorts unselected for personal or family cancer history suggests that pathogenic and likely pathogenic germline variants in moderate- to high-penetrance cancer predisposition genes underlie around 8% of all cancers (Huang et al. 2018). Site-specific estimates vary between 2-4% in cholangiocarcinoma and acute myeloid leukemia and

19-23% in ovarian cancer and pheochromocytoma and paraganglioma. Surprisingly, similar rates of pathogenic germline findings have been reported in individuals referred for clinical hereditary cancer testing. Around 8-11% of individuals who receive targeted multigene germline testing on the basis of personal and/or family cancer history are found to carry pathogenic or likely pathogenic germline variants in known cancer predisposition genes (Kurian et al. 2014; LaDuca et al. 2014; Susswein et al. 2016; Couch et al. 2017). Although clinical multigene panels have demonstrated clinical utility through the identification of actionable germline variants that would result in a change in clinical management, these findings also suggest a discordance in phenotype-based clinical testing criteria that may reduce the sensitivity of molecular diagnosis for inherited cancer predisposition (Desmond et al. 2015). Despite widespread use of multigene NGS panels in hereditary cancer testing, 67-73% of individuals referred for index genetic testing receive uninformative results (LaDuca et al. 2014).

1.3   Phenotypic heterogeneity in cancer predisposition syndromes

Phenotype-based clinical genetic testing guidelines are used to identify individuals who are most likely to carry pathogenic germline variants with clinical actionability. This may ultimately allow opportunities for cascade carrier testing in family members, increased cancer screening, prophylactic surgery or use of targeted therapies. However, due to incomplete disease penetrance, uninformative family structure, and/or incomplete family cancer history, many hereditary cancer families remain undiagnosed. The Ashkenazi Jewish (AJ) population is one of the best studied populations for HBOC due to the high prevalence of pathogenic germline variants in *BRCA1* and *BRCA2*, occurring in over 2% of the population (Hartge et al. 1999). Population screening for three common AJ founder variants (*BRCA1* c.185delAG, *BRCA1*

c.5382insC and *BRCA2* c.6174delT), which account for the majority of HBOC kindreds in this population, increased the number of carriers identified by 40-63% compared to current delivery models of genetic testing based on personal or family cancer history (Levy-Lahad et al. 1997; Kauff et al. 2002; Manchanda et al. 2015; Walsh et al. 2017). Breast and ovarian cancer risks are similar among carriers for *BRCA1* and *BRCA2* obtained through population genetic screening or based on family history, indicating that carriers identified through population screening may equally benefit from enhanced cancer prevention and screening strategies (King et al. 2003; Gabai-Kapara et al. 2014). Although the current semi-opportunistic model for genetic testing aims to offer genetic testing to individuals with the greatest probability of harbouring germline variants, only 38-45% of carriers identified through population genetic testing meet clinical genetic testing criteria (Metcalfe et al. 2010; Metcalfe et al. 2013). Therefore, many carriers who would otherwise benefit from early detection and personalized cancer risk management are missed.

1.3.1 Challenges of incomplete penetrance

Despite the significant cancer risks associated with high-penetrance cancer predisposition syndromes, only a minority are associated with almost complete cancer penetrance. These include familial adenomatous polyposis (FAP), associated with a 93% risk of CRC by 50 years, and Li Fraumeni syndrome, associated with a nearly 100% risk for malignancy by 70 years (Jasperson et al. 2010; Mai et al. 2016). However, many cancer syndromes show incomplete penetrance that may be influenced by modifying genetic factors, lifestyle, environment and/or stochastic processes. For example, the *RAD51* c.135G>C allele, occurring in the 5' untranslated region (5' UTR), increases breast cancer risk in carriers for *BRCA2*, with a hazard ratio of 3.17 in CC

homozygotes (Levy-Lahad et al. 2001; Antoniou et al. 2007). This is due to the activation of an alternative splice site in the 5' UTR, and was shown to have a greater effect in homozygotes than heterozygotes. Although retrospective studies assessing the influence of environmental and lifestyle factors on cancer risk have inherent selection and ascertainment biases, some consistent findings have been found among carriers for *BRCA1* or *BRCA2* (Milne and Antoniou 2016). Tamoxifen use in *BRCA1* and *BRCA2* carriers and oral contraceptive use in *BRCA1* carriers are protective from breast and ovarian cancer, respectively. Reproductive history, including breast feeding, age at menarche and age at first full-term pregnancy, has also been consistently reported as protective from breast cancer in *BRCA1* carriers (Friebel et al. 2014).

Incomplete penetrance in hereditary cancer families may also result from the contribution of moderate-penetrance genes conferring modest increases in cancer risk. Conventionally, germline variants associated with moderate-penetrance cancer susceptibility are defined as alleles conferring a two to fivefold higher cancer risk (Tung et al. 2016). Several moderate-penetrance cancer predisposition genes have been described to date, where *ATM* and *CHEK2* are among the most common. Early observations of families with autosomal recessive ataxia-telangiectasia supported an association between *ATM* carrier status and breast cancer susceptibility, reporting a high prevalence of breast cancers in heterozygous carriers (Swift et al. 1987; Renwick et al. 2006). Although around 1% of the general population harbour heterozygous pathogenic variants in *ATM*, cohort studies suggest that *ATM* carrier status confers susceptibility to breast and pancreatic cancers (Broeks et al. 2000; Thompson et al. 2005; Roberts et al. 2012). Germline variants in the serine/threonine kinase *CHEK2* were first reported in several families with Li Fraumeni syndrome without causal variants in *TP53* (Bell et al. 1999). Similar to causal variants in *ATM*, the most common pathogenic variant in *CHEK2* (1100delC) is present in around 1% of unaffected

individuals but is enriched in breast cancer families. The 1100delC allele is associated with bilateral breast cancer and family breast cancer history, suggesting that individual cancer risks associated with moderate-penetrance cancer susceptibility genes depend on other individual- or family-specific factors (Meijers-Heijboer et al. 2002; Vahteristo et al. 2002). Due to the more modest increases in lifetime cancer risk, clinical management for carriers of moderate-penetrance variants requires careful consideration of personal and family cancer history to determine appropriate recommendations for cancer screening or prophylactic interventions. Accordingly, modified counseling frameworks for moderate-penetrance genes have been established to avoid unnecessary harm in affected families (Tung et al. 2016).

Beyond the contributions of high- and moderate-penetrance germline variants in well-characterized cancer predisposition genes, genome-wide associations studies (GWAS) in large cancer and control cohorts support a role for low-penetrance genetic variation in cancer susceptibility (Easton et al. 2007; Houlston et al. 2008). These findings indicate that cancer results from multiple genetic, epigenetic, and environmental factors, which may interact additively or synergistically to alter cellular function (Eichler et al. 2010). GWAS have identified several genes conferring small increases in risk for various cancer types, demonstrating that genotype- and tissue-specific differences in gene expression and function may underlie some interindividual variability in disease susceptibility (Cookson et al. 2009; Wu et al. 2018). Common variants in *cis*-regulatory elements may affect transcriptional regulation by directly promoting or reducing transcription factor binding or disrupting physical interactions between regulatory elements (Gallagher and Chen-Plotkin 2018). Alternatively, polymorphisms in coding or splice regions may be associated with altered protein function or alternative splicing, respectively.

Intriguingly, common variants in otherwise high- and moderate-penetrance cancer predisposition genes have also been associated with small increases in cancer risk. Two low-penetrance alleles in *APC* have been described in patients with multiple colorectal adenomas and CRC: I1307K, an AJ founder allele, and E1317Q (Laken et al. 1997; Frayling et al. 1998; Lamlum et al. 2000; Liang et al. 2013). The I1307K allele has been associated with a relative risk of 1.5-1.7 for colorectal neoplasia, and may directly contribute to 3-4% of all CRC in the AJ population (Gryfe et al. 1999). Similarly, the European *CHEK2* founder allele I157T has been associated with increased risks for breast, colon, kidney, prostate and thyroid cancers (Cybulski et al. 2004). Although the identification of low-penetrance germline variants may improve clinical management for individuals from populations with well-established cancer risk estimates, the clinical utility of low-penetrance alleles in the general population is controversial, given the potential harm of interventions indicated for individuals at high risk.

1.4 Biological mechanisms underlying high-penetrance cancer syndromes

1.4.1 Molecular and morphological characteristics

As regulators of cellular proliferation and guardians of genome integrity, many high-penetrance cancer predisposition genes have been associated with distinct genetic, cellular, and histological cancer subtypes. For example, pathogenic germline variants in mismatch repair (MMR) genes are associated with MMR protein deficiency in Lynch syndrome-related tumours (Aaltonen et al. 1993; Hampel et al. 2005a). MMR deficiency is uncommon in sporadic cancers, observed in 3.8% of all cancers and 15% of CRC (Boland and Goel 2010; Bonneville et al. 2017). Universal screening for MMR deficiency in all new colorectal and endometrial cancer diagnoses has thus been widely adopted by health authorities to identify possible Lynch syndrome families

who may benefit from clinical intervention and cancer risk management (Aaltonen et al. 1998; Boland et al. 1998). Similar clinicopathological features are observed in breast cancer, where expression of the estrogen receptor (ER), progesterone receptor (PR) and human epidermal growth factor receptor (HER2) can inform prognosis and indicate the use of specific therapies (Fisher et al. 1989; Dent et al. 2007; Dunnwald et al. 2007; Tischkowitz et al. 2007; Liedtke et al. 2008). Triple-negative breast cancers characterized by the absence of ER, PR and HER2 by immunohistochemistry (IHC) have been associated with carrier status for *BRCA1* and *BRCA2* (Foulkes et al. 2003; Hartman et al. 2012; Couch et al. 2015).

Diffuse and lobular histological subtypes of gastric and breast cancer, respectively, are established criteria for the identification of families with HDGC. Germline variants in *CDH1*, the gene encoding epithelial-specific transmembrane protein E-cadherin, account for the majority of families meeting International Gastric Cancer Linkage Consortium (IGCLC) criteria for HDGC (Fitzgerald et al. 2010; van der Post et al. 2015). Loss of E-cadherin has been specifically associated with signet ring cell (SRC) morphology in HDGC and sporadic SRC carcinoma, suggesting a common mechanism of pathogenesis is associated with *CDH1* inactivation in both hereditary and sporadic disease (Humar et al. 2009; Pernot et al. 2015).

1.4.2 Mutational signatures

The pattern of mutations in individual cancer genomes caused by various endogenous and exogenous processes are termed somatic mutational signatures. Defined by the particular base substitution and flanking 5' and 3' nucleotide sequences, mutational signatures have recently provided insights into the complex aetiology of cancer (Alexandrov et al. 2013; Alexandrov et al. 2020). Genomic signatures provide a basis for tissue-specific molecular subtyping with potential

prognostic and therapeutic implications, and several mutational signatures observed in large cancer genomics cohorts have been characterized in the Catalogue for Somatic Mutations in Cancer (COSMIC) (Nik-Zainal and Morganella 2017; Van Hoeck et al. 2019). Although the molecular mechanisms underlying tumour development in many cancer syndromes are not well understood, tumours caused by germline and somatic inactivation of genes involved in DNA repair show distinct mutational profiles. Genomic instability at DNA microsatellites, small tandem repeats of one to six nucleotides in length, is a common feature of tumours with biallelic loss in one of four MMR genes. Microsatellite instability (MSI) is associated with global hypermutation (>100,000 somatic mutations) and is observed in more than 90% of Lynch syndrome-related tumours (Ionov et al. 1993; Aaltonen et al. 1998). Deficiency in HR, a cellular process with roles in DNA repair, DNA replication-fork rescue, meiotic chromosome segregation and telomere maintenance, has similarly been characterized in cancers associated with germline variants in *BRCA1* and *BRCA2* (Davies et al. 2017). Due to their crucial role in double-stranded break repair (DSBR), BRCA1 and BRCA2 deficiency results in large-scale genomic events caused by inaccurate repair of DSBs. Characterizing the morphological and molecular features of tumours associated with high-penetrance cancer syndromes may thus improve identification, diagnosis and treatment of hereditary cancer families.

1.5   Challenges in molecular diagnosis

Despite the advances in NGS technologies and implementation of clinical multigene sequencing panels, many individuals with a strong personal or family history of cancer receive uninformative genetic testing results that may lead to an uncertainty in cancer risk, nonspecific recommendations for cancer screening, and missed opportunities for the use of targeted therapies

or predictive genetic testing in at-risk relatives. For example, causal germline variants are identified in around 70% of cases meeting Amsterdam I or II criteria for Lynch syndrome, which require at least three family members with a spectrum cancer (Vasen et al. 1991; Vasen et al. 1999; Lynch et al. 2009). Individuals with strong phenotypic indications of high-penetrance cancer predisposition syndromes may harbour rare germline variants that are undetectable by targeted sequencing assays, such as SVs, non-coding variants, and epigenetic mutations (epimutations). The limitations of panel-based approaches to hereditary cancer testing were highlighted by Rhees *et al.* (2014), who identified a frequent *MSH2* inversion in several Lynch syndrome families who remained undiagnosed by standard clinical assays (Rhees et al. 2014). Recent technological advancements have resulted in decreasing costs for genome sequencing (GS), allowing for its wider application and future replacement of comparable but currently more cost-effective exome sequencing (ES) (Schwarze et al. 2018). GS may provide immediate benefits to the areas of precision oncology and rare disease in particular to improve the rate of genetic diagnosis in individuals with strong clinical phenotypes.

1.5.1   Non-coding variation causes aberrant gene regulation

Beyond variants in coding and canonical splice site regions, genetic variation in intronic and regulatory regions has also been reported in hereditary cancer families. Consistent with a role for non-coding variation in the pathogenesis of cancer predisposition syndromes, a 40 kb duplication upstream of bone morphogenic protein (BMP) antagonist *GREM1* was identified as the cause of hereditary mixed polyposis syndrome (HMPS) in several affected families (Jaeger et al. 2012). Similar to other gastrointestinal polyposis syndromes, HMPS is associated with variable colorectal polyposis conferring increased risks for CRC. The causal duplication contains predicted

enhancer elements, which show strong interactions with the *GREM1* promoter and result in an increase in allele-specific expression. Deleterious germline variants in regulatory elements have also been described in families with gastric adenocarcinoma and proximal polyposis of the stomach (GAPPS), an autosomal dominant gastric cancer syndrome characterized by extensive polyposis in the fundus of the stomach (Worthley et al. 2012). Pathogenic variants in the promoter 1B of *APC* reduce transcription factor binding, resulting in reduced expression from the variant allele (Li et al. 2016). These findings demonstrate that short- and long-range interactions between *cis*-regulatory elements contribute to allele-specific expression of cancer predisposition genes.

1.5.2   Epimutations remain undetected by genome sequencing

Epigenetic regulation of transcription occurs through reversible chemical modification of DNA and histones, the structural protein component of chromatin that helps maintain genome structure, accessibility and integrity. Methylation is the most common DNA modification, occurring at cytosine bases and playing important roles in embryonic development, genomic imprinting and transcriptional regulation (Greenberg and Bourc'his 2019). Allele-specific promoter methylation and transcriptional repression have also been proposed to contribute to a proportion of the missing heritability of complex disease (Meaburn et al. 2010). This mechanism of gene inactivation has been identified as an alternative cause of Lynch syndrome in rare families (Suter et al. 2004). Constitutional methylation of the *MLH1* promoter has been described in inherited and *de novo* cases, occurring either as primary epimutations or as a consequence of underlying germline variants (Pinto et al. 2018). *De novo* constitutional methylation has been especially associated with early onset cancer and history of multiple primary tumours, although constitutional methylation at the *MLH1* promoter conferring autosomal dominant cancer

susceptibility has been associated with both primary and secondary epimutations (Leclerc et al. 2018).

1.5.3   Tissue-specific mechanisms of tumourigenesis

Due to reduced penetrance, broad cancer spectrum, and molecular heterogeneity observed in many cancer predisposition syndromes, some genetically undiagnosed hereditary cancer families may harbour pathogenic germline variants in high-penetrance genes that may not be detectable in standard clinical assays or cannot be interpreted based on available functional evidence. RNA sequencing (RNA-seq) has been shown to improve the rate of diagnosis in Mendelian disorders where standard clinical testing was uninformative by evaluating the functional impact of genetic variation on gene expression, alternative splicing, or allele-specific expression (Cummings et al. 2017; Kremer et al. 2017). In cancer predisposition syndromes, tumour genome and transcriptome sequencing can provide insights into individual-, gene- and tissue-specific mechanisms of disease pathogenesis. Integrating cellular and molecular tumour characteristics into variant interpretation may thus provide evidence supporting the classification of germline VUS or implicate previously undetected variants missed by targeted NGS (Lu et al. 2015; Huang et al. 2018; Shirts et al. 2018). This indicates a need for integrated clinical and molecular approaches to improving the identification and molecular diagnosis of carriers for high-penetrance cancer predisposition syndromes.

1.6   Research hypothesis and objectives

Despite widespread use of multigene NGS panels, many individuals meeting phenotype-based clinical genetic testing criteria receive uninformative genetic testing results. This may be due in part to the inherent limitation of targeted NGS to detect balanced rearrangements, characterize precise SV breakpoints, and identify variants affecting *cis*-regulatory elements involved in the regulation of gene expression. Characterizing genetic variation in both coding and noncoding regions of known cancer predisposition genes is necessary to inform clinical management guidelines for individuals from subpopulations with varying levels of cancer risk, including local populations, patient populations at various stages of disease, and founder populations. As GS becomes more widely used with improved technologies and decline in sequencing costs, GS will improve rates of diagnosis in cancer predisposition syndromes and other Mendelian disorders. Limitations in biological and clinical knowledge regarding variant pathogenicity indicate that tumour sequencing may improve variant interpretation for individuals with suspected high-penetrance cancer syndromes. For example, tumour-normal sequencing may detect potential second hits in classical tumour suppressor genes, identify the presence of characterized somatic mutational signatures, or characterize tissue-specific transcriptional regulation associated with carrier status for cancer predisposition genes.

Identification and molecular diagnosis of carriers for high-penetrance cancer syndromes has critical implications for the health and well-being of patients and their families, for cancer prevention, early cancer detection and individual health empowerment. Due to reduced penetrance and variable cancer spectrum associated with many cancer predisposition syndromes, carriers of clinically actionable germline variants may remain undetected. While undetected or unclassified germline variants may underlie cancer predisposition in some families, germline

variation is associated with distinct tumour phenotypes that can be evaluated to improve genetic

diagnosis, cancer risk stratification and cancer treatment. Therefore, a broader understanding of

clinical phenotypes and molecular heterogeneity of cancer susceptibility may improve the

ascertainment and management of cancer families. I hypothesized that germline variation in

cancer predisposition genes is associated with distinct genomic, transcriptional and pathological

signatures that contribute to disease pathogenesis. The main goal of this research was thus to

investigate the functional, pathogenic and clinical implications of germline variants in cancer

predisposition genes in an unselected patient population and in undiagnosed hereditary cancer

families. These were approached through the following aims:

i.   Explore molecular mechanisms associated with carrier status for high-penetrance

cancer predisposition genes in patient tissues and organoid model systems.

ii.   Investigate the utility of long-read genome sequencing in germline variant resolution

and genetic diagnosis of suspected hereditary cancer families.

iii.   Describe the clinical utility of GS in patient populations through phenotypic

characterization of moderate- to high-penetrance cancer syndromes.

Chapter 2: Mechanisms of molecular pathogenesis in high-penetrance cancer predisposition syndromes

2.1   Introduction

Although the interpretation of genetic variation is often hindered by a lack of functional evidence to support a role in gene function or the regulation of gene expression, pathogenic germline variants in cancer predisposition genes have been associated with distinct mechanisms of pathogenesis compared to sporadic cases of the same cancer type. High-penetrance genes involved in several DNA repair pathways, including MMR, homologous recombination (HR) and base excision repair (BER), are associated with specific patterns of mutations across individual cancer genomes that reveal inherent genetic aetiologies for cancer progression (Alexandrov et al. 2013; Pilati et al. 2017; Viel et al. 2017). Mutational signatures associated with moderate-penetrance cancer predisposition genes have been investigated in site-specific cancer genomics cohorts. This approach identified uncharacterized roles for *PALB2* and *RAD51* in breast cancer pathogenesis through HR deficiency (Polak et al. 2017).

Single base substitution (SBS), doublet base substitution (DBS) and indel signatures resulting from impaired DNA damage repair have been broadly characterized across resectable cancers (Alexandrov et al. 2020). I hypothesized that high- and moderate-penetrance germline variants underlie global genomic instability across multiple cancer types and contribute to tumour progression. Using GS, we explored the landscape of germline variants associated with DNA repair deficiency across advanced cancers and demonstrate the utility of targeted tumour sequencing in clinical patient populations with suggestive tumour features. Ultimately, our

findings may allow opportunities for improved clinical management and cancer treatment in affected families.

2.2 Materials and Methods

2.2.1 Whole-genome sequencing

Germline genome sequencing, tumour genome sequencing and tumour transcriptome sequencing were performed for 705 adult patients with primarily metastatic cancers participating in BC Cancer's Personalized OncoGenomics (POG) program in Vancouver, British Columbia, Canada (NCT02155621). This study was approved by the University of British Columbia Research Ethics Committee, and written informed consent was obtained for all participants (H12-00137, H14-00681, H16-00291). Tissue collection, nucleic acid extraction and short-read sequencing library preparation have been previously described (Pleasance et al. 2020). Briefly, DNA was extracted from peripheral blood and from tumour biopsy sections embedded in optimal cutting temperature compound. PCR-free genome libraries were prepared for paired-end genome sequencing, which was performed on the Illumina HiSeq to an average coverage of 40X for peripheral blood and 80X for tumour samples. mRNA was purified from tumour biopsy specimens, converted to cDNA, and paired-end sequencing of strand-specific libraries was performed on Illumina HiSeq instruments to a mean depth of approximately 200 million reads. Reads were aligned to the human reference genome version hg19 using BWA-MEM v0.7.6 (Li and Durbin 2009). All sequencing data for the POG cohort have been deposited in the European Genome-phenome Archive under accession EGAS00001001159.

2.2.2   Germline variant calling

Small germline variants were identified using SAMtools v0.1.17, and region-based filtering was performed to prioritize candidate variants across 99 known cancer predisposition genes (Supplementary Table 2.1) (Li et al. 2009; Cingolani et al. 2012; Auton et al. 2015; Landrum et al. 2018). Small germline variants were annotated using ANNOVAR, and semi-automated variant classification according to ACMG/AMP guidelines was performed using InterVar (Wang, Li & Hakonarson, 2010; Li & Wang, 2017). Germline SNVs and indels were prioritized according to known or predicted clinical impact, predicted functional impact and population frequency. Variants with ClinVar annotations or InterVar predictions of pathogenic or likely pathogenic, variants with conflicting interpretations in ClinVar, and variants predicted to result in nonsense, frameshift, abnormal splicing and start loss alterations were prioritized for manual review. Germline copy number variants were identified using ControlFREEC, and structural variants identified using DELLY v0.7.3 and Manta v1.0.0 were aggregated with mRNA fusion events detected by Trans-ABySS v1.4.10 (Simpson et al. 2009; Robertson et al. 2010; Boeva et al. 2012; Rausch et al. 2012; Chen et al. 2016). Joint annotation and filtering of SVs and fusions was performed using MAVIS (Reisle et al. 2019).

2.2.3   Somatic mutation and copy number analysis

Somatic mutations were identified using SAMtools v0.1.17 and v4.3.5, and Strelka v1.0.6 as previously described (Li et al. 2009; Ding et al. 2012; Saunders et al. 2012; Pleasance et al. 2020). Somatic SNVs and indels were classified by base substitution and 5' and 3' nucleotide context into one of several categories using a published framework (Alexandrov et al. 2013). The contribution of mutational signatures defined in COSMIC versions 2 and 3 to each

tumour's mutation profile was then calculated by solving non-negative least squares problems using the R package MutationalPatterns (Blokzijl et al. 2018). The contribution of DBS signatures and indel signatures defined in COSMIC version 3 was similarly calculated using the R package Palimpsest (Shinde et al. 2018). Somatic copy number calling and LOH prediction were performed using CNAseq v0.0.6 and APOLLOH v0.1.1, respectively, and LOH status for pathogenic and likely pathogenic germline variants was determined through manual review in the Integrated Genomics Viewer (IGV) v2.7.0 (Jones et al. 2010; Ha et al. 2012). Tumour genome data were manually reviewed to identify secondary somatic events in carriers for moderate- to high-penetrance germline variants, including copy neutral or deletion LOH, non-synonymous small somatic mutations, and somatic SVs.

2.2.4   Targeted tumour-normal sequencing

Between June 2018 and December 2019, individuals referred to the BC Cancer Hereditary Cancer Program on the basis of a Lynch syndrome-related cancer with MMR deficiency as assessed by IHC were offered tumour and germline genetic testing using Ambry Genetics' TumorNext-Lynch assay (https://www.ambrygen.com/clinician/genetic-testing/18/oncology/tumornext-lynch). Electronic pathology and genetics laboratory reports were reviewed to describe tumour type and histology, age of cancer diagnosis, and results from IHC, MSI testing, *MLH1* promoter hypermethylation testing and genetic testing. Family cancer history was reviewed to identify cases meeting clinical testing criteria defined by the Amsterdam I, Amsterdam II and revised Bethesda guidelines or with other personal and/or family history of Lynch syndrome-related cancers (Vasen et al. 1991; Vasen et al. 1999; Umar et al. 2004).

2.3   Results

2.3.1   Prevalence of moderate- to high-penetrance germline variants in advanced cancers

The tumour landscape is defined by inherited germline events in addition to tumour-specific somatic mutations. To investigate the overall contribution of high- and moderate-penetrance germline variants to the molecular pathogenesis of advanced tumours, germline variants in 99 cancer predisposition genes were identified by GS and manually curated according to ACMG/AMP guidelines. Among 705 advanced cancer patients, 13.8% ($n = 97$) of individuals were carriers of moderate- to high-penetrance germline variants in one or more cancer predisposition genes (Figure 2.1). This was comparable to published prevalence estimates between 12.2% and 17.8% in similar advanced cancer cohorts (Schrader et al. 2016; Mandelker et al. 2017; Robinson et al. 2017; Bertelsen et al. 2019). Pathogenic and likely pathogenic germline variants were identified across 17 cancer types spanning 32 genes, with variants in *CHEK2* ($n = 14$), *BRCA2* ($n = 14$), *MUTYH* ($n = 13$), *BRCA1* ($n = 12$) and *ATM* ($n = 9$) accounting for 57.9% of all variants. Germline variants associated with rare high-penetrance cancer predisposition syndromes, including FAP, Lynch syndrome, *MUTYH*-associated polyposis (MAP), Li Fraumeni syndrome and TSC, contributed to a small number of cases in this cohort.
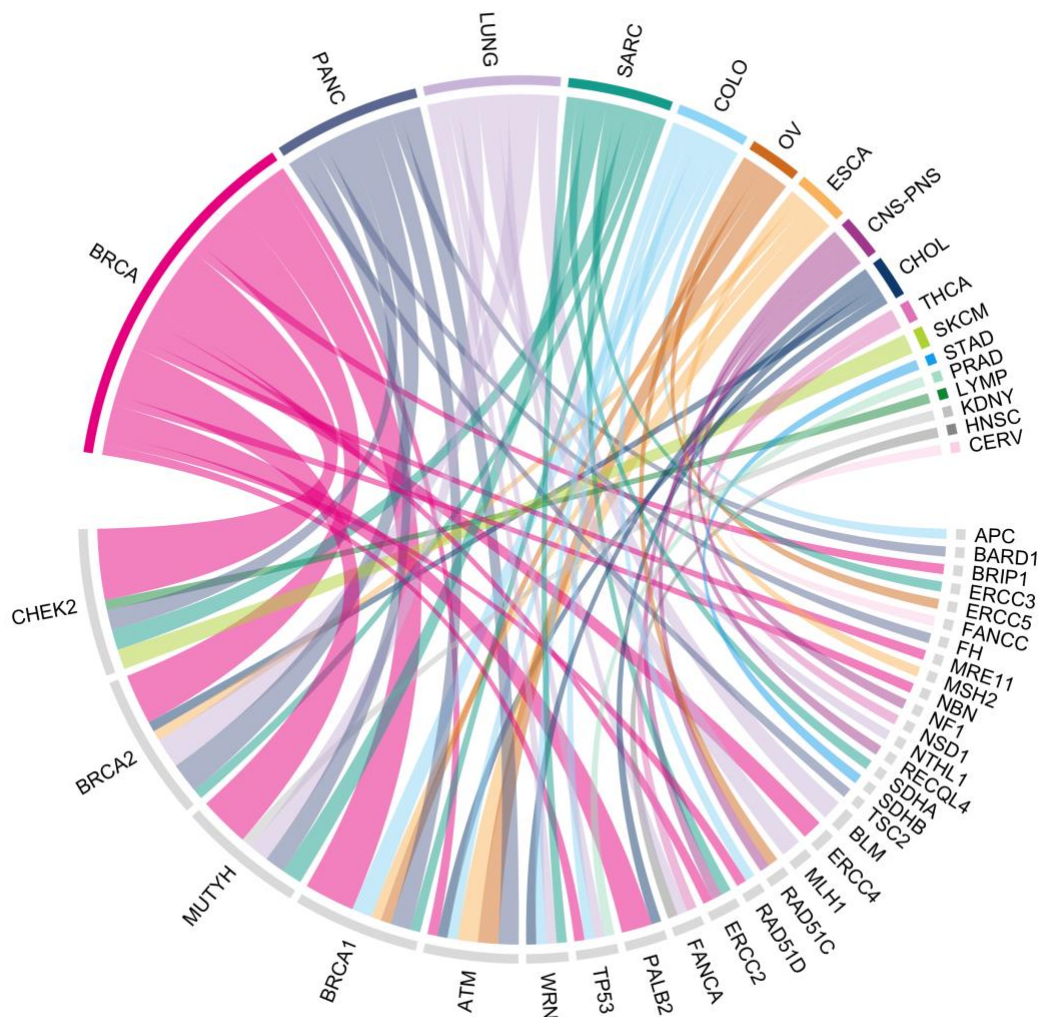
Figure 2.1. Landscape of germline variants across advanced cancers. Chord diagram showing cancer cases where pathogenic or likely pathogenic germline variants in cancer predisposition genes were identified. The width of each link represents the number of carriers with a given cancer diagnosis (top) and for a given gene (bottom). BRCA: breast cancer; PANC: pancreatic cancer; LUNG: lung cancer; SARC: sarcoma; COLO: colorectal cancer; OV: ovarian cancer; ESCA: esophageal adenocarcinoma; CHOL: cholangiocarcinoma; CNS-PNS: neuroendocrine cancer; THCA: thyroid cancer; SKCM: melanoma; LYMP: blood and lymphoid cancer; STAD: stomach adenocarcinoma; PRAD: prostate adenocarcinoma; KDNY: kidney cancer; HNSC: head and neck cancer.

Consistent with a two-hit model of tumorigenesis, 46% of germline variants were associated with a secondary somatic genomic event, such as somatic small mutations, somatic SVs, or LOH, at the same locus. Accordingly, 47% of all carriers had an identifiable second hit in at least one gene, including one case with biallelic germline variants in *MUTYH*. Secondary somatic alterations were common across *ATM* (67%), *BRCA1* (67%), *BRCA2* (64%), and *CHEK2* (50%), with deletion and copy neutral LOH contributing to the majority of secondary genomic events in these genes. Although a broad cancer spectrum was observed among carriers of high-penetrance variants in *BRCA1* and *BRCA2* and moderate-penetrance variants in *ATM* and *CHEK2*, we found an enrichment of breast cancers among *CHEK2* carriers showing LOH compared to those without an identifiable secondary event ($P = 0.03$, Fisher's exact test). While modest *CHEK2*-associated increases in lifetime cancer risk have been reported for several cancer types, including those of the breast, colon, and prostate, these findings suggested a specific mechanism underlying breast cancer pathogenesis in *CHEK2* carriers. Consistent with previous studies, we also observed a higher prevalence of triple-negative breast cancers in carriers for *BRCA1* (4 of 5) compared to carriers for *BRCA2* (0 of 5) (Atchley et al. 2008). In contrast, carrier status for moderate-penetrance breast cancer genes was associated with hormone receptor-positive breast cancer (De Bock et al. 2004; Schmidt et al. 2016).

2.3.2 Mutational patterns of inherited DNA repair deficiency

Several mutational processes were associated with carrier status for high-penetrance genes associated with DNA damage repair. Germline variants in *BRCA1* and *BRCA2*, which play central roles in DSBR through the HR and Fanconi anemia pathways, showed greater contributions from genomic signatures associated with HR deficiency compared to cases with

somatic loss or no genetic alterations in these genes (Figure 2.2a-c). Cases with both germline and somatic alterations in *BRCA1* and *BRCA2* showed higher contributions from COSMIC signature 3, associated with defective HR, compared to germline carriers without second genomic hits or cancers with only somatic alterations in these genes (Alexandrov et al. 2013). Tumours with compound germline and somatic events in *BRCA1* and *BRCA2* showed elevated contributions from doublet base signature DBS4 and DBS6, characterized respectively by GC/TC>AA and TG>AT/CT substitutions, compared to tumours without loss-of-function mutations in those genes (Figure 2.2c).

As homologous and repeat sequences are integral to various DNA repair pathways, specific indel signatures have also been associated with global DNA repair deficiency during tumourigenesis (Helleday et al. 2014). These were substantiated by the finding of high contributions from indel signature 8 (ID8), characterized by deletions > 5 bp at repetitive regions suggestive of error-prone DNA double-stranded break repair by non-homologous end-joining (NHEJ), in tumours associated with germline and compound germline and somatic events in *BRCA1* and *BRCA2* (Alexandrov et al. 2020; de Witte et al. 2020). HR is also required for the formation of telomere-specific structures that are essential for the protection of DNA ends (Dunham et al. 2000; Wang et al. 2004; Verdun and Karlseder 2006). Global genomic instability assessed by the presence of genome-wide LOH, telomeric allelic imbalance, and large-scale state transitions, characterized as chromosomal breaks generating fragments > 10 Mb, was also greater in cases with germline and somatic alterations in *BRCA1* and *BRCA2*.

a

COSMIC V2 Signature 3 (percent of mutations)

COSMIC V3 SBS3 (percent of mutations)

BRCA = 7.3%
non-BRCA = 4.1%

BRCA = 0%
non-BRCA = 0%

BRCA (n = 158)   non-BRCA (n = 547)

germline & somatic (n = 7)
germline (n = 3)
compound somatic (n = 9)
somatic (n = 64)
none (n = 75)
germline & somatic (n = 10)
germline (n = 6)
compound somatic (n = 7)
somatic (n = 135)
none (n = 389)

b

COSMIC V2 Signature 8 (percent of mutations)

COSMIC V3 SBS8 (percent of mutations)

non-BRCA = 16.7%
BRCA = 11.0%

non-BRCA = 5.7%
BRCA = 5.2%

BRCA (n = 158)   non-BRCA (n = 547)

germline & somatic (n = 7)
germline (n = 3)
compound somatic (n = 9)
somatic (n = 64)
none (n = 75)
germline & somatic (n = 10)
germline (n = 6)
compound somatic (n = 7)
somatic (n = 135)
none (n = 389)

c

COSMIC V3 DBS4 (percent of mutations)

all = 9.5%

COSMIC V3 DBS6 (percent of mutations)

all = 9.1%

COSMIC V3 ID8 (percent of mutations)

all = 19.1%

BRCA1, BRCA2 mutation status

germline & somatic (n = 17)
germline (n = 9)
compound somatic (n = 16)
somatic (n = 199)
none (n = 464)

d

COSMIC V3 SBS3 (percent of mutations)

8.6e-06   1.5e-09   1.4e-05

all = 0%

COSMIC V3 SBS8 (percent of mutations)

0.0025   0.0060   0.0204

all = 5.6%

COSMIC V3 ID8 (percent of mutations)

0.0011   5.7e-05   7.2e-05

all = 19.1%

PALB2, BRCA2, BRCA1 carrier status

germline *PALB2* (n = 4)
germline *BRCA2* (n = 14)
germline *BRCA1* (n = 12)
none or other (n = 675)

28

Figure 2.2. *PALB2* **is associated with genomic homologous recombination deficiency in breast cancer independent of** *BRCA1* **and** *BRCA2*. **a.** Percent contribution of signature 3 and single base substitution signature 3 (SBS3) in COSMIC versions 2 and 3, respectively, to breast (BRCA) and other (non-BRCA) tumours associated with germline and somatic alterations in *BRCA1* and *BRCA2*. Black horizontal bars show the median percent contribution for each group, and dashed horizontal lines show the median contribution in BRCA and non-BRCA cohorts. **b.** Percent contribution of signature 8 and single base substitution signature 8 (SBS8) in COSMIC versions 2 and 3, respectively, to breast (BRCA) and other (non-BRCA) tumours associated with germline and somatic alterations in *BRCA1* and *BRCA2*. **c.** Percent contribution of doublet base signatures 4 (DBS4) and 6 (DBS6) and indel signature 8 (ID8) across all cancers in the POG cohort according to *BRCA1* and *BRCA2* mutation status. Median percent contribution for each signature across the entire POG cohort (all) is shown by dashed horizontal lines and annotated to the right. **d.** SBS3, SBS8 and ID8 are elevated in *PALB2* carriers compared to carriers for *BRCA1* and *BRCA2*. Significance between tumours from *PALB2*, *BRCA1* or *BRCA2* carriers compared to tumours from individuals without pathogenic or likely pathogenic germline variants in those genes is shown (*P*-value, Wilcoxon rank-sum test).

Carrier status for most other genes involved in DSBR did not show consistent contributions from COSMIC signature 3. However, as an established breast cancer susceptibility gene, *PALB2* has been recently associated with mutational signatures indicative of HR deficiency (Polak et al. 2017). These findings were consistent across metastatic tumours from four *PALB2* carriers identified here, including three breast cancers and one cholangiocarcinoma. These tumours showed a high proportion of somatic mutations attributed to SBS signatures 3 and 8, SBS3 and SBS8, respectively (Figure 2.2d). This was independent of mutation status in *BRCA1* or *BRCA2*. Notably, all *PALB2* carriers consistently showed greater contributions from SBS8 and ID8 compared to *BRCA1* or *BRCA2* carriers. Although the aetiology of these two signatures has not been well-described, SBS8 has been associated with nucleotide excision repair (NER) deficiency in breast tumours and in *in vivo* and *in vitro* models (Jager et al. 2019). Given that the HR, NER and Fanconi anemia pathways play complementary roles in DNA repair, namely at DNA interstrand crosslinks and DNA-protein crosslinks, SBS8 may reflect multiple cellular processes (De Silva et al. 2000; Nakano et al. 2007).

The genomic landscape of tumours associated with constitutional variants in MMR genes and LOH showed a high mutational burden and MSI, consistent with their known function in maintaining genomic stability at microsatellite repeats across the human genome (Figure 2.3). These findings appeared independent of cancer type, as neither esophageal carcinoma nor lung adenocarcinoma are considered part of the spectrum of Lynch syndrome-related cancers. The molecular origins of Lynch syndrome in these cases included a large deletion spanning *EPCAM* and the 5' region of *MSH2* and constitutional methylation of the *MLH1* promoter. As expected, the Lynch syndrome-related tumours showed a high burden of small somatic mutations compared to tumours without germline inactivation in MMR genes. These tumours displayed

SBS signatures characteristic of MMR deficiency (SBS6 and SBS44) and signatures without

characterized associations with germline variants in MMR genes (SBS7c and SBS57) (Figure

2.3a). Tumours with both germline and somatic alterations in MMR genes also showed high

contributions from DBS3, putatively associated with somatic mutations in DNA polymerase

epsilon (*POLE*), and DBS7, ID2 and ID7, characteristic of MSI. The presence of other driver

mutations may underlie variability in the presence of specific mutational signatures, as

demonstrated for example in MMR-deficient tumours with *POLD1* mutations and cancer

genomes characteristic of COSMIC signature 20 (Haradhvala et al. 2018). While no *POLD1*

mutations were found, mutations in *POLL* and *POLE2* identified in a lung adenocarcinoma

associated with constitutional *MLH1* promoter methylation could not exclude possible

contributions from other members of DNA polymerase to the characteristic signature 20

(SBS44) in this case.

Figure 2.3. Genomic signatures of mismatch repair. a. Percent contribution of single base substitution signatures SBS6, SBS7c, SBS44 and SBS57, doublet base substitution signatures DBS3 and DBS7, and indel signatures ID2 and ID7. Black bars show the median within each group. **b.** Representative single base substitution mutation profile in an esophageal carcinoma associated with a germline structural variant (SV) at the locus of *EPCAM* and *MSH2* and LOH. The mutational signature in this case reflects a composition of several established COSMIC version 3 signatures, representative figures of which are shown below.

Consistent with global deficiency in MMR resulting from biallelic inactivation in *MLH1* and *MSH2*, both Lynch syndrome tumours showed genome-wide MSI (47% and 65%,

respectively). MSI was rare in MMR-proficient tumours and tumours with somatic homozygous loss of MMR genes, suggesting that the high burden of MSI in Lynch syndrome-related tumours was associated with long-term accumulation of genome instability. Supporting increased microenvironmental immune response associated with a high burden of somatic mutations and potential neoantigens, tumour infiltrating lymphocytes were abundant in the microsatellite unstable Lynch syndrome-related tumours, with T cell infiltration above the 94th and 97th percentiles across all tumours (Smyrk et al. 2001; Phillips et al. 2004). Notably, neither HR nor MMR deficiency was indicated by mutational signatures alone for one individual with metastatic lung adenocarcinoma and pathogenic germline variants in *BRCA1* and *MLH1*. These findings suggested that the presence of specific mutational signatures associated with cancer predisposition syndromes may be mediated by other individual-, gene- or tissue-specific factors.

Despite recent studies describing a somatic mutational signature associated with NER, higher relative contributions of SBS8 was not observed in advanced tumours with biallelic loss in *ERCC* genes (Jager et al. 2019). Mutational signatures associated with defects in BER have been previously described in biallelic *MUTYH* carriers with MAP, characterized by multiple colorectal adenomas and an increased risk for CRC, indicating a role for global BER deficiency in pancreatic carcinogenesis (Rashid et al. 2016; Thibodeau et al. 2019).

2.3.3   Clinical actionability of germline variants in an advanced cancer population

Within patient- or population-based genetic screening programs, identifying individuals with clinically actionable germline variants conferring defined increases in cancer risk may have immediate implications for clinical translation. Moderate- to high-penetrance germline variants were identified in 13.8% of adult participants in the POG program, including carrier status for

genes associated with autosomal recessive cancer predisposition syndromes. Variants in *BRCA1*, *BRCA2*, *ATM*, and *CHEK2* were among the most common, accounting for the majority of carriers in this cohort. Moderate-penetrance genes in particular may benefit from accurate assessment of family history to account for individual factors contributing to cancer susceptibility and guide recommendations for clinical management (Daly et al. 2017).

Among moderate-penetrance genes evaluated in the POG program, variants in *ATM* and *CHEK2* were the most common, identified 3% of all cases. Although cancer risk estimates for *ATM* have been best described in association with breast cancer, increased lifetime risk for several other cancers has been associated with carrier status for *ATM* (Hall et al. 2020). Prostate, gastric and pancreatic cancers all have a reported odds ratio greater than twofold among *ATM* carriers (Angèle et al. 2004; Roberts et al. 2012; Helgason et al. 2015). However, breast, prostate, gastric or pancreatic cancers accounted for only three of nine *ATM* carriers identified here. In contrast, a current or prior personal history of breast cancer was reported in nine of 14 *CHEK2* carriers and three of four *PALB2* carriers, consistent with lifetime risks for breast cancer of 37% and 35% by 70 years of age, respectively (Nevanlinna and Bartek 2006; Weischer et al. 2008; Antoniou et al. 2014). Recent studies also suggest that certain variants in moderate-penetrance genes are associated with greater increases in lifetime cancer risk that are clinically meaningful (Southey et al. 2016).

Therefore, we defined clinically actionable germline variants as pathogenic or likely pathogenic germline variants in cancer susceptibility genes with published and non-conflicting estimates of cancer risk. Based on considerations of other gene-specific clinical guidelines, 9.6% of participants in the POG program were found to be carriers of clinically actionable germline variants (Figure 2.4). The overall rate of referral to the BC Cancer Hereditary Cancer Program

for these cases was 76% (52 of 68), including individuals referred prior to participation in the

POG program for index or carrier testing ($n = 34$). Several carriers met clinical testing criteria for

rare high-penetrance autosomal dominant cancer syndromes, including FAP ($n = 1$), Lynch

syndrome ($n = 2$), and Li Fraumeni or Li Fraumeni-like syndrome ($n = 3$). Given the advanced

stage of disease among participants in the POG program, lack of referral may be attributed to

patient medical status, uninformative personal or family history, or undefined clinical guidelines.

Among individuals referred for clinical assessment with positive germline findings, 53% (27 of

52) were informed by germline GS in POG. The significance of these findings was further noted

for 67% of carriers (18 of 27) for clinically actionable germline variants identified through the

POG program who were not otherwise eligible for provincially-funded genetic testing based on

current clinical guidelines.
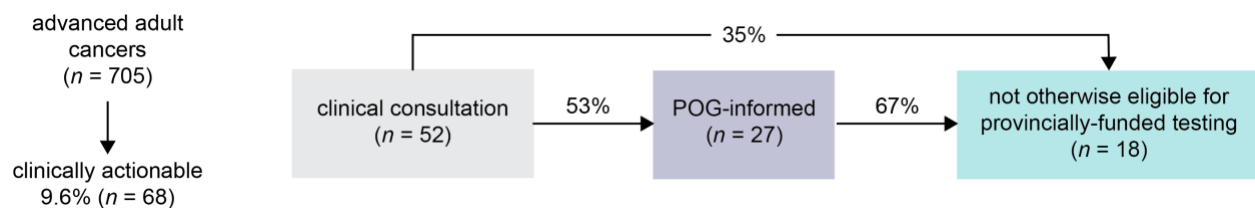
Figure 2.4. Summary of the impact of germline findings for cancer susceptibility in the Personalized OncoGenomics program

Carrier status for genes without clear estimates for autosomal dominant cancer risk was

indicated in 4% of individuals without other actionable germline findings. These included

variants in autosomal recessive genes, as well as genes with conflicting or minimal support for

cancer susceptibility, such as *BARD1* and *MRE11* (Tung et al. 2016). Personal and family

medical history should be carefully considered in these cases to determine the appropriateness of

variant disclosure for cancer susceptibility. Identification of low- to moderate-penetrance cancer predisposition variants may present an additional challenge due to limited evidence-based guidelines regarding effective clinical management and cancer risk reduction strategies. Similarly, the risks and benefits of returning carrier status for autosomal recessive cancer susceptibility genes should be evaluated on a gene- and case-specific basis.

2.3.4   Tumour-normal NGS in MMR-deficient tumours improves clinical management for potential Lynch syndrome families

Universal screening for MMR deficiency in CRC and EC is used to identify individuals with Lynch syndrome who may benefit from increased cancer screening, cancer prevention, or the use of targeted therapies. Germline variants are identified in 25-67% of individuals with MMR-deficient CRC, suggesting that remaining cases represent sporadic cancer occurrences or cases with cryptic germline variants undetected by standard clinical multigene sequencing (Hampel et al. 2005b; Hampel et al. 2008). Therefore, targeted tumour-normal genetic testing in cases with MMR-deficient Lynch syndrome-related cancers may improve molecular diagnosis of suspected families and identify sporadic cancer cases that may not require intensive cancer risk management strategies. Paired tumour-normal targeted NGS was performed in 76 cases referred to the BC Cancer Hereditary Cancer Program on the basis of an MMR-deficient Lynch syndrome-related cancer (Table 2.1). The majority of cancers were of primary colorectal (59%) or endometrial (32%) origin, but also included sebaceous adenomas ($n = 3$), small bowel cancers ($n = 2$), gastric cancer ($n = 1$), and renal cell carcinoma ($n = 1$).

Table 2.1. Clinical characteristics and screening results for individuals receiving targeted tumour-normal NGS by cancer type

			Cancer Type	
	Index cases (%)	Colorectal (%)	Endometrial (%)	Other (%)
Total	76	45	24	7
Family history				
Amsterdam I/II	4 (5)	3 (7)	1 (4)	0
Revised Bethesda	52 (68)	29 (64)	17 (71)	5 (71)
CRC or EC in 1° or 2° relatives	5 (7)	4 (9)	1 (4)	0
None	16 (21)	9 (20)	5 (21)	2 (29)
MMR IHC				
MLH1/PMS2	34 (45)	27 (60)	4 (17)	3 (43)
MSH2/MSH6	20 (26)	9 (20)	8 (33)	3 (43)
MSH6	8 (11)	1 (2)	6 (25)	1 (14)
PMS2	10 (13)	6 (13)	4 (17)	0
Other	4 (5)	2 (4)	2 (8)	0
Microsatellite instability				
Unstable	70 (92)	44 (98)	20 (83)	6 (86)
Stable	5 (7)	0	4 (17)	1 (14)
Not specified	1 (1)	1 (2)	0	0
MLH1 **hypermethylation**				
Absent	53 (70)	26 (58)	21 (88)	6 (86)
Present	23 (30)	19 (42)	3 (13)	1 (14)

Most individuals (68%) referred for tumour-normal testing met revised Bethesda criteria based on age of diagnosis and MMR deficiency identified through universal screening. In contrast, few individuals met Amsterdam I or II criteria on the basis of multiple primary tumours and/or strong family history of Lynch syndrome spectrum cancers. Overall, targeted germline genetic testing identified a pathogenic or likely pathogenic variant in 28% of cases, and constitutional methylation of *MLH1* was confirmed through orthogonal testing in one additional case meeting

Amsterdam II criteria (Table 2.2). Across the largest cancer cohorts in this study, Lynch syndrome was diagnosed in 24% and 29% of CRC and EC, respectively. A molecular diagnosis was determined for three of four cases meeting Amsterdam I or II criteria, while three carriers did not meet any clinical testing criteria. PMS2 deficiency by IHC was associated with the highest prevalence of pathogenic and likely pathogenic germline variants compared to the other MMR genes, identified in 80% (8 in 10) of PMS2-deficient cancers. The remaining two cases were likely sporadic based on the presence of double somatic mutations in *PMS2*.

Table 2.2. Clinical characteristics and screening results for individuals receiving targeted tumour-normal NGS by predicted origin

	Predicted origin			
	Germline (%)	Sporadic (%)	DNAme (%)	Uncertain* (%)
Total	21	21	17	17
Family history				
Amsterdam I/II	3 (14)	0	0	1 (6)
Revised Bethesda	15 (71)	16 (76)	7 (41)	13 (76)
CRC or EC in 1° or 2° relatives	0	2 (10)	3 (18)	0
None	3 (14)	3 (14)	7 (41)	3 (18)
Cancer type				
CRC	11 (52)	12 (57)	14 (82)	8 (47)
EC	7 (33)	9 (43)	2 (12)	6 (35)
Other	3 (14)	0	1 (6)	3 (18)
MMR IHC				
MLH1/PMS2	5 (24)	8 (38)	16 (94)	5 (29)
MSH2/MSH6	5 (24)	8 (38)	0	7 (41)
MSH6	2 (10)	2 (10)	0	4 (24)
PMS2	8 (38)	2 (10)	0	0
Other	1 (5)	1 (5)	1 (6)	1 (6)
Microsatellite instability				
Unstable	20 (95)	20 (95)	17 (100)	13 (76)
Stable	1 (5)	0	0	4 (24)
Not specified	0	1 (5)	0	0
***MLH1* hypermethylation**				
Absent	18 (86)	19 (90)	0	16 (94)
Present	3 (14)	2 (10)	17 (100)	1 (6)

*Tumours with uncertain origin include cases with MMR protein loss that could not be explained by *MLH1* promoter methylation or two somatic hits (e.g. double somatic mutations, somatic mutations with loss of heterozygosity, or somatic mutations with hypermethylation) in genes consistent with IHC results.

DNAme, DNA methylation

Notably, tumour-normal sequencing confirmed a sporadic cancer occurrence in 28% of cases, mediated by double somatic mutations, copy number alterations and/or LOH. Hypermethylation of the *MLH1* promoter, which is often associated with sporadic CRC and presence of the oncogenic *BRAF* V600E mutation, further indicated a likely sporadic cancer occurrence in 22% of cases overall and 47% of MLH1-deficient tumours (Weisenberger et al. 2006). Current clinical guidelines do not recommend increased cancer screening for individuals with tumour *MLH1* hypermethylation unless otherwise indicated by strong or suspicious personal and/or family history (Poynter et al. 2008). Although the majority of cases with pathogenic or likely pathogenic germline findings did not show hypermethylation at the *MLH1* promoter, *MLH1* hypermethylation was identified in tumours from two *PMS2* carriers. While MLH1/PMS2 deficiency in one case without a characterized second hit suggested that methylation of *MLH1* was an early event driving tumour initiation, retained MLH1 expression and biallelic inactivation of *PMS2* in the other case indicated that *MLH1* hypermethylation may have been a later event in tumour progression. Among genetically confirmed Lynch syndrome cases, second somatic mutations or LOH were identified in 62%. This included one index case meeting multiple revised Bethesda criteria and found to carry *MSH2* H610P. Although initially classified as a VUS at the time of referral, this variant was subsequently reclassified based on its association with tumour LOH, MSH2 deficiency and MSI in multiple individuals meeting phenotype-based Lynch syndrome criteria.

Based on these findings, we developed an evidence-based framework for integrating tumour-normal sequencing into a provincial Lynch syndrome screening program (Figure 2.5). Tumour sequencing may also inform the reclassification of VUS, particularly in the context of a clinical diagnosis of Lynch syndrome, by providing a functional read-out of MMR-associated

mutation patterns or identifying loci showing LOH. Germline NGS should not be excluded on the basis of *MLH1* hypermethylation or double somatic mutations alone for individuals with personal and/or family history suggestive of Lynch syndrome. Cases without germline variants or double somatic alterations may represent sporadic cases resulting from other genetic or epigenetic events, such as somatic methylation of the *MSH2* promoter, a frequent somatic event not captured by DNA sequencing (Nagasaka et al. 2010). Alternatively, these cases may harbour pathogenic germline variants that are undetectable by targeted NGS. Among the cases presented here, germline or sporadic origin could not be confirmed in 22%, including one index case meeting Amsterdam I criteria. In such cases, GS could be considered in order to characterize potential non-coding or structural variants that remain undetectable by panel-based sequencing.
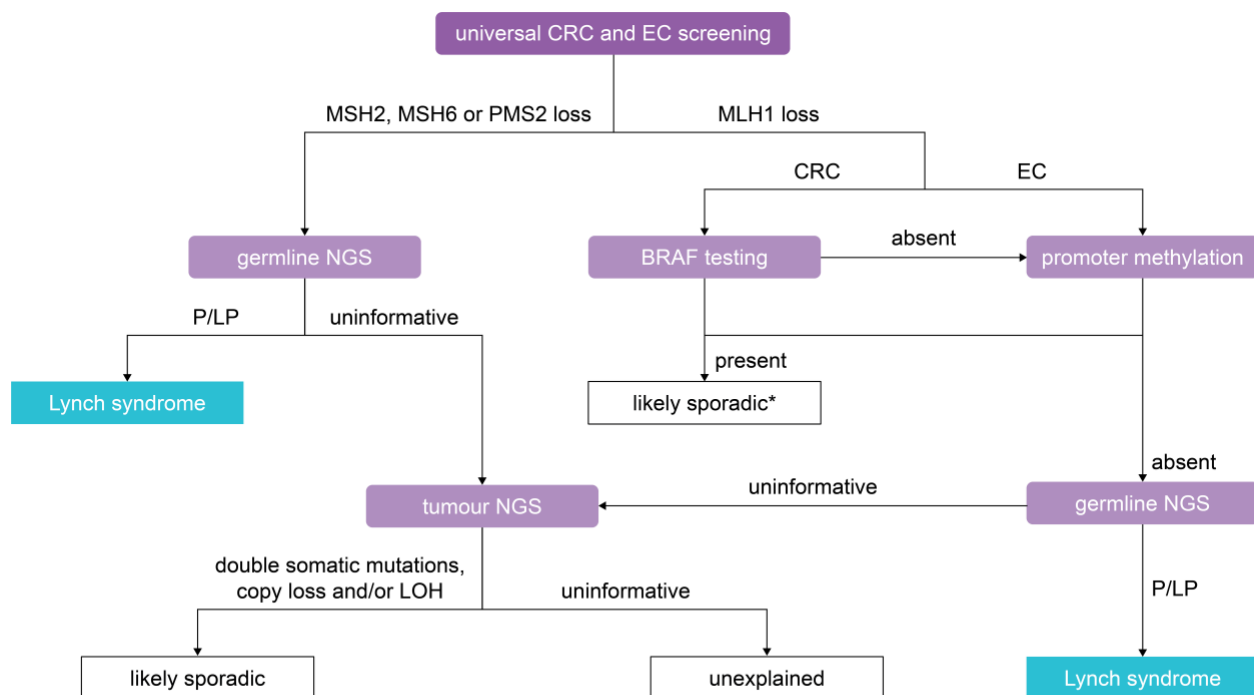
*germline NGS and/or constitutional promoter methylation testing should be performed if indicated clinically

Figure 2.5. Modified framework for Lynch syndrome screening in MMR-deficient CRC and EC cases. CRC: colorectal cancer; EC: endometrial cancer; LOH: loss of heterozygosity; LP: likely pathogenic; NGS: next-generation sequencing; P: pathogenic.

2.4   Discussion

Identifying tissue-specific molecular alterations associated with constitutional genetic variation can provide a better understanding of the biological pathways underlying disease pathogenesis. This may ultimately allow opportunities for the development of targeted therapies or targeted screening strategies aimed at identifying early lesions in patients at high risk. Accordingly, molecular signatures of the genome, transcriptome and proteome have been described in infectious disease, neurodegenerative disorders and in cancer (Campbell and Ghazal 2004; Diaz-Castro et al. 2019). Cancer genomes showing an accumulation of specific patterns of DNA damage may implicate endogenous defects in DNA repair caused by pathogenic germline

42

variants in one of several high-penetrance cancer predisposition genes. However, distinct mutational patterns have not been associated with other high-penetrance genes, thus requiring alternative approaches to studying cancer progression in individuals with cancer predisposition syndromes.

Molecular tumour signatures in HBOC and Lynch syndrome have been the best characterized to date, with current implications for carrier identification, variant interpretation, and therapeutic selection. *BRCA1* carriers in particular show an increased incidence of hormone receptor-negative breast cancers that can improve ascertainment of potential HBOC families in unselected patient populations. Tumours with germline inactivation of *BRCA1* and *BRCA2*, associated with genome-wide HR deficiency, show a better response to platinum-based therapies and PARP inhibitors (Kaufman et al. 2015).

Pathogenic germline variants in *PALB2* and *RAD51C* were recently shown to be associated with similar genome-wide signatures of abnormal HR in breast cancers (Polak et al. 2017). These observations support known roles for BRCA1, BRCA2 and PALB2 during HR in facilitating the displacement of the single-stranded DNA binding protein RPA and recruitment of RAD51 to DNA damage-induced nuclear foci (Scully et al. 1997; Sharan et al. 1997). We showed that germline *PALB2*-associated HR deficiency may not be limited to breast cancers, occurring in three *PALB2*-related breast tumours and a *PALB2*-related cholangiocarcinoma. These findings were in contrast to tissue-specific contributions of *BRCA1*- and *BRCA2*-related HR deficiency, characterized by strong contributions from COSMIC SBS3 and SBS8, in breast cancers compared to other tumour types. Differential gene expression between tissues has not explained the increased risk for certain cancer types associated with carrier status in cancer predisposition genes (Lage et al. 2008; Schneider et al. 2017). Our results support hypotheses

that cell type- or tissue-specific pathway activity may at least partially confer susceptibility to site-specific cancers.

Germline variants in moderate- and high-penetrance cancer predisposition genes show variable expressivity and incomplete penetrance of cancer and non-cancer phenotypes. As a result, many carriers of clinically actionable genetic variation may not be identified through current models of ascertainment based on personal and family cancer history. Among carriers of actionable germline variants who were identified through the POG program, 35% who received clinical genetics assessment did not meet current clinical criteria for provincially-funded genetic testing. Conversely, many individuals with clinical phenotypes suggestive of inherited cancer susceptibility do not receive a molecular diagnosis by standard clinical testing. Tumour-normal sequencing thus offers opportunities for improved diagnosis by identifying somatic alterations associated with inherited deficiencies in relevant cancer pathways. This was significant for two individuals with familial breast cancer and global tumour HR deficiency, each with uninformative clinical genetic testing at the time of initial referral. Retrospective germline analysis revealed that these were related to previously uncharacterized pathogenic variants in *BRCA2* and *PALB2*, ultimately allowing for cascade carrier testing in family members.

Lynch syndrome-related tumours are caused by pathogenic germline variants in *MLH1*, *MSH2*, *MSH6* and *PMS2*, leading to global MSI and a high burden of somatic mutations. Due to the potential for encoding many tumour-specific antigens, MMR-deficient tumours are sensitive to immune checkpoint blockades, such as the PD-1 blockade pembrolizumab (Le et al. 2015; Le et al. 2017). These features appear independent of primary tumour origin, shown here through genome-wide mutation analysis in esophageal carcinoma and lung adenocarcinoma in two unrelated Lynch syndrome patients. Accordingly, MSI may indicate germline testing for Lynch

syndrome across a broader spectrum of cancer types (Latham et al. 2019). Notably, HR- and MMR-deficient tumours associated with germline variants show greater contributions of their respective mutational signatures compared to tumours with somatic inactivation in these genes. In individuals with biallelic variants in MMR genes, cancers are associated with ultra-hypermutation and distinct mutational signatures compared to cancers associated with monoallelic germline variants or somatic mutations in those genes (Shlien et al. 2015).

According to the revised Bethesda guidelines, all CRC and EC patients under 60 years of age with an MMR-deficient tumour are eligible for clinical genetic testing regardless of family history (Umar et al. 2004). The identification of sporadic cancer cases through tumour sequencing may improve clinical management for individuals with a history of an MMR-deficient cancer (Hampel et al. 2018). Among index cases receiving tumour-normal sequencing for MMR-deficient Lynch syndrome-related cancers, 50% of all cases were likely sporadic cancers caused by double somatic alterations or *MLH1* methylation. These findings helped inform a modified framework for Lynch syndrome testing, incorporating tumour sequencing as a secondary test following uninformative germline testing. Similar approaches have been investigated in unselected patients with CRC and EC, together indicating that identifying likely sporadic cancer occurrences may reduce the need for intensive cancer screening in patients who do not have an otherwise suspicious personal or family cancer history (Hampel et al. 2018; Salvador et al. 2019). In cases with phenotypic indications of Lynch syndrome, more comprehensive germline testing may be performed to exclude other possible causal variants. Future studies are needed to assess the clinical utility, positive and negative predictive values and cost-effectiveness of tumour sequencing in the genetic diagnosis and clinical management of hereditary cancer families.

Chapter 3: Tissue-specific modeling aids in molecular characterization of hereditary diffuse gastric cancer

3.1   Introduction

HDGC is an inherited cancer predisposition syndrome associated with a significant lifetime risk of diffuse gastric cancer (DGC) and lobular breast cancer. Genetic testing is recommended for individuals and families meeting clinical testing guidelines defined by the International Gastric Cancer Linkage Consortium (IGCLC), and pathogenic germline variants in *CDH1* are identified in 25-40% of HDGC families (Guilford et al. 1998; van der Post et al. 2015). Carriers have a 56-70% cumulative risk of DGC by 80 years, while female carriers also have a 42% cumulative risk of lobular breast cancer (Hansford et al. 2015). Notably, the average age of gastric cancer onset in HDGC is 38 years, which necessitates intensive cancer screening and prophylactic surgery in carriers and at-risk family members. HDGC accounts for 1-3% of all gastric cancer and is associated with high grade, advanced stage and poor prognosis at the time of diagnosis. Therefore, understanding the biological mechanisms underlying the early stages of disease may help identify strategies to improve early cancer detection and inform the development of targeted therapies.

Given the distinct genomic signatures in tumours with inherent DNA repair deficiency, I sought to investigate transcriptional signatures associated with inactivation of *CDH1*, a characterized tumour suppressor gene encoding the cell-cell adhesion molecule E-cadherin. HDGC is characterized by the presence of signet ring cells (SRCs), defined as cells with prominent cytoplasmic mucin and crescent-shaped nuclei, indicating that *CDH1* inactivation is associated with specific cellular and morphological features (Pernot et al. 2015). These may

result from molecular alterations in normal gastric epithelium that precede neoplastic transformation and tumour development. I hypothesized that homozygous loss of *CDH1* is associated with differential transcriptional regulation in normal gastric epithelium. Using a murine organoid model of HDGC and single cell RNA sequencing (scRNA-seq), we explored the single-cell landscape of *Cdh1*-deficient gastric organoids across gastroesophageal cell lineages and investigate candidate biomarkers of early SRC lesions in *CDH1* carriers.

3.2   Methods

3.2.1   Single-cell RNA sequencing

Murine gastric organoids were generated from *CD44*-cre/*Cdh1loxP/loxP*/*tdTomato* mice as previously described (Bougen-Zhukov et al. 2019). Briefly, organoids were generated from minced stomach tissue obtained from neonatal mice using air-liquid interface culture with myofibroblast co-culture. *Cdh1* deletion was induced at day 0 post-seeding with 5 μM endoxifen in 3 mL of growth medium containing F-12, 20% fetal bovine serum (FBS) and 50 μg/mL Gentamycin. An equivalent volume of DMSO was added to controls. Organoids were disaggregated into single cell suspension with trypsin and stored in cryopreservation medium containing growth medium, FBS, DMSO and ROCK inhibitor (Y-27632) for transportation. Thawed cell suspensions were washed twice with FBS and filtered through a cell strainer prior to incubation in DAPI for fluorescent-activated cell sorting to enrich for live cells. Single cells were loaded on the 10x Genomics Single Cell Controller and 3' scRNA-seq libraries were prepared using the Chromium Single Cell 3' Reagent v2 Chemistry Kit (10x Genomics) according to the manufacturer's protocol. Libraries were sequenced on the Illumina NextSeq, and demultiplexing

and sequence alignment were performed using Cell Ranger 2.0 (10x Genomics). Sequencing

metrics are provided in Table 3.1.

Table 3.1 Sequencing metrics for mouse gastric organoids analyzed by single-cell RNA sequencing

Replicate	Condition	Cells sequenced	Mean reads per cell	Median genes per cell
1	KO	4,036	150,641	4,833
	WT	5.357	98,010	4,415
2	KO	2,323	106,784	5,081
	WT	1,588	155,199	5,108

3.2.2   Data processing

Knockout and wild-type cells within each experiment were processed together. Cell- and

gene-level quality metrics were calculated using the R package scater (McCarthy et al. 2017).

Cells were excluded from analysis if the library size or number of detected genes was more than

three mean absolute deviations below the median of all cells or if the percentage of

mitochondrial transcripts was more than three mean absolute deviations above the median of all

cells. Genes that were not expressed in any cells were removed. To reduce cell-specific

sequencing or capture bias, prescaled size factors were calculated following the removal of poor

quality cells using the R package scran using rough clusters generated from raw transcript counts

(Lun et al. 2016). Log normalization was then performed based on cell-specific size factors using

scater.

3.2.3 Clustering and marker identification

Datasets from two independent experimental replicates were combined prior to cluster analysis and marker identification, rescaling counts to account for biases in library size using batchelor (Haghverdi et al. 2018). Adjustment of size factors and log-normalized counts was performed for a common set of transcripts expressed in both datasets, and genes with positive biological components of variance in log-transformed expression across replicates were selected for principal component analysis and batch correction ($n = 6,426$). Batch correction was performed using the mutual nearest neighbour (MNN) method. Graph-based clustering of single cells was performed with scran, selecting iterative values for cluster density (k) and using highly variable genes described above. A minimal cluster solution was selected by identifying the smallest number of clusters that could distinguish between outlier cell populations, resulting in 12 clusters (k = 25). Between individual cell clusters and larger cell populations, defined by the expression of known marker genes, genes showing a greater than twofold relative expression compared at least one other cluster with a false discovery rate (FDR) ≤ 0.01 were identified using scran.

3.2.4 Differential expression

Differential expression was evaluated in 9,834 epithelial cells (clusters 2, 3, 4, 5, 6, 7, 8, 10 and 12) using MAST (Finak et al. 2015). Adjusting for batch effects between independent experiments, we fit a hurdle model using raw count data to compare differences between *Cdh1* knockout and wild-type cells. Statistical significance of gene-wise comparisons between conditions was estimated using a likelihood ratio test. This procedure was subsequently performed within each major epithelial cell type: basal cells (clusters 6, 7 and 10), suprabasal

cells (clusters 2, 5 and 12), suprabasal-like cells (clusters 4 and 8), and mucous progenitor cells (cluster 3). Gene ontology analysis was performed for genes with an absolute fold-change ≥ 1.5 and P-value ≤ 0.01 using the R package ReactomePA.

3.2.5   Immunohistochemistry

Use of human tissue specimens in this study was approved by the BC Cancer Research Ethics Board (H19-02571). Staining was performed on the Ventana platform using monoclonal antibodies for CXCL5 (#MAB254, R&D Systems), CXCL7 (#MAB393, R&D Systems), IGFBP-3 (#MAB305, R&D Systems). Staining for CK7 (#GA619, Dako) and CK19 (#GA615, Dako) was performed on the Dako Omnis platform according to standard clinical protocols. Expression of candidate markers was evaluated in a gastric cancer tissue microarray and in whole gastric tissue sections from known *CDH1* carriers. To account for heterogeneous staining within gastric cancer cores, semiquantitative scoring considering both intensity of staining and percentage of positive cells was used (Remmele and Stegner 1987; Fedchenko and Reifenrath 2014). Intensity of each case was assigned a score of 0 (no staining), 1 (light), 2 (moderate), or 3 (strong), while percent positivity for each value of intensity was assigned a score of 0 (0%), 1 ($\leq$10%), 2 (11-50%), 3 (51-80%), or 4 (>80%).

3.3   Results

3.3.1   Distinguishing epithelial cell types in murine gastric organoids

Conditional *Cdh1* knockout organoids were generated from stomach tissue obtained from *CD44*-cre/Cdh1loxP/loxP/*tdTomato* neonatal mice as previously described (Nouri 2019). A total of four organoids from two independent experimental replicates, each including one uninduced

Cdh1 wild-type (*Cdh1wт*) and one *Cdh1* knockout (*Cdh1-/-*) organoid, were analyzed by scRNA-seq on the 10x Genomics Chromium platform. Excluding low-quality cells, a total of 11,364 cells were used for downstream analysis (Figure 3.1a-b). To characterize biologically distinct cell subpopulations, batch correction and clustering of single cells from both wild-type and knockout organoids was performed. We observed several populations representative of early gastrointestinal cell lineages, which were broadly categorized as epithelial cells, marked by the expression of *Epcam*, and non-epithelial cells, marked by the expression of vimentin (*Vim*) (Figure 3.1c). Non-epithelial cell clusters could be further resolved by the expression of the fibroblast-specific genes *Bgn* and *Col1a1* and the leukocyte marker *Ptprc*. Single cells within fibroblast and leukocyte cell clusters did not aggregate into condition-specific subclusters, suggesting that conditional deletion of *Cdh1* did not affect the global transcriptome of non-epithelial cells. Universal proliferation markers *Birc5* and *Mki67* were highly expressed in a small population of epithelial cells (cluster 6) and fibroblasts (cluster 11), independently of *Cdh1* status (10.3% *Cdh1-/-* and 11.5% *Cdh1wт*). This finding was recently reported in another gastric organoid model suggesting that around 6% of fibroblasts are actively proliferating (Chen et al. 2019).
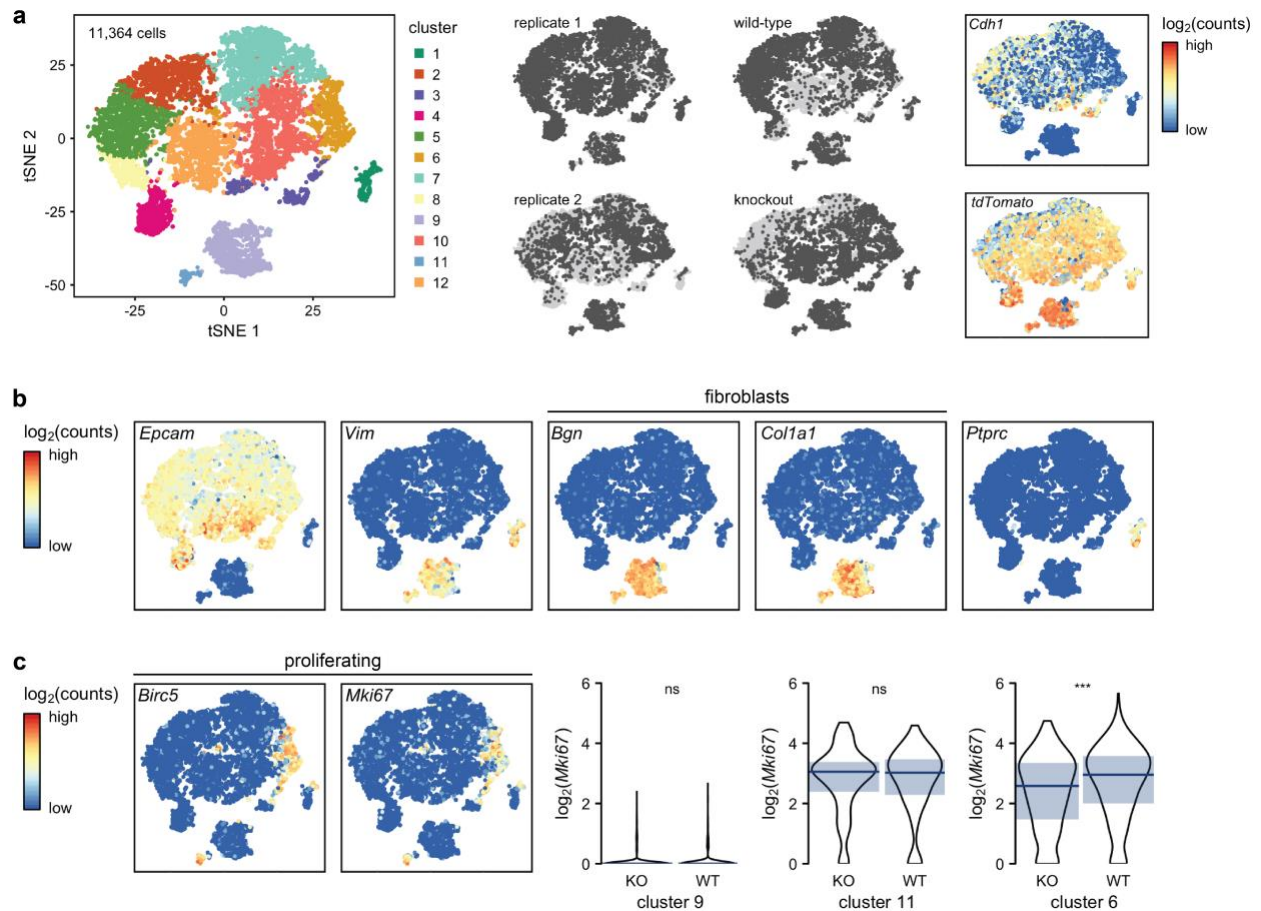
Figure 3.1. Single-cell RNA sequencing of *Cdh1WT* and *Cdh1-/-* murine gastric organoids. a.
T-distributed stochastic neighbor embedding (tSNE) projection of 11,364 high-quality single

cells sequenced across four murine gastric organoids, coloured by cell cluster, replicate,

condition and normalized expression of *Cdh1* and *tdTomato*, a fluorescent marker indicating

successful deletion of *Cdh1*. **b.** Normalized expression of epithelial (*Epcam*), non-epithelial

(*Vim*), fibroblast (*Bgn*, *Col1a1*) and leukocyte (*Ptprc*) marker genes differentiates between

epithelial and non-epithelial cell clusters. **c.** tSNE and violin plots showing the distribution of

Mki67 expression in fibroblast cells (clusters 9 and 11) and epithelial cell cluster 6. The median

level of expression is shown by horizontal bars, and shaded boxes show the respective

interquartile range for each cluster. *P*-value, Wilcoxon rank-sum test: ns, not significant; ***, *P* ≤ 0.001.

The mature mammalian stomach is composed of glands with varying contributions from several cell lineages. Four major differentiated epithelial cells types exist in the adult stomach and express lineage-restricted marker genes: surface mucous or pit cells (*Muc5ac*), parietal or mucous gland neck cells (*Muc6*), zymogenic chief cells (*Atp4b*), and enteroendocrine cells (*Gast*). Given that stomach maturation, gland development, and gastric cell lineage differentiation are incomplete at birth in mice, these differentiated cell markers were not expressed in our gastric organoid model. However, all of the epithelial cell clusters expressed *Sox2*, a transcription factor required for early foregut development, and *Foxq1*, a transcription factor necessary for *Muc5ac* expression in mature gastric pit cells and for gastric acid secretion in parietal cells (Verzi et al. 2008). Neither the chief cell lineage-restricted transcription factor *Mist1* (*Bhlha15*) nor enteroendocrine progenitor-specific transcription factor *Neurog3* were expressed, suggesting that the majority of cells derived from *Cdh1loxP/loxP* organoids had been specified to the mucous-secreting cell lineages (Bjerknes and Cheng 2006; Lennerz et al. 2010). Therefore, this model likely reflects the early stages of gastric development and gastric cell lineage specification.
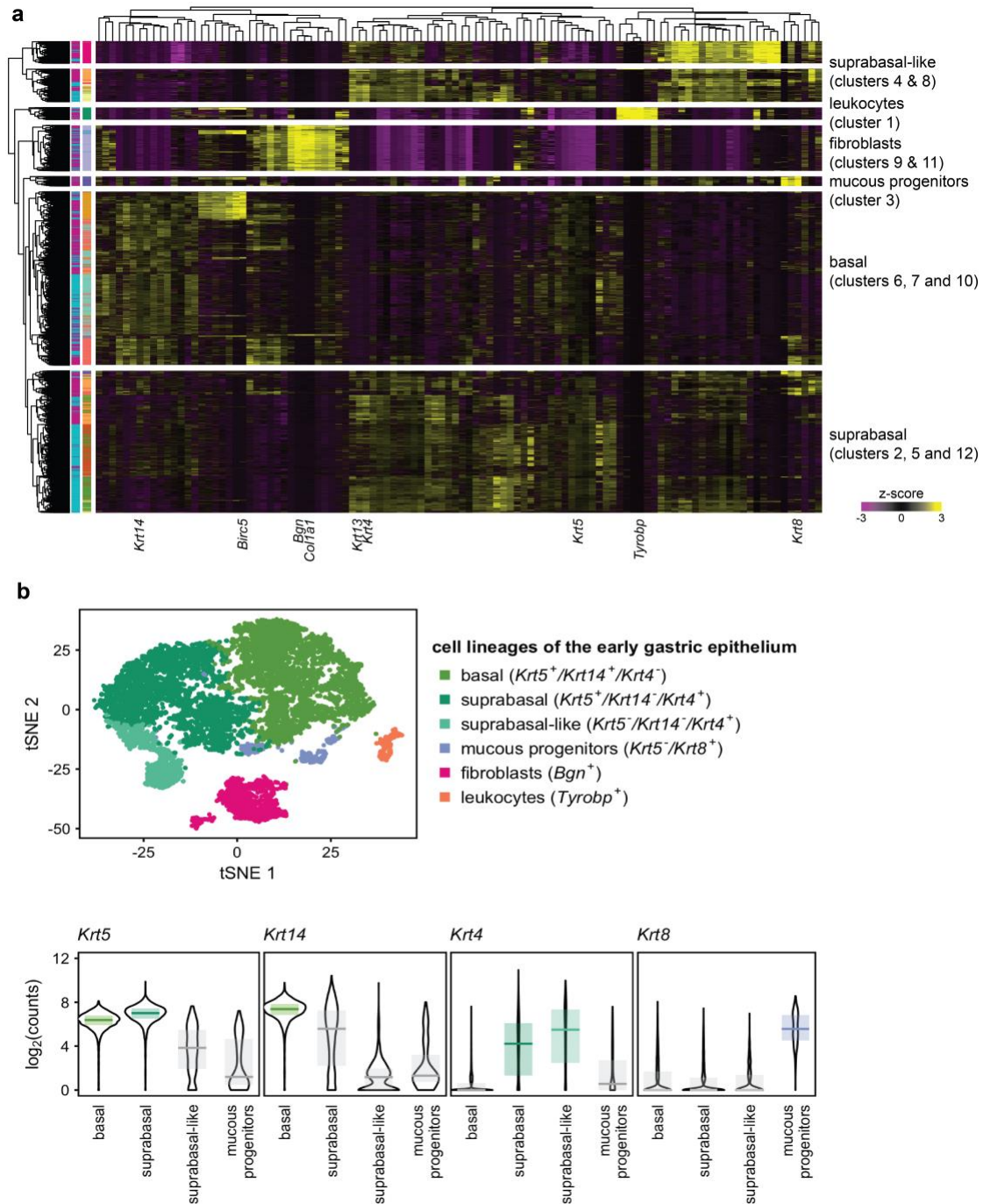
Figure 3.2. Lineage-specific genes of stratified squamous epithelium distinguish basal, suprabasal and mucous-secreting cell progenitors in mouse gastric tissue organoids. a. Heatmap showing scaled normalized expression of candidate cluster-specific marker genes. **b.** Original tSNE projection showing predicted cell populations based on the differential expression of basal epithelial marker *Krt5*, basal-specific marker *Krt14*, suprabasal-specific marker *Krt4* and luminal marker *Krt8*.

The proximal mouse stomach is lined by keratinized squamous epithelium, while the distal glandular stomach is composed of columnar cells. The squamous-columnar junction, which in humans is located between the esophagus and stomach, is characterized by the differential expression of cytokeratins 5 (*Krt5*), 7 (*Krt7*), and 8 (*Krt8*) (Jiang et al. 2017). *Krt5*, a marker of basal and suprabasal cell layers in stratified squamous epithelium, was almost universally expressed in epithelial cell clusters from the murine gastric organoids (Figure 3.2a) (Ramaekers et al. 1987; Evans et al. 2001). Among *Krt5+* cells, we further identified distinct subpopulations of cells expressing the basal cell markers *Trp63* and *Krt14* and suprabasal cell markers *Krt13* and *Krt4* (Figure 3.2b,c) (Bragulla and Homberger 2009). We observed discrete clusters of wild-type and *Cdh1* knockout cells in both *Krt14+* basal cells (clusters 6, 7 and 10) and *Krt13+/Krt4+* suprabasal cells (clusters 2, 5, and 12). Expression of the luminal cell marker *Krt8* and cardia mucosa marker *Cldn18* were limited to a small cluster of cells (cluster 3) that also expressed the trefoil factor *Tff1* and gastrokine *Gkn1*, markers of mucous-secreting cells (Jovov et al. 2007). *Cldn18* expression is restricted to lung and stomach epithelial cells, each expressing tissue-specific isoforms through alternative splicing (Niimi et al. 2001). This cluster was suspected to represent a small population of differentiating cells destined to contribute to the glandular hindstomach.

3.3.2   Putative biomarkers in hereditary diffuse gastric cancer

Given the general distinction between wild-type and knockout cells within the epithelial cell populations, differential expression of several genes was associated with *Cdh1* loss in the gastric organoids (Figure 3.3a-c). Compared to wild-type cells, 133 transcripts showed significant aberrant expression in *Cdh1-/-* cells (P-value ≤ 0.01, absolute fold-change ≥ 1.5).

Among 67 genes with lower expression in knockout cells, several were involved in the regulation of cell proliferation, response to stress, or chemotaxis. 66 genes showing higher expression compared to wild-type cells were enriched for genes with biological roles in the humoral immune response and cell differentiation and development. Although these findings indicated that *Cdh1* loss was associated with differential regulation of genes involved in various nonspecific cellular processes, several genes with known or potential clinical significance were identified among those showing higher expression in knockout cells. These included the cytokeratins *Krt7* and *Krt19*, chemokines *Cxcl5* and *Cxcl7* (*Ppbp*), and insulin-like growth factor *Igfbp3*, which were selected for evaluation by IHC in human tissues (Figure 3.3d,e).
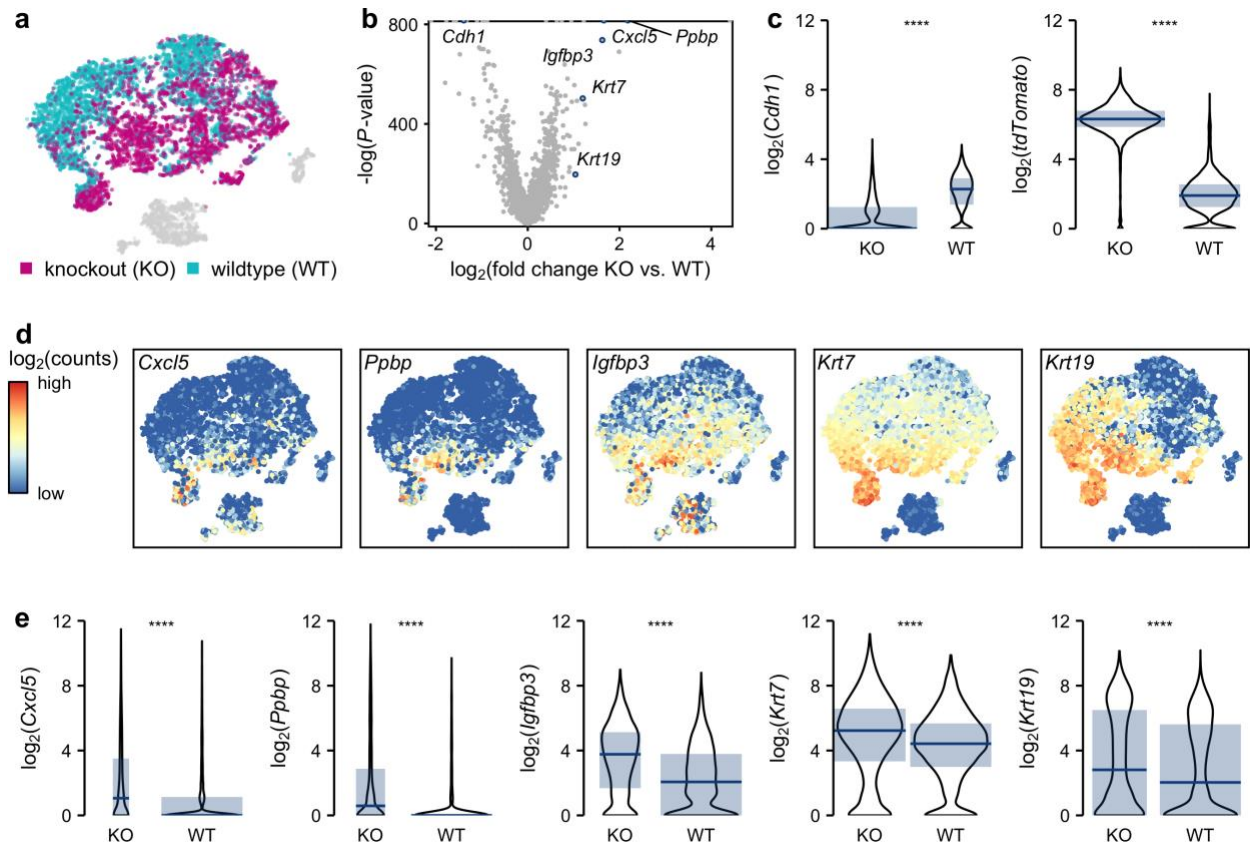
Figure 3.3. Global differential expression in *Cdh1* knockout single cells from mouse gastric organoids. a. Spatial distribution of *Cdh1* knockout and wild-type single cells in an integrated analysis of high-quality mouse stomach tissue organoids. **b.** Volcano plot of differential gene expression assessed through pseudo-bulk analysis of single cells. Fold-change and *P*-values are shown for knockout cells relative to wild-type cells. ****, $P \leq 0.0001$. **c.** Expression of *Cdh1* and *tdTomato* in knockout and wild-type cells. **d.** Log-transformed and normalized counts of five candidate marker genes upregulated in *Cdh1* knockout cells. **e.** Expression of candidate marker genes in knockout and wild-type cells. N.B. Due to the high number of replicates (individual cells) inherent in single-cell data, *P*-values may appear inflated and should be interpreted cautiously.

Results of IHC performed across 202 gastric cancer cases in a human tissue microarray are summarized in Table 3.1. KRT7 (CK7) is a broadly used clinical marker in gastrointestinal pathology. KRT7+ specimens distinguish primary gastric tumours from tumours of other gastrointestinal origins, shown in the differential diagnosis of metastatic gastric and colorectal cancers through the differential expression of KRT7 and KRT20 (Park et al. 2002). As expected, KRT7 did not differentiate between intestinal and diffuse gastric cancers evaluated in the gastric cancer microarray ($P = 0.2873$). While absent to patchy expression of KRT7 has been reported in the gastric cardia, KRT7 expression is absent from the gastric body and antrum (Couvelard et al. 2001; Jovanovic et al. 2002; Mohammed et al. 2002). IHC in prophylactic gastrectomy specimens from human *CDH1* carriers revealed negative or patchy expression of KRT7 in normal gastric epithelium, with strong cytoplasmic staining of SRCs against predominantly negative normal background gastric glands in one case (Figure 3.4). In contrast, both normal and cancer tissues stained strongly for KRT19 (CK19), consistent with its universal expression in epithelial tissues.

Table 3.2. Summary of immunohistochemistry results for four candidate markers in a microarray of human gastric cancer tissues

	Total number of cases	Number of positive cases (%)			
		CXCL7	IGFBP-3	KRT7	KRT19
Total	202	48 (24)	63 (31)	164 (81)	198 (98)
Intestinal	119	30 (25)	42 (35)	101 (85)	117 (98)
Diffuse	55	10 (18)	11 (20)	43 (78)	53 (96)
Mixed	28	8 (29)	10 (36)	20 (71)	28 (100)

Several genes involved in humoral immune response upregulated in knockout cells, including *Cxcl5*, *Cxcl7* and *Igfbp3*, did not show a specific association with SRC morphology or

diffuse gastric cancer in human tissues (Figure 3.4). The presence of cells expressing CXCL7

and IGFBP-3 in both normal and malignant cells in the gastric glands suggest that these ligands

may play a role in mediating the immune response in gastric epithelium. Staining for these

markers was often restricted to a few glands or to individual cells within a gland. Between

intestinal and diffuse gastric cancer subtypes, a higher proportion of intestinal-type gastric cancer

were positive for IGFBP-3, although this finding was not significant ($P = 0.0513$). Together with

the absence of a notable association between CXCL5, CXCL7 and IGFBP-3 with early invasive

SRC carcinoma, these findings suggest a nonspecific response through activation of

inflammatory or immune response pathways. Further studies of the molecular characteristics of

CXCL7+ and IGFBP-3+ cells in the gastric gland may be warranted given the potential

application of chemokines in cancer therapeutics (Nagarsheth et al. 2017).
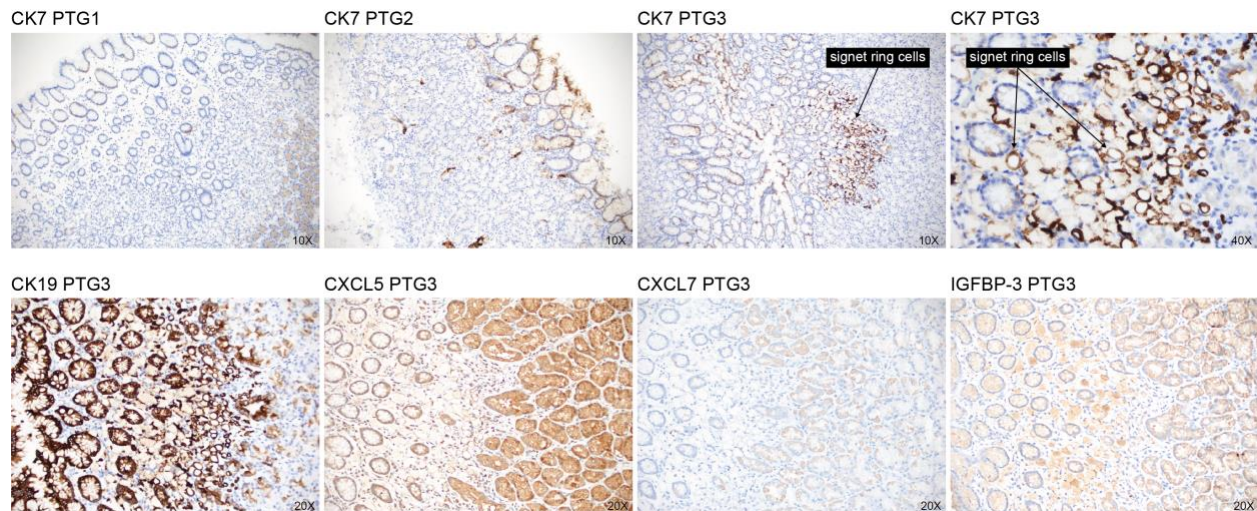
Figure 3.4. CK7 shows strong staining of early invasive signet ring cells in a human *CDH1*

carrier. Immunohistochemistry (IHC) for KRT7 (CK7) was performed in sectioned formalin-

fixed paraffin-embedded biopsy specimens from three known *CDH1* carriers. Signet ring cells

were identified in a representative section from one carrier (PTG3), and representative images

from two carriers not known to have early neoplastic lesions are shown (PTG1 and PTG2). Representative images of KRT19 (CK19), CXCL5, CXCL7 and IGFBP-3 IHC in PTG3.

3.3.3 Deregulation of luminal and basal cell markers is associated with *Cdh1* loss

KRT7 has been associated with intestinal metaplasia and has a characterized role in the pathogenesis of Barrett's esophagus (Mohammed et al. 2002; Shen et al. 2002). In murine models of Barrett's esophagus, Krt7 expression is restricted to the transitional epithelium, marked by the expression of squamous cell markers Krt5 and Trp63, at the squamous-columnar junction (Wang et al. 2011; Jiang et al. 2017). Trp63 is critical for normal development of the esophageal epithelium and undergoes sequential changes in expression during the differentiation of progenitor cells into mature basal cells (Daniely et al. 2004). In epithelial cell of the mouse gastric organoids, *Krt7* expression was negatively associated with the expression of *Trp63*, the expression of which is restricted to progenitor populations from both the developing forestomach and hindstomach in mice (Spearman's correlation = -0.51). This observation is consistent with the absence of *Krt7* expression in the proximal mouse stomach (Wang et al. 2011; Jiang et al. 2017).

Given the upregulation of *Krt7* in *Cdh1$_{-/-}$* cells, we evaluated differential gene expression between knockout and wild-type cells within each of the four major epithelial cell populations (Figure 3.5a). Overall, *Krt5$_-$/Krt14$_-$/Krt4$_+$* suprabasal-like cells had the highest number of differentially expressed genes compared to other epithelial cell populations. These cells clustered independently from basal and suprabasal cells, suggesting that they may reflect a unique and less well-defined subpopulation susceptible to genetic loss of *Cdh1* (Figure 3.2a). Notably, previously detected marker genes were differentially regulated in knockout cells across the

various epithelial subpopulations (Figure 3.5b). Mucous progenitor cell-specific markers were especially influenced by *Cdh1* loss, as the proportion of differentially expressed transcripts in cell type-specific markers was more than twofold greater than differentially expressed markers that were not restricted to this population. The luminal cell marker *Krt8*, the expression of which was restricted to mucous progenitor cells, was upregulated in *Cdh1-/-* cells across basal and suprabasal epithelial cell populations. Corresponding downregulation of *Trp63* suggested that loss of *Cdh1* may promote a luminal-like pattern of gene expression.

During development, prostate stem cell antigen *PSCA* has been shown to be expressed in differentiating gastric epithelial cells and is highly expressed in the adult human stomach (Bahrenberg et al. 2000; Sakamoto et al. 2008). Intriguingly, genetic polymorphisms in *PSCA* have also been associated with gastric cancer susceptibility, suggesting that variability in *PSCA* expression may contribute to interindividual variability in cancer risk. Within our data, *Psca* was highly expressed in a *Krt5-/Trp63-/Krt7+* population of predominantly *Cdh1-/-* cells. This pattern differs from prior analysis of cytokeratin molecules in prostate epithelium, where PSCA-positive cells retain KRT5 and KRT14 but lose TP63 (Tran et al. 2002). Although a specific relationship between E-cadherin and gastrointestinal cytokeratin molecules has not been described, our findings suggest that loss of *Cdh1* in gastric lineage progenitors may disrupt normal tissue morphogenesis and organization.
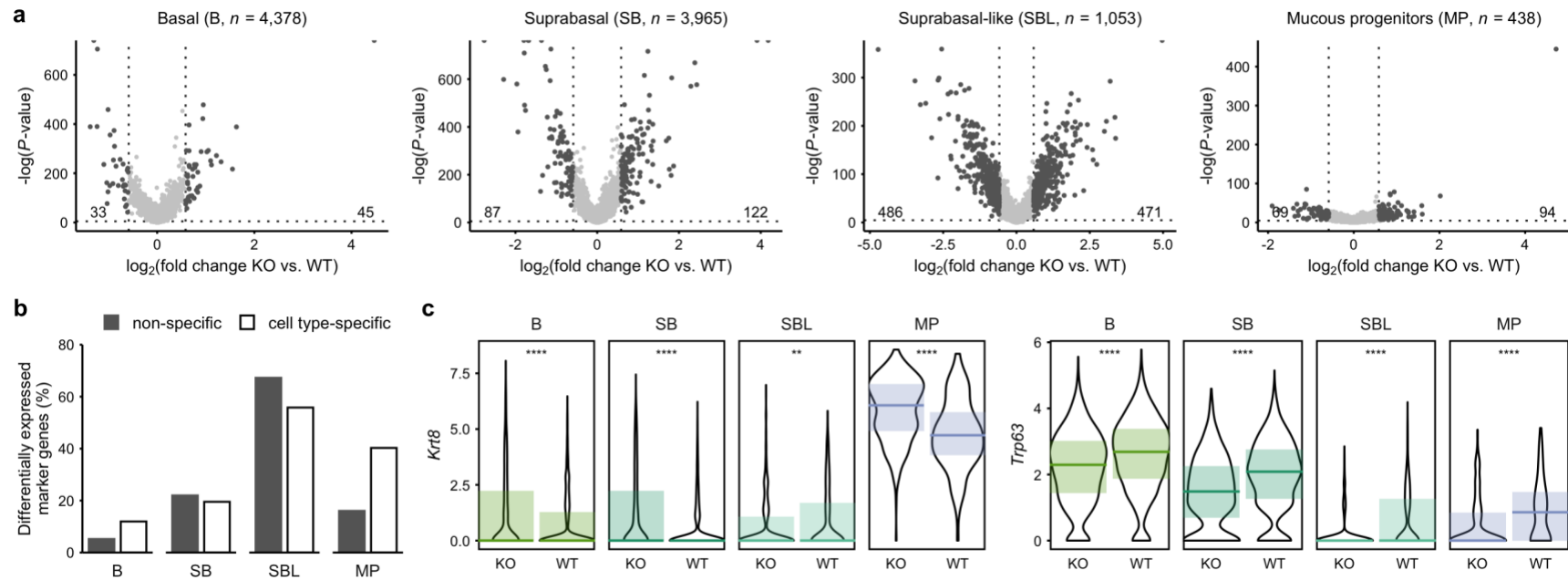
Figure 3.5. *Cdh1* loss promotes the expression of luminal genes in basal and suprabasal cell types. a. Volcano plots showing differential gene expression between *Cdh1* knockout (KO) and wild-type (WT) cells within four epithelial cell types. **b.** Percent of non-specific and cell type-specific marker genes differentially expressed in KO cells. **c.** Violin plots showing the distribution of mucous progenitor cell-specific marker *Krt8* and basal and suprabasal cell marker *Trp63* in KO and WT cells within epithelial cell types. *P*-value, Wilcoxon rank-sum test: **, $P \leq 0.01$; ****, $P \leq 0.0001$.

3.3.4 Prognostic gene signatures differentiate between gastrointestinal epithelial cell types

Specific patterns of gene expression have been shown to stratify molecular and prognostic gastric cancer subtypes (Cristescu et al. 2015). Poor prognosis mesenchymal-like tumours in particular are predominantly of the diffuse histological subtype and are microsatellite stable. To investigate whether *Cdh1* loss in murine gastric organoids was associated with an epithelial-to-mesenchymal transition (EMT)-like transcriptional program, we assessed the presence of an established gene signature comprised of 150 mesenchymal-like genes and 161 epithelial-like genes (Loboda et al. 2011). Overall, the expression of mesenchymal ($n = 143$) and epithelial ($n = 149$) genes differentiated epithelial from non-epithelial cell clusters, among which fibroblasts showed the highest expression of mesenchymal genes (Figure 3.6). However, epithelial subpopulations showed varying contributions of this gene signature, with basal cells showing a more mesenchymal-like pattern of gene expression compared to suprabasal, suprabasal-like or mucous progenitor cells. Within each epithelial cell population, the contributing clusters showed varying contributions of the EMT signature. For example, basal cell cluster 10 showed higher expression of mesenchymal-like genes compared to other basal cell clusters. These results were consistent with findings that genes involved in extracellular matrix reorganization were overexpressed in basal cells compared to other epithelial cell populations.
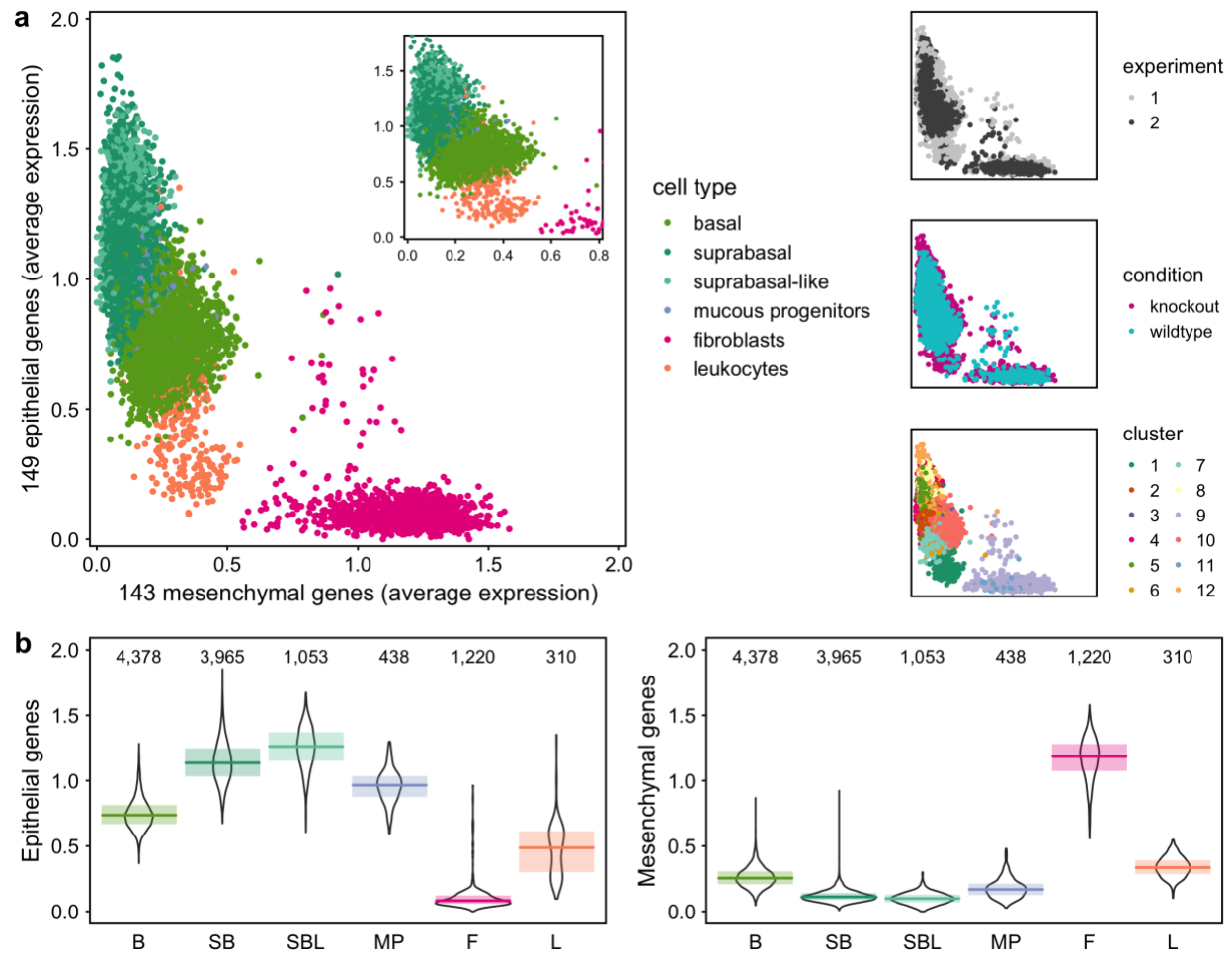
Figure 3.6. Epithelial basal cells show a mesenchymal-like gene signature. a. Average $\log_2$-normalized expression across epithelial- and mesenchymal-like genes identified by Loboda *et al*. (2010) and coloured by cell type, experiment, condition, and cell cluster. **b.** Cell type-specific differences in the expression of EMT gene signatures. The number of cells is noted above for each cell type. B, basal; SB, suprabasal; SBL, suprabasal-like; MP, mucous progenitors; F, fibroblasts; L, leukocytes.

3.4   Discussion

Given the limited understanding of disease pathogenesis and few effective treatment options in many cancer syndromes, identifying molecular markers associated with the early stages of carcinogenesis may identify opportunities for improved clinical management. In this study, we characterized a murine organoid model of HDGC using single-cell transcriptome sequencing. These data established distinct epithelial cell populations that show cell type-specific patterns of gene expression and differential response to genetic loss of *Cdh1*, ultimately allowing the identification of a putative marker of early SRC foci in gastric epithelial tissue from *CDH1* carriers. Overall, our findings demonstrate the potential utility of three-dimensional organoid development and single-cell technologies in studying the molecular events associated with homozygous loss of tumour suppressor genes in non-diseased tissue.

E-cadherin is a member of the cadherin protein family and a critical component of adherens cell junctions. Loss of E-cadherin, and transcriptional upregulation of the mesenchymal neural (N)-cadherin, is a signature of neoplastic cells undergoing EMT. In gastric adenocarcinoma, genetic loss of *CDH1* is associated with genomic stability and is enriched in histologically diffuse-type gastric cancers (Bass et al. 2014). Although these observations support a specific mechanism of disease pathogenesis in *CDH1* carriers, limited understanding of the initiating molecular events underlying diffuse gastric cancer prevent the identification of predictive, prognostic or therapeutic biomarkers associated with cancer onset. Given the crucial role of E-cadherin in maintaining the structural integrity of epithelial tissue, global deregulation of cytokeratins in *Cdh1*-/- cells may reflect early events in cytoskeletal reorganization. Specific upregulation of glandular mucosa markers *Krt8* and *Krt7*, along with downregulation of the basal cell marker *Trp63*, suggests that *Cdh1* loss may mediate epithelial cell differentiation towards a

simple columnar cell fate. However, the mechanisms underlying global and cell type-specific transcriptional regulation require further study.

HDGC is associated with a significant lifetime risk of diffuse gastric cancer and lobular breast cancer and is characterized by early age of onset, advanced stage of disease and poor prognosis. Due to the lack of specific symptoms at the early stages of disease and potential for rapid cancer progression, prophylactic total gastrectomy is recommended for most *CDH1* carriers. Total gastrectomy is associated with significant morbidity, including both immediate and long-term effects such as weight loss, anastomotic leakage, bile reflux and dumping syndrome (Lang et al. 2000; Strong et al. 2017; van der Kaaij et al. 2018). In cases where prophylactic gastrectomy is delayed or for individuals who elect not to undergo preventive surgery, regular endoscopic surveillance may be recommended 5-10 years earlier than the age of the youngest cancer onset in the family (Barber et al. 2008). However, pre-malignant lesions are identified in around 90% of tissue specimens from individuals undergoing prophylactic total gastrectomy, indicating inherent challenges to endoscopic screening for diffuse gastric cancer. Identification of biomarkers for early neoplastic lesions, such as our finding that KRT7 specifically stained SRC foci in a prophylactic gastrectomy specimen from one *CDH1* carrier, may provide a basis for minimally-invasive screening and early cancer detection.

Describing the molecular events associated with carrier status for high-penetrance cancer predisposition syndromes has implications for the identification of carriers in unselected populations, variant interpretation and genetic diagnosis in suspected cancer families, and choice of therapies in affected carriers. Modeling the early events in tumourigenesis may ultimately provide mechanistic insights into the site-specificity of cancer predisposition genes, recapitulating the cellular composition of tumours associated with carrier status for high-

penetrance genes (Rosenbluth et al. 2020). Translation between animal models and human tissues is facilitated through tissue-specific model systems such as organoid cultures that promote use of cell type-specific techniques including scRNA-seq. While *in vitro* models are limited in their ability to replicate *in vivo* tumour development, this approach may allow the dissection of cell-, tissue-, and gene-specific involvement in the molecular pathogenesis of cancer predisposition syndromes.

Chapter 4: Long-read sequencing improves variant interpretation and genetic diagnosis for cancer susceptibility

4.1   Introduction

A significant amount of genetic variation in the human genome is due to SVs, such as deletions, duplications, inversions and translocations (Sudmant et al. 2015; Jain et al. 2018). Genome sequencing allows high-resolution gene-agnostic analysis of variants in known and novel disease genes, and thus genome sequencing may improve rates of molecular diagnosis by overcoming some of the limitations of targeted clinical assays. NGS, or sequencing by synthesis, is the most widely used sequencing technology, and is based on the generation of short (50-300 bp) reads that are aligned to a reference genome or assembled into longer contiguous sequences (contigs) prior to alignment. Accurate alignment and variant calling in NGS is challenging due to regions of low sequence complexity, repetitive elements and strong GC bias in the human genome, reducing the sensitivity and specificity for novel variant discovery. This indicates a need for improved approaches to characterize genetic variation, particularly for large or complex variants.

Moderate- to high-penetrance germline variants in cancer predisposition genes underlie a small proportion of all cancers. However, the prevalence of SVs in clinical and research cancer cohorts is likely underestimated due to the technical and computational limitations of multigene panel, exome and genome sequencing (Cheng et al. 2017). Recently, third-generation or long-read sequencing has been used to characterize complex genetic variation in human genomes and aid in the diagnosis of rare disorders (Merker et al. 2018; Sanchis-Juan et al. 2018). Long-read sequencing may thus improve the molecular diagnosis of suspected hereditary cancer families.

To investigate the contribution of germline SVs to cancer susceptibility, long-read genome sequencing was performed for individuals with known or suspected cancer predisposition syndromes. Our results demonstrate the nanopore long-read sequencing improves the resolution of germline SVs identified by short-read genome sequencing, and complementary sequencing-based approaches may improve the differential molecular diagnosis of individuals who remain genetically undiagnosed following clinical panel-based NGS.

4.2   Materials and Methods

4.2.1   Illumina genome sequencing

Short-read genome sequencing was previously performed on Illumina HiSeq platforms in normal tissue samples for 705 advanced cancer patients enrolled in the POG program. Putative SVs were identified in genomes aligned to the human reference genome version hg19 using multiple copy number and SV calling tools. Illumina genome sequencing reads were aligned to the human reference genome version hg19 using BWA-MEM v0.7.6, and duplicate reads were removed using Picard tools v1.92 (http://broadinstitute.github.io/picard/) (Li and Durbin 2009). To improve the sensitivity of SV detection, two computational pipelines were implemented to identify potential pathogenic and likely pathogenic germline SVs. Large copy number variants were called using the read depth-based tool Control-FREEC, and region-based filtering was used to identify variants overlapping 99 cancer predisposition genes. Known and recurrent technical artifacts were subsequently filtered prior to manual review. SV calling was performed using DELLY v0.7.3 and Manta v1.0.0 and were aggregated with mRNA fusion events detected by Trans-ABySS v1.4.10. Putative variants identified by each tool were compared, merged and

annotated with gene and functional information using MAVIS (Robertson et al. 2010; Rausch et al. 2012; Chen et al. 2016; Reisle et al. 2019).

4.2.2   Germline variant curation

Gene-based filtering and filtering based on predicted impact to protein-coding regions was performed to identify non-synonymous variants in candidate cancer predisposition genes. Manual review of germline and tumour Illumina genome sequencing data was performed using IGV v2.7.0 to flag suspected technical artifacts and prioritize candidate variants for assessment by Oxford Nanopore long-read sequencing. These variants were used to determine the sensitivity of SV calling from short-read genome sequencing and guide manual data curation of novel variants. Fourteen SVs that were predicted to have a deleterious impact on gene expression or function were subsequently identified through manual review in IGV. Variants in five known carriers previously identified by clinical guideline-based testing were used to evaluate the sensitivity of SV calling through Illumina genome sequencing. Filtering, prioritization and review for small variants using short-read GS was performed as described in Chapter 2.

4.2.3   Oxford Nanopore sequencing

Long-read sequencing was performed on 13 of 14 POG cases with candidate germline SVs and for whom archived DNA was available, as well as for three probands from suspected hereditary cancer families. Genome libraries were constructed for high-molecular weight DNA purified from peripheral blood and sequenced on the Oxford Nanopore Technology MinION or PromethION. Base calling and read alignment were performed using Guppy version 3 and Minimap2, respectively, and alignments were visualized in IGV (Li 2018; Wick et al. 2019).

Variant calling was performed for samples sequenced on the PromethION using Sniffles v1.0.11 (Sedlazeck et al. 2018). Paired tumour genome sequencing and RNA-seq were assessed for somatic mutations, LOH, mutational signatures, alternative splicing and fusion transcript expression as previously described and detailed in Chapter 2 (Pleasance et al. 2020).

4.2.4 Breakpoint sequence analysis

Repetitive elements overlapping breakpoints predicted by Illumina or Nanopore genome sequencing were identified using the annotated RepeatMasker dataset obtained from the University of California Santa Cruz Table Browser for the reference genome version hg19 (http://genome.ucsc.edu/) (Smit et al. 1996; Karolchik 2004). Sequence identity within ±150 bp of predicted breakpoints was evaluated through pairwise sequence alignment using EMBOSS Needle (Needleman and Wunsch 1970). Percent identity and gaps in pairwise alignments between each corresponding 5' and 3' breakpoint were noted, and each alignment was manually reviewed for regions of microhomology. Genomic features at breakpoint junctions were similarly evaluated through pairwise sequence alignment and manual review, comparing short-read contig sequences, when available, and expected junctional sequences based on the reference genome.

4.2.5 Sanger sequencing

Primers were designed for PCR and Sanger sequencing across the canonical splice junctions of *MSH2* exons 13-15 using Primer3Plus (http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi) and the UCSC In Silico PCR tool (https://genome.ucsc.edu/cgi-bin/hgPcr). Primer sequences were as followed: *MSH2* exons 13-

14, 5'-CTTGGCCAATCAGATACCAAC-3' (forward, F) and 5'-

CATATCCTTGCGATTCTCCAA-3' (reverse, R); *MSH2* exons 14-15, 5'-

CCCTGGAACTTGAGGAGTTTC-3' (F) and 5'-CAGTAAAGGGCATTTGTTTCAC-3' (R);

and *MSH2* exons 13-15, 5'-CTTGGCCAATCAGATACCAAC-3' (F) and 5'-

CAGTAAAGGGCATTTGTTTCAC-3' (R). Peripheral blood RNA collected in PAXgene Blood

RNA Tubes (PreAnalytix) was extracted according to manufacturer's instructions using the

RNeasy Mini Kit (Qiagen). cDNA conversion and PCR were performed using the SuperScript

IV Reverse Transcriptase (Thermo Fisher) and Platinum PCR SuperMix High Fidelity (Thermo

Fisher), respectively, according to manufacturer's instructions. PCR products were analyzed by

gel electrophoresis and sent to GENEWIZ (South Plainfield, New Jersey, USA) for Sanger

sequencing.

4.2.6   RNA-seq analysis

For advanced cancer cases sequenced as part of the POG program, paired-end tumour

RNA-seq reads were aligned to the hg19 reference genome using Trans-ABySS version 1.4.10,

and duplicate reads were marked with Picard. mRNA read support for aberrant splicing and

fusion transcript expression associated with germline SVs was computed using TAP, a pipeline

for targeted assembly and realignment (Chiu et al. 2018). Briefly, we classified and filtered

RNA-seq reads matching target gene reference sequences and performed *de novo* assembly using

Trans-ABySS. Contigs were aligned to the reference genome and transcriptome using BWA-

MEM to characterize splicing events and fusion transcripts, and read support across known and

novel splice and fusion junctions was calculated from the number of reads mapping to each

contig sequence. Detailed methods for RNA-seq analysis are included in Appendix B.

4.3   Results

4.3.1   Short-read GS identifies putative germline SVs in cancer predisposition genes

Among advanced cancer patients unselected for personal or family cancer history, twelve candidate germline SVs were detected in fourteen individuals by short-read GS (Table 4.1). Five individuals (Cases 8, 10, 11, 13 and 14) were known carriers of high-penetrance germline variants who had received prior clinical variant confirmation through panel-based NGS or multiplex ligation-dependent probe amplification (MLPA). Eight deletions, two inversions and two complex rearrangements were predicted to disrupt the coding sequence of at least one known cancer predisposition gene. Although most variants were detected by multiple short-read SV calling tools and inferred through contig-level read support, three variants were identified by only one tool, including one with prior clinical validation (Table 4.1 and Supplementary Table 4.1). Surprisingly, three unrelated individuals without medical histories suggestive of TSC were found to carry a recurrent and predicted pathogenic event on chromosome 16p13 identified through short-read genome sequencing (Figure 4.1). Long-read sequencing performed in Cases 1-3 revealed that an inverted duplication of an Alu element from *TSC2* intron 16 into *IFT140* intron 30 was miscalled by both DELLY and Manta and could not be resolved through manual review, consistent with ambiguous alignment of short reads at these loci. This finding, in addition to the lack of clinical phenotype consistent with TSC in any of the carriers, led to the classification of this variant as likely benign.

Table 4.1. Variant information and patient characteristics for germline structural variants predicted or known to be deleterious by short-read genome sequencing

Case ID	Resolved variant	SRS evidence	Descriptive utility of LRS	Coding sequence impact	ACMG/AMP classification (criteria)	Indication for clinical genetics assessment[a]
Cases 1-3	NC_000016.9:g.1566535_1566536ins2119755_2119863inv	variant miscalled[b]	variant re-interpretation and confirmation of false-positive finding	none	likely benign (BS2)	no referral
Case 4	NC_000005.9:g.176441544_176441555delins176409841_176603468inv	PR, SR, contig	resolution of variant configuration	NSD1 5'UTR-exon 2 duplication	likely benign (BS2)	no referral
Case 5	NC_000016.9:g.2093921_2214187delins2126780_2212350inv	PR, SR, contig	resolution of variant configuration	TSC2 5'UTR-exon 25 deletion	pathogenic (PVS1, PM2, PP4)	Tuberous sclerosis complex
				NTHL1 5'UTR-exon 3 deletion	pathogenic (PVS1, PM2)	autosomal recessive NTHL1-associated polyposis[d]
Case 6	NM_000051.3(ATM):c.2467-527_8851-2114del	read depth	resolution of breakpoints near flanking repetitive elements	ATM exons 17-61 deletion	pathogenic (PVS1, PM2)	ATM-associated cancer susceptibility
Case 7	NM_058216.2(RAD51C):c.706-1013_837+296delins706-469_837+296inv	SR[c]	resolution of 5' breakpoint and flanking deletion	RAD51C exon 5 deletion	likely pathogenic (PVS1 [strong], PM2)	moderate-penetrance ovarian cancer susceptibility
Case 8	NM_000051.3(ATM):c.1065+647_1236-369del	contig	confirmation	ATM exon 9 deletion	pathogenic (PVS1, PM2)	ATM-associated cancer susceptibility
Case 9	NC_000017.10:g.41217614_41295110del	PR, SR, contig, read depth	confirmation	BRCA1 5' UTR-exon 17 deletion	pathogenic (PVS1, PM2)	HBOC
Case 10	NM_007294.3(BRCA1):c.547+946_4186-1194del	read depth	confirmation	BRCA1 exons 9-12 deletion	pathogenic (PVS1, PM2)	HBOC
Case 11	NC_000002.11:g.47545553_47674137del	PR, SR, contig, read depth	confirmation	EPCAM deletion MSH2 5'UTR-exon 7 deletion	pathogenic (PVS1, PM2)	Lynch syndrome
Case 12	NM_000135.2(FANCA):c.792+452_1826+222del	PR, SR, contig, read depth	confirmation	FANCA exons 9-20 deletion	pathogenic (PVS1, PM2)	autosomal recessive Fanconi anemia[c]

Table 4.1. Variant information and patient characteristics for germline structural variants predicted or known to be deleterious by short-read genome sequencing (continued from previous page)

Case 13	NM_024675.3(PALB2):c.2835-282_3113+1377del	PR[b]	confirmation	*PALB2* exons 9-10 deletion	pathogenic (PVS1, PM2)	moderate-penetrance breast cancer susceptibility
Case 14	NM_000546.5(TP53):c.-28-252_920-15del	PR, SR, contig	NA	*TP53* exons 2-9 deletion	pathogenic (PVS1, PM2, PP4)	Li Fraumeni syndrome

[a]Indication for referral for hereditary cancer risk assessment on the basis of the variant identified. Detailed personal and family cancer history for each case reported here is included in Appendix B.

[b]The predicted variant, NC_000016.9:g.1566535_2119866inv, was miscalled by short-read genome sequencing based on paired reads, split reads and contigs in three unrelated cases. This variant was subsequently found by nanopore sequencing to reflect an inverted duplication of an Alu element from *TSC2* intron 16 into intron 30 of *IFT140*.

[c]Germline variants in cases 7 and 13 were additionally supported by multiple lines of read evidence in matched tumour tissue.

[d]Clinical referral on the basis of carrier status for recessive syndromes should be considered in the context of family structure and medical history that may have differing indications for XY, XX and XO relatives at risk.

FHx, family history; HBOC, hereditary breast and ovarian cancer; LRS, long-read genome sequencing; NA, not applicable; PR, paired reads; SR, split reads; SRS, short-read genome sequencing
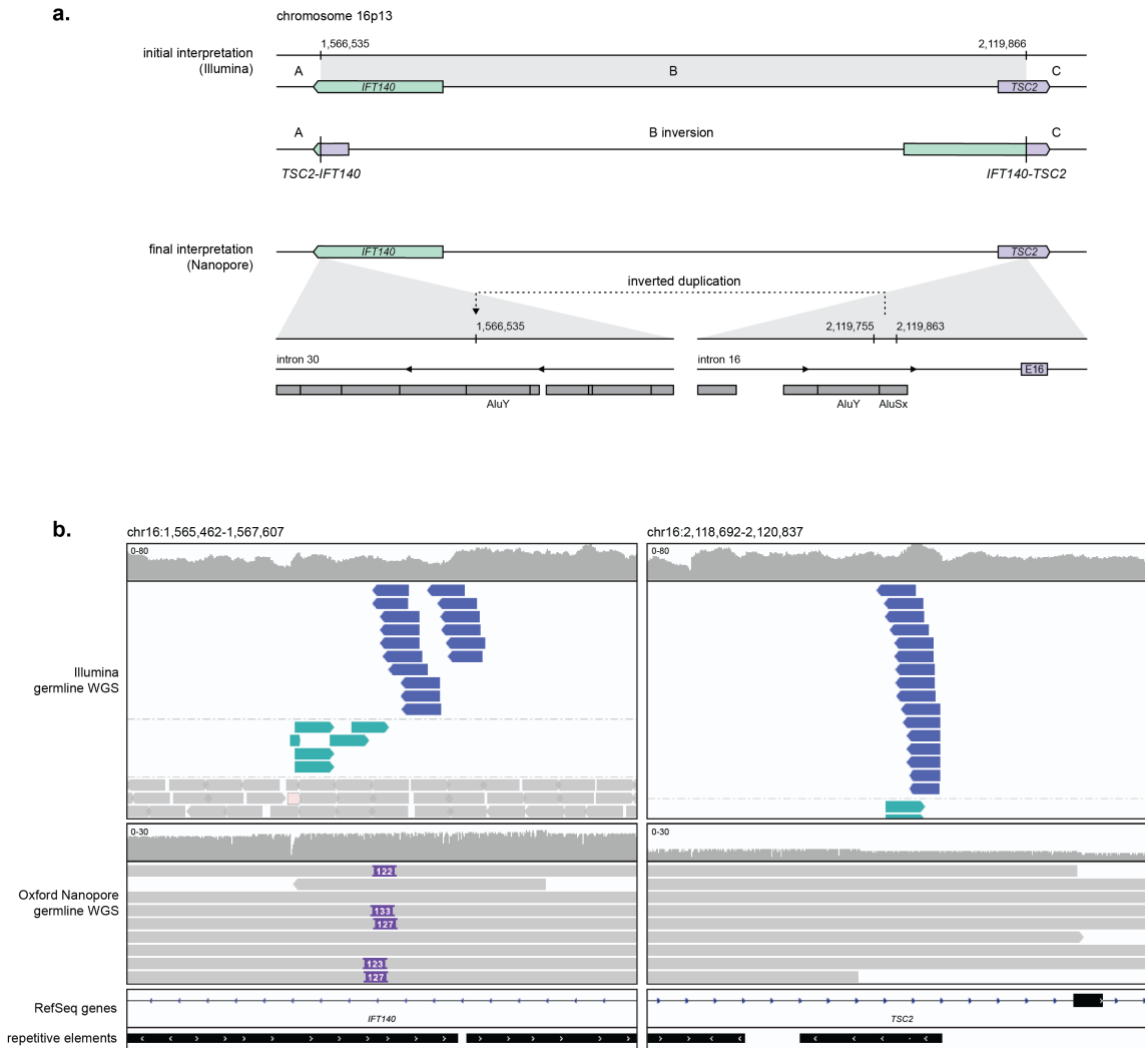
Figure 4.1. A recurrent germline variant resolved using long-read sequencing. a. Schematic of a recurrent event identified in Cases 1, 2 and 3 and predicted to be pathogenic was reinterpreted as a likely benign intronic variant based on Oxford Nanopore sequencing. Illumina short-read genome sequencing data supported a long-range inversion on chromosome 16p13 with breakpoints in *IFT140* and *TSC2* (upper), while Nanopore sequencing data showed an insertion in intron 30 of *IFT140* likely arising from an Alu element in intron 16 of *TSC2* (lower). **b.** Illumina and Oxford Nanopore genome sequencing data for Case 3 visualized using IGV at the loci of *IFT140* and *TSC2*. Paired-end reads mapping to intron 30 of IFT140 and intron 16 of *TSC2* are shown in parallel and coloured by strand. 133 bp and 136 bp insertions were found in two Nanopore reads, with sequences mapping to Alu elements at the locus of the *TSC2* breakpoint predicted by Illumina short-read sequencing.

4.3.2 Complex genetic rearrangements resolved by nanopore sequencing

A novel complex rearrangement was identified on chromosome 5q35 in Case 4, who was shown to carry a 194 kb inverted duplication flanked by a small indel at the breakpoint junction (Figure 4.2). Two fusion transcripts supporting the breakpoints, *NSD1-UIMC1* and *UIMC1-ZNF346*, were identified by RNA-seq. However, the configuration of the variant determined from LRS indicated that, in addition to their partial duplication, undisrupted copies of both *NSD1* and *UIMC1* were maintained on the variant allele. Given the individual's unremarkable medical history, with no known clinical features consistent with Sotos syndrome, this variant was classified as likely benign. In contrast, nanopore sequencing in Case 5 indicated that a complex variant identified on chromosome 16p13.3 involved an 85 kb inversion with breakpoints in *TSC2* and *TRAF7* flanked by two deletions, resulting in partial loss of *NTHL1* and *TSC2* (Figure 4.3). Furthermore, LOH at the locus in the individual's tumour indicated that the complex germline rearrangement involved only one allele. This case had a prior history of TSC and has been previously described (Wong et al. 2018). The complex SV in this case was thus associated with TSC and carrier status for *NTHL1*-associated polyposis, caused by partial heterozygous germline loss of *TSC2* and *NTHL1*, respectively.
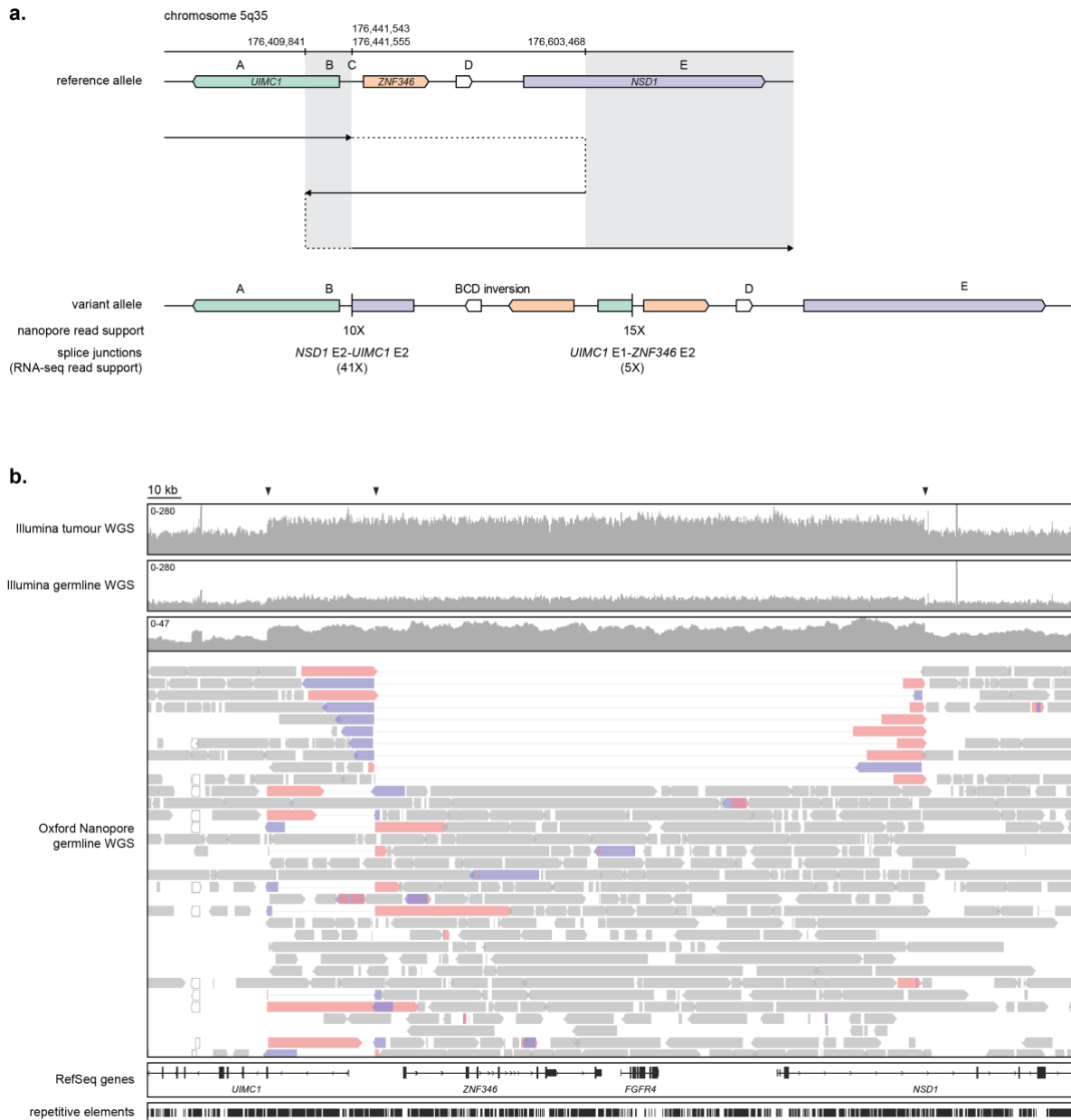
Figure 4.2. Long-read sequencing resolves variant configuration and interpretation in Case 4. **a.** Schematic of a likely benign complex germline SV identified in Case 4. **b.** Illumina and Oxford Nanopore genome sequencing data for Case 4 visualized using IGV at the locus of *UIMC1* and *NSD1*. Split Nanopore reads spanning the breakpoint junctions are shown mapping to flanking regions of the predicted breakpoints, denoted by black arrows, and connected by a thin gray line. Read segments coloured red and blue denote split reads mapping to both plus and minus strands, indicating a probable inversion event.
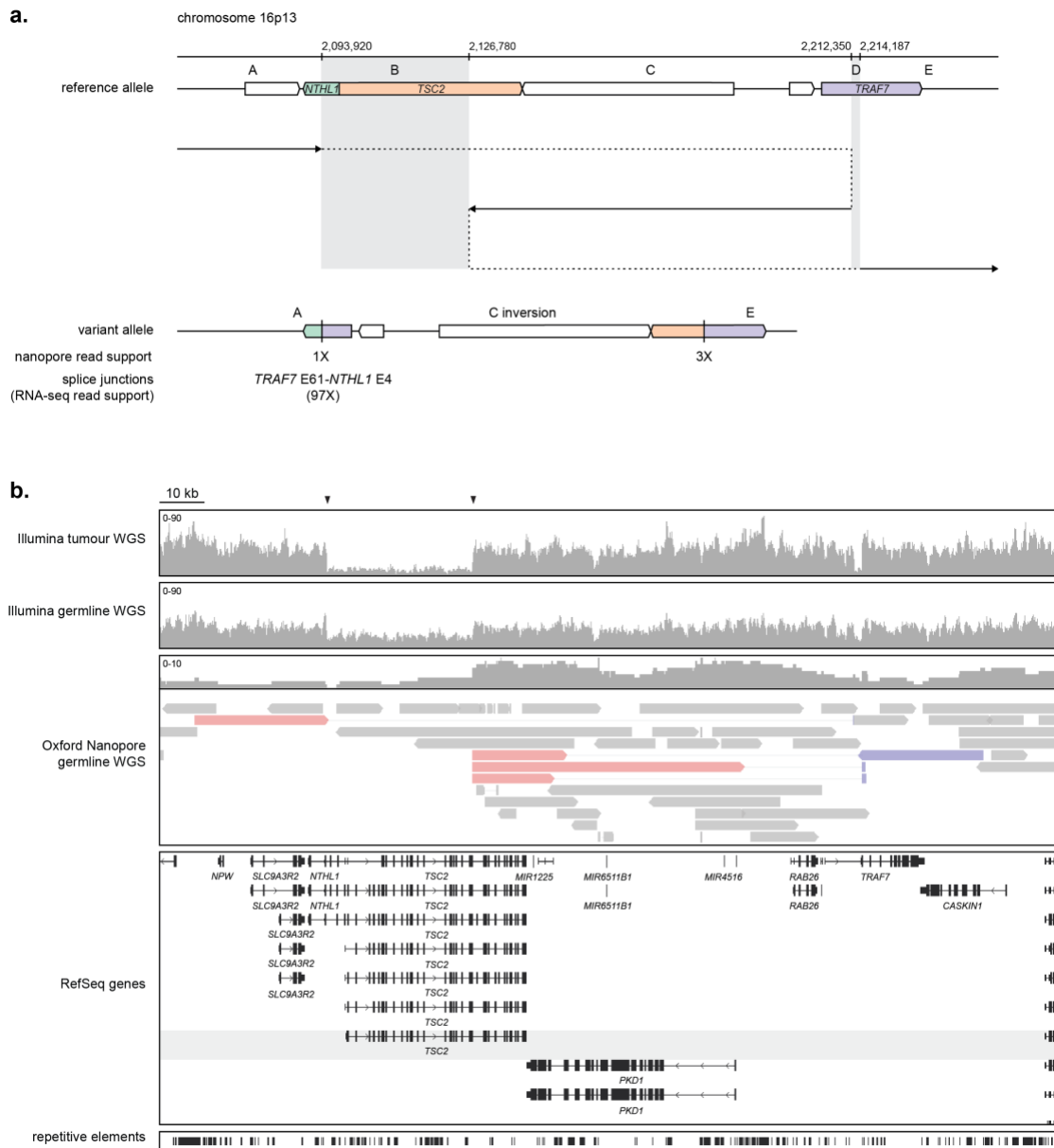
Figure 4.3. Long-read sequencing resolves configuration of a complex SV in Case 5. **a.** Schematic of a pathogenic germline SV identified in Case 5. **b.** Illumina and Oxford Nanopore genome sequencing data for Case 5 visualized using IGV at the locus of *TSC2* and *NTHL1*. Split Nanopore reads spanning the breakpoint junctions are shown mapping to flanking regions of the predicted breakpoints (black arrows) connected by a thin gray line. Read segments coloured red and blue denote split reads mapping to both plus and minus strands, indicating a probable inversion event.

Nanopore sequencing further informed SV breakpoints in two cases and confirmed simple deletions in six additional cases. Sequence analysis at the breakpoint junctions found that repetitive elements were present at most breakpoints, suggesting that they contributed to both the formation of large SVs and miscalling of a recurrent variant (Table 4.2). Long tracks of homology in two cases indicated that variant formation may have been a consequence of break-induced replication. Notably, the breakpoints of a partial *ATM* deletion in Case 6 were predicted to occur near two long interspersed nuclear elements (LINEs), of which a single copy could be mapped to two PromethION reads (Supplementary Figure 4.1). Many SV breakpoints had simple blunt ends or small indels in the absence of microhomology, short regions of shared nucleotide identity, characteristic of products of non-homologous end joining (Carvalho and Lupski 2016). Microhomology near the breakpoints in Cases 4, 7 and 11 suggested that these events may have arisen through microhomology-mediated end joining or microhomology-mediated break-induced replication. Perhaps as the results of sequence homology, a 544 bp deletion at the 5' breakpoint of a *RAD51C* exon 5 inversion in Case 7 was not confidently captured by short-read sequencing (Supplementary Figure 4.2).

Table 4.2. Repetitive elements and sequence similarity at breakpoint junctions for germline structural variants detected through short-read genome sequencing listed in Table 4.1

		5' breakpoint			3' breakpoint		Breakpoint sequence analysis (±150 bp)			
Case ID	Position	Repeat name (class)	Length (strand)	Position	Repeat name (class)	Length (strand)	Identity	Gaps	MH	Junction features
Cases 1, 2 and 3	16:1,566,535	AluY (SINE)	303 bp (+)	16: 2,119,755	AluY (SINE)	295 bp (-)	65.8%	15.4%		unknown
				16: 2,119,836	AluSx (SINE)	133 bp (-)	61.3%	26.6%		unknown
Case 4	5:176,441,543	NA	NA	5:176,603,468	AluJo (SINE)	167 bp (-)	39.9%	31.5%	yes	indel
	5:176,409,841	AluSx (SINE)	286 bp (-)	5:176,441,555	NA	NA	50.7%	22.0%	yes	indel
Case 5	16:2,126,780	NA	NA	16:2,214,187	(CGTG)n (Simple repeat)	55 bp (+)	44.6%	28.6%		indel
	16:2,093,920	NA	NA	16:2,212,350	NA	NA	48.5%	23.5%		blunt ends
Case 6	11:108,137,370	L1PA2 (LINE)	6,017 bp (+)	11:108,233,694	L1PA2 (LINE)	6,036 bp (+)	41.3%	31.5%		unknown
Case 7	17:56,786,207	AluSx3	112 bp(-)	17:56,786,751	NA	NA	55.1%	38.2%	yes	unknown
	17:56,786,751	NA	NA	17:56,787,647	AluSg (SINE)	316 bp (+)	40.9%	50.4%	yes	unknown
Case 8	11:108,118,496	AluSg (SINE)	306 bp (-)	11:108,121,054	AluSg (SINE)	257 bp (-)	74.4%	8.3%		blunt ends
Case 9	17:41,217,614	AluSp (SINE)	308 bp (+)	17:41,295,110	(TTTA)n (Simple repeat)	23 bp (+)	32.0%	41.3%		blunt ends
Case 10	17:41,235,786	NA	NA	17:41,250,846	AluSp (SINE)	302 bp (-)	39.9%	31.5%		unknown
Case 11	2:47,545,553	AluSp (SINE)	284 bp (-)	2:47,674,137	AluSq2 (SINE)	296 bp (-)	56.3%	24.0%	yes	blunt ends
Case 12	16:89,844,986	AluSg (SINE)	164 bp (+)	16:89,869,214	L1MA5 (LINE)	474 bp (+)	32.6%	60.1%		blunt ends

Table 4.2. Repetitive elements and sequence similarity at breakpoint junctions for germline structural variants detected through short-read genome sequencing listed in Table 4.1 (continued from previous page)

Case 13	16:23,631,306	AluSz6 (SINE)	292 bp (-)	16:23,634,733	AluSx3 (SINE)	301 bp (-)	77.6%	5.2%	indel
Case 14	17:7,576,941	NA	NA	17:7,580,192	L2 (LINE)	179 bp (+)	43.2%	29.5%	blunt ends

4.3.3   Mechanisms of variant formation and implications for tumourigenesis

Among the ten cases with pathogenic or likely pathogenic SVs identified in this cohort, seven were associated with LOH and four tumours showed significant contributions from somatic SNV signatures with characterized genetic aetiologies: signature 30 was associated with homozygous loss of *NTHL1* in Case 5, signature 3 suggested homologous recombination deficiency caused by loss of *BRCA1* or *PALB2* in Cases 9 and 13, respectively, and signature 6 supported mismatch repair deficiency in Case 11 (Alexandrov et al. 2013). Tumour RNA-seq demonstrated aberrant splicing in several cases with intragenic SVs and sufficient read coverage at the splice junction, thus providing additional support for variant pathogenicity in these cases.

4.3.4   Undetected germline SVs in suspected Lynch syndrome families

Given the opportunities afforded by long-read sequencing to resolve SVs identified using short-read or long-read GS may also elucidate pathogenic SVs underlying cancer predisposition syndromes that are not found by short-read sequencing technologies. To investigate the utility of long-read sequencing in the genetic diagnosis of individuals with strong phenotypic indications of high-penetrance cancer predisposition syndromes, nanopore sequencing was performed for two index cases ascertained based on a personal history of Lynch syndrome-related cancer with MMR deficiency assessed through IHC and suspicious family cancer history. Targeted panel-based sequencing performed in peripheral blood and paired tumour biopsy tissues did not identify pathogenic or likely pathogenic germline variants in the coding or splice site regions of *MLH1*, *MSH2*, *MSH6*, or *PMS2* in either case, nor were somatic mutations, somatic copy number alterations, LOH or *MLH1* hypermethylation found in the tumours that could have explained their MSI phenotype (Chapter 2).

Case 15 had a strong personal and family history of multiple Lynch syndrome-related cancers, meeting phenotype-based Amsterdam I and II criteria (Figure 4.4a). Loss of MSH2 protein by IHC was observed in colorectal tumours from the index case and her affected son, indicating that an undetected germline variant in *MSH2* may segregate with disease in this family. Nanopore sequencing in this case allowed genome-wide variant calling and manual variant curation to identify potential causal variants. A heterozygous 287 bp insertion was detected in *MSH2* at a position 18 bp upstream of the canonical splice acceptor of exon 15 (Figure 4.4b). Analysis of the inserted sequence and flanking breakpoint sequences revealed that an 11 bp region of homology at the breakpoint likely mediated the insertion of a highly conserved Alu element at this locus. Although this variant was not predicted to impact the coding sequence of *MSH2*, *in silico* splice site analysis suggested that this variant may result in the activation of a novel branchpoint or acceptor site. PCR across the splice junctions between exons 14 and 15 and exons 15 and 16 in cDNA derived from peripheral blood did not reveal differences in product size compared to a control. Although skipping of exon 15 is predicted to produce a premature nonsense variant in exon 16, this variant would not be expected to be targeted by NMD given its position in the transcript's last exon. Accordingly, exon 15 appeared to be retained in peripheral blood, indicating that the variant did not disrupt the canonical splice donor site. Sanger sequencing across the native splice junction between exons 14 and 15 did not identify an insertion between annotated canonical splice sites, suggesting that the Alu insertion did not promote retention of an intronic sequence in the variant transcript through the creation of a novel branchpoint or splice acceptor site (Figure 4.4c).
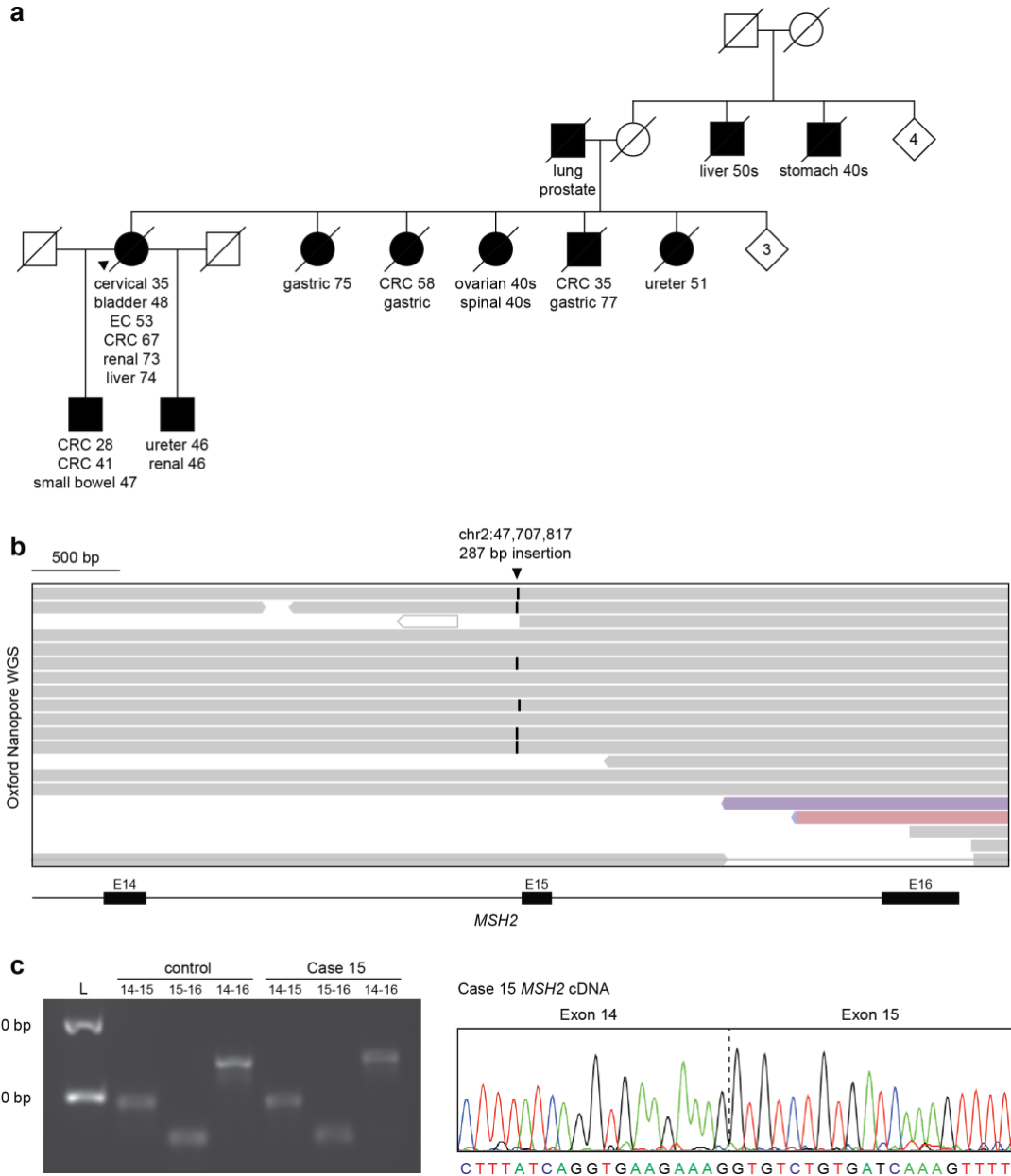
Figure 4.4. Candidate germline SV in an index case from a suspected Lynch syndrome family. a. Pedigree for Case 15. **b.** Oxford Nanopore GS performed in the index case identified a 287 bp insertion in intron 14. The predicted position of insertion along reads containing the variant is shown by black vertical lines. **c.** Gel electrophoresis and Sanger sequencing across the canonical splice junctions of exons 14, 15 and 16. CRC, colorectal cancer; EC, endometrial cancer.

Case 16 had a personal history of cervical cancer and CRC at 29 and 34 years of age, respectively (Figure 4.5). Stomach and endometrial cancers were additionally reported in the patient's extended family. The index case thus met clinical testing criteria on the basis of early onset cancer and family history of multiple Lynch syndrome-related tumours. Combined MLH1/PMS2 deficiency observed by IHC and absence of somatic mutations and hypermethylation of the *MLH1* promoter in the colorectal tumour of the index case suggested that this individual may harbour a potential germline variant in *MLH1*. Unfortunately, no candidate variants were identified through variant calling or by manual review of Oxford Nanopore long reads at the *MLH1* locus. These findings suggested that other non-coding germline variants, constitutional epigenetic changes or undetected somatic alterations, indicating a likely sporadic cancer occurrence, may underlie the early onset CRC in this case. Given the current limitations of long-read sequencing in the identification of small variants, these could not be excluded.
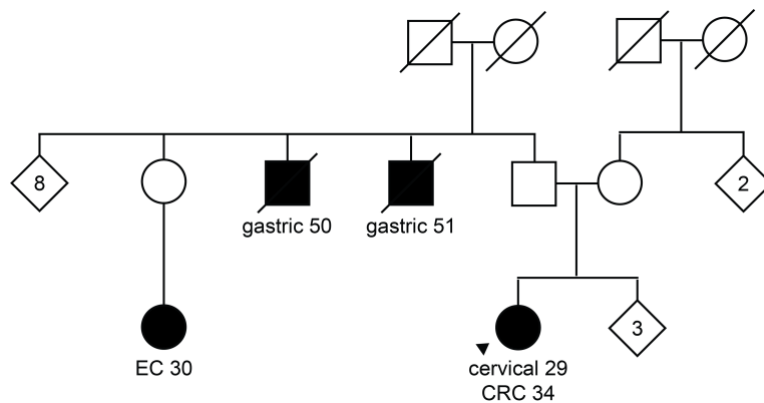
Figure 4.5. Pedigree of molecularly undiagnosed Lynch syndrome in Case 16.

4.3.5   Long-read sequencing to assess causal germline variants in familial pancreatic cancer

While Lynch syndrome offers a unique opportunity for the application of long-read sequencing given the relatively high rate of germline genetic diagnosis among MMR-deficient tumours, other cancer syndromes show more complex or heterogeneous clinical phenotypes that may be challenging to diagnose. For example, an increased risk for pancreatic cancer is associated with several high-penetrance cancer predisposition syndromes, including Lynch syndrome, Peutz-Jeghers syndrome, and familial atypical multiple mole melanoma (FAMMM). However, many families with multiple first-degree relatives affected by pancreatic cancer who do not meet criteria for other inherited cancer syndromes, termed familial pancreatic cancer (FPC), remain molecularly undiagnosed. Germline variants in *BRCA2*, *PALB2* and *ATM* are identified in 15-20% of FPC, and variants in *CDKN2A* underlie pancreatic cancer susceptibility in some FAMMM kindreds without a known history of melanoma (FAMMM-PC) (Bartsch et al. 2012). However, the causal genetic variants in most FPC kindreds remain unknown.

To investigate the molecular basis of cancer susceptibility in FPC, tumour GS and RNA-seq of fresh-frozen pancreatic tumour biopsies were performed for two affected siblings (III-9 and III-12) from an FPC kindred (Figure 4.6a). The index case (III-12) met clinical testing criteria for *CDKN2A*-associated cancer predisposition based on a history of pancreatic cancer reported in eight individuals across three generations and absence of other cancer types. Neither clinical panel sequencing nor germline short-read GS identified a causal variant in coding or splice regions of known cancer predisposition genes. Tumour GS and RNA-seq revealed that the FPC tumours were characteristic of the stable genomic subtype and classical mRNA subtype. These tumours shared common driver mutations implicating known pancreatic cancer pathways

in cancer progression: shared *KRAS* G12R mutations suggested aberrant RAS signaling, focal copy number alterations in *CDK4* and *CDKN2A* indicated deregulation of the p16-mediated cell cycle pathway, altered TGF-β signaling was supported by copy number alterations in *TGFBR2* and *ACVR1B*, and homozygous deletions in *ARID1B* implicated the SWI/SNF-mediated chromatin remodeling complex in tumourigenesis (Figure 4.6b,c).
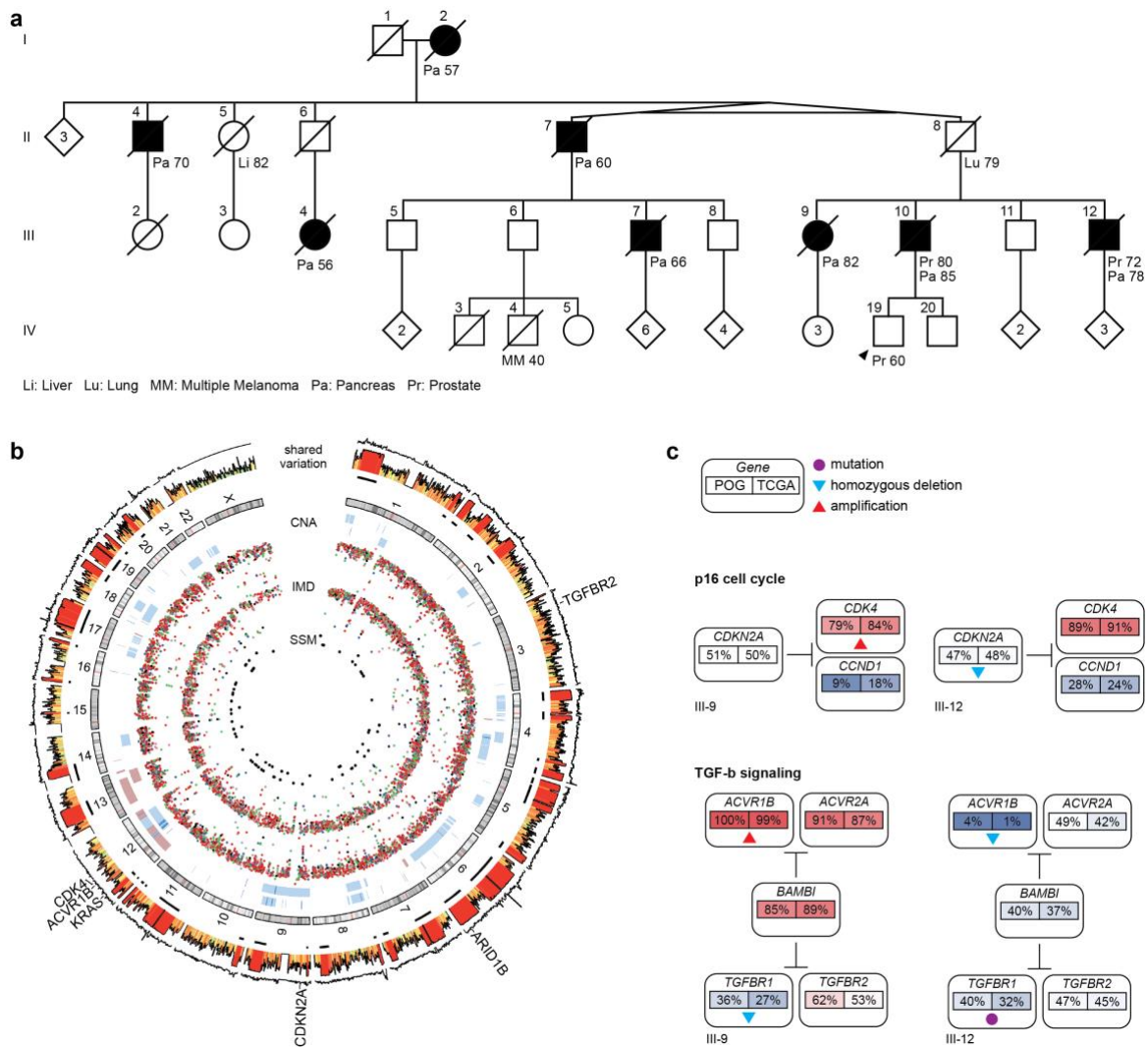
Figure 4.6. Tumour genome and transcriptome landscape in familial pancreatic cancer. a. Pedigree of a genetically undiagnosed kindred with familial pancreatic cancer. Monozygotic twins (II-7 and II-8) are shown by diagonal lines originating from the same point, linked by a horizontal line. **b.** Circos plot comparing germline variation and somatic genomic alterations in tumours from individuals III-9 and III-12. From outer to inner ring: number of shared germline SNVs per Mb, proportion of shared germline SNVs per Mb, segments of shared germline variation, chromosome ideogram, somatic copy number alterations (CNA) for III-9, somatic CNA for III-12, waterfall plot and intermutation distances (IMD) for III-9, waterfall plot and IMD for III-12, coding simple somatic mutations (SSM) in III-9, and coding SSM in III-12. **c.** Schematic summary of somatic alterations and expression percentiles in genes involved in the

p16 cell cycle and TGF-β signaling pathways. Percentiles were calculated for each familial tumour compared to POG PDAC ($n = 44$) and TCGA PAAD ($n = 150$) tumours.

Together with lack of global LOH, telomeric allelic imbalance and large-scale state transitions, the absence of strong contributions from COSMIC signature 3 in the familial tumours suggested that undetected or uncharacterized germline variants in *BRCA2* or *PALB2* were unlikely to underlie pancreatic cancer in this family. Similarly, absence of genome-wide MSI and COSMIC signature 6 similarly did not indicate causal germline variants in *MLH1*, *MSH2*, *MSH6* and *PMS2*. Germline long-read sequencing subsequently performed in III-12 did not identify possible causal SVs in known cancer predisposition genes, including putative variants at the locus of *CDKN2A*.

4.3.6   Exploring novel disease genes in the molecular pathogenesis of FPC

In complex clinical cases without pathogenic small variants or SVs in known cancer predisposition genes, functional evidence supporting a role for variants in novel genes may inform molecular diagnosis. Using tumour RNA-seq, we evaluated possible pathway deregulation by identifying gene expression outliers in resected FPC compared to unrelated pancreatic ductal adenocarcinomas (PDAC or PAAD) from the POG PDAC and TCGA PAAD cohorts (Figure 4.7a). This approach identified 145 genes expressed above the 95_{th} percentile in both familial pancreatic tumours compared to POG and TCGA. Several of these genes were also found to be statistically overexpressed in the familial tumours compared to identically sequenced POG PDAC tumours (fold-change ≥ 1.5, $P \leq 0.01$). Gene ontology analysis indicated an enrichment for genes involved in the regulation of insulin secretion, including insulin expression, processing and storage, in the FPC tumours (Figure 4.7b). Tumours from III-9 and III-12 showed higher expression of several genes encoding proteins involved in glucose-stimulated insulin

secretion, including the pancreatic β-cell transcription factor MAFA (*MAFA*), secretory granule-associated molecules IA-2 (*PTPRN*), PACAP (*ADCYAP1*) and ZNT-8 (*SLC30A8*), and glucose-6-phosphatase catalytic subunit G6PC2 (*G6PC2*) (Saeki et al. 2002; Portela-Gomes et al. 2003; Chimienti et al. 2004; Zhang et al. 2005; Hutton and O'Brien 2009). Consistent with aberrant regulation of glucose metabolism, expression of both insulin (*INS*) and glucagon (*GCG*) were higher in tumours from the FPC kindred.

Overexpression of *INS* and *GCG*, secreted respectively by β-cells and α-cells of pancreatic islets, may indicate a relative enrichment of endocrine cells in the tumour biopsy samples rather than biological differences in gene expression. Therefore, we estimated the proportion of six major pancreatic cell types, namely exocrine ductal and acinar cells and endocrine α-, β-, γ- and δ-cells, in tumour RNA-seq using single-cell transcriptome data from healthy human pancreatic tissue (Segerstolpe et al. 2016; Wang et al. 2019). Despite showing high expression of genes encoding the hormones GCG, INS, pancreatic polypeptide (PPY, γ-cells) and somatostatin (SST, δ-cells), the familial tumours did not show elevated proportions of endocrine cells compared to other PDACs (Figure 4.7c-e). In contrast, transcriptome-based cell type decomposition in pancreatic neuroendocrine tumours (PNETs) revealed major contributions of these endocrine cell types. Furthermore, expression of the ductal cell marker *KRT19* and acinar cell marker *PRSS1* did not differ between familial and unrelated PDACs, showing low relative expression in PNETs (Figure4.7e) (Muraro et al. 2016). Given clinical diagnoses of hyperlipidemia in III-9 and type II diabetes in III-12, these findings ultimately suggested an association between hyperinsulinemia, insulin resistance and pancreatic cancer in this family (Shanik et al. 2008).
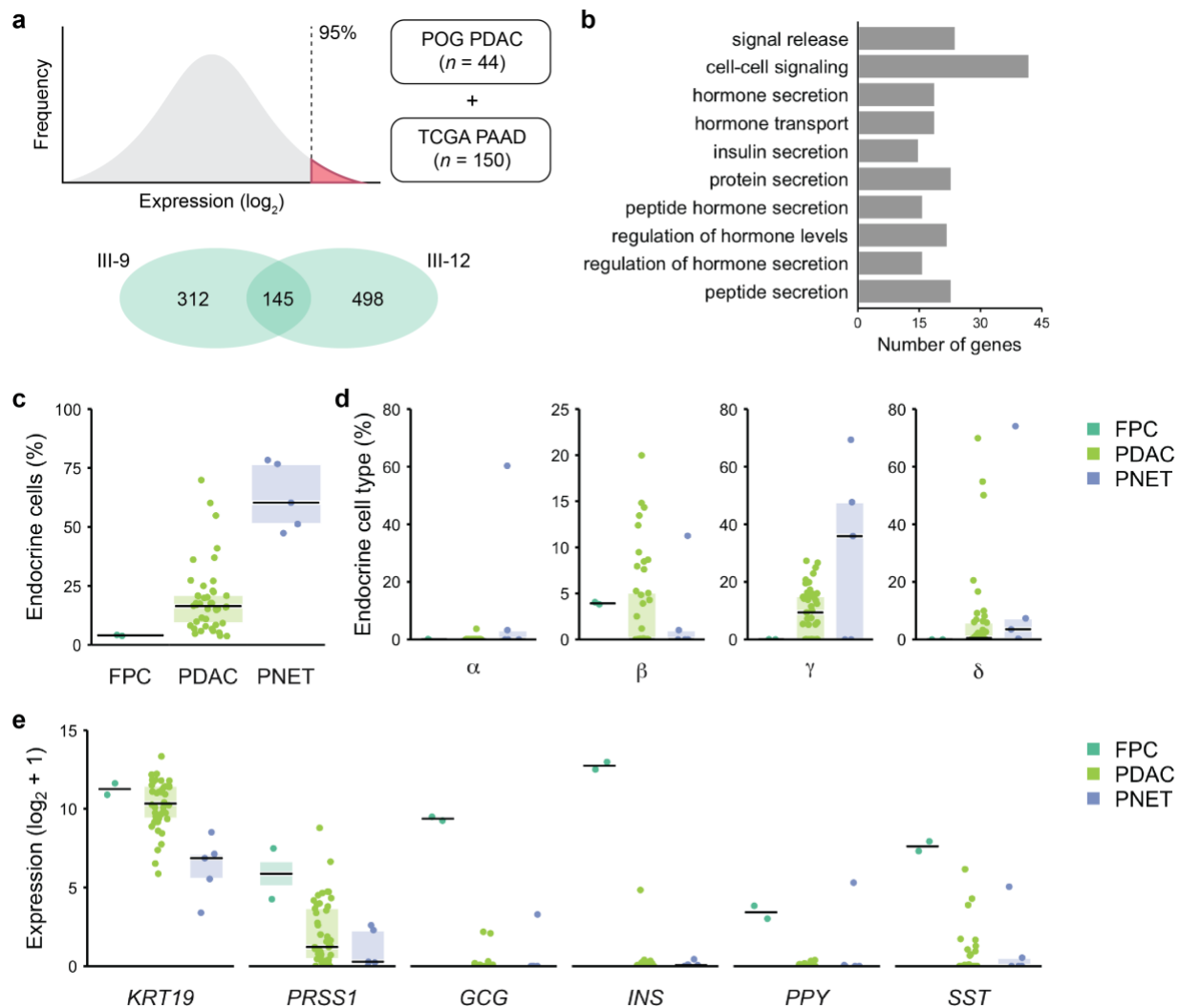
Figure 4.7. Tumour transcriptome sequencing indicates aberrant glucose metabolism in FPC pathogenesis. a. Schematic of outlier expression analysis between familial tumours ($n = 2$) and unrelated pancreatic ductal adenocarcinomas from POG (POG PDAC) and TCGA (TCGA PAAD). **b.** Top 10 most significant pathways enriched among 145 genes overexpressed in FPC tumours. **c,d.** Decomposition of endocrine cell types from bulk RNA-seq across FPC, POG PDAC ($n = 44$) and POG pancreatic neuroendocrine tumours (PNET) ($n = 5$). The percent composition of six major pancreatic cell types was estimated from scRNA-seq data using MuSiC, including four major endocrine cell types (**c**): α-, β-, γ- and δ-cells (**d**). **e.** Expression of marker genes associated with one of six major pancreatic cell types: ductal cells (*KRT19*), acinar cells (*PRSS1*), α-cells (*GCG*), β-cells (*INS*), γ-cells (*PPY*) and δ-cells (*SST*). Values are shown as log2-transformed transcripts per million.

Based on observations from previous studies of non-syndromic FPC kindreds, private genetic variants with unknown functional or clinical significance may underlie cancer susceptibility in molecularly undiagnosed families (Roberts et al. 2016). In the absence of pathogenic germline variants in known moderate- to high-penetrance pancreatic cancer predisposition genes, we investigated the potential contribution of novel disease genes in FPC by reviewing rare heterozygous variants in protein-coding and splice regions. Due to the presence of other pancreatic cancer risk factors in this family, including history of acute pancreatitis, smoking and alcohol consumption, we prioritized variants identified in at least two siblings to account for a possible phenocopy. Among 12 candidate germline variants with predicted loss-of-function, a rare splice site variant was identified in a member of the class II phosphoinositide 3 kinase (PI3K), *PIK3C2G* c.1429+1G>C (Table 4.3). Class II PI3Ks have previously been implicated in vesicle trafficking but their biological functions are not fully understood (Martini et al. 2014). However, recent *in vivo* models suggest that PIK3C2G plays a specific role in hepatic glycogen accumulation and regulation of glucose metabolism (Braccini et al. 2015).

Table 4.3. Candidate causal germline variants in an undiagnosed familial pancreatic cancer kindred

| Carriers | Gene | | Variant | | |
	Symbol	Description	cDNA change	Impact	gnomAD AF
III-9, -10 & -12	*ZP4*	zona pellucida protein	c.876_877delAG	frameshift	0
III-9 & -10	*NAPRT*	NAD biosynthesis enzyme	c.1213C>T	nonsense	9.71E-05
	XIRP2	actin-binding protein	c.8416G>T	nonsense	3.24E-05
III-9 & -12	*CHIT1*	chitin	c.963_991del	frameshift	0
	OR5J2	olfactory receptor	c.427delG	frameshift	0
	PIK3C2G	lipid kinase	c.1429+1G>C	splicing	0.0003
	SKOR1	transcriptional corepressor	c.42+2T>A	splicing	0.0009
	SLC47A2	solute transporter	c.341delG	frameshift	0
III-10 & -12	*MUC19*	mucin	c.21325-2A>C	splicing	0.0036
	PAPLN	ECM glycoprotein	c.54+1G>A	splicing	0.0019
	SH2D4A	intracellular signaling molecular	c.706+1G>A	splicing	0.0012

AF, allele frequency; cDNA, complementary DNA; gnomAD, genome aggregation database

PIK3C2G c.1429+1G>C occurs at the canonical splice donor site of exon 10 and is predicted to result in a shift in reading frame and expression of a premature termination codon in exon 11. mRNA transcripts harbouring premature truncating variants, including frameshift, nonsense and splice site variants, are classically targeted by nonsense-mediated mRNA decay (NMD), a post-transcriptional modulator of normal gene expression that plays important roles in embryonic development, cell differentiation and in response to cellular stress (Lykke-Andersen and Heick Jensen 2015). Tumour RNA-seq demonstrated exon 10 skipping and allelic imbalance at a shared heterozygous SNP (rs12312266) in both carriers, supporting aberrant splicing and NMD-mediated degradation of the variant transcript. Neither LOH nor secondary somatic mutations were identified in *PIK3C2G* to implicate this gene as a classical tumour suppressor;

however, aberrant hormone expression, altered insulin signaling, and co-segregation of hyperlipidemia and type II diabetes mellitus suggested that a potential genetic susceptibility to insulin resistance may mediate pancreatic cancer susceptibility in some FPC kindreds.

4.4   Discussion

The average human genome contains approximately 5-28 thousand SVs, including balanced rearrangements such as inversions and translocations, and unbalanced rearrangements such as large deletions, duplications and insertions (Sudmant et al. 2015; Chaisson et al. 2019). SVs larger than ~3 Mb are found at a high frequency in certain disorders and have historically been assessed using karyotyping or microarrays. However, submicroscopic SVs require molecular approaches with a higher resolution in order to determine variant configuration and to allow for accurate clinical interpretation. Our findings suggest that germline SVs are a rare cause of cancer susceptibility, underlying 1.4% of all cases in an advanced adult cancer cohort ($n = 705$) and 10% of cases associated with moderate- to high-penetrance germline variants in known cancer predisposition genes ($n = 97$) (Chapter 2). Short-read GS detected known variants in five carriers with prior clinical genetic testing, and identified pathogenic and likely pathogenic variants in five additional cases without prior genetic diagnoses. However, short-read sequencing was insufficient to accurately and fully resolve the configuration of three SVs, including two variants that were ultimately classified as likely benign.

Insertions, balanced SVs and complex rearrangements that consist of three or more breakpoints are particularly difficult to characterize using NGS given the inferential nature of SV detection through contig-, split read-, flanking read- or depth of coverage-based approaches. Recently, long-read sequencing has allowed the molecular diagnosis of SVs causing Mendelian

disease in cases where clinical assays or short-read GS have been unsuccessful (Merker et al. 2018; Sanchis-Juan et al. 2018). Here, long-read sequencing confirmed three simple variants and resolved a complex rearrangement in a genetically undiagnosed individual with TSC and carrier status for *NTHL1*. Notably, one individual was a carrier of a *RAD51C* inversion with breakpoints within introns 4 and 6 that would have been missed through targeted NGS. As demonstrated by Rhees *et al.* (2014), the precise characterization of SV breakpoints is critical in order to guide the development of targeted clinical assays for familial, recurrent or founder variants that may be undetectable through standard clinical assays in known or suspected hereditary cancer families (Rhees et al. 2014). This was illustrated by the detection of a 301 bp intronic insertion in *MSH2*, occurring 18 bp upstream of the canonical splice acceptor site of exon 15, in an index case with a personal history of MSH2- and MMR-deficient CRC and family cancer history suggestive of Lynch syndrome.

Although many carriers in our unselected patient cohort had a personal and/or family history suggestive of moderate- to high-penetrance inherited cancer susceptibility, four carriers (40%) did not have a personal or family history that would have indicated prior referral for genetic counseling and testing. This finding is consistent with previous reports suggesting that less than half of carriers identified through population genetic testing meet current clinical testing criteria (Metcalfe et al. 2010). The significance of accurate variant interpretation, particularly in individuals who do not meet phenotype-based testing criteria, was highlighted by Case 3, who was referred for clinical testing on the basis of the miscalled inversion in *TSC2* and LOH in their tumour. At the time of referral, PCR-based validations of the predicted breakpoint junctions were unsuccessful; however, nanopore sequencing later characterized the true variant as a small inverted duplication in a deep intronic region of *IFT140*. On the basis of accurate

variant resolution, classifications for this variant and a complex rearrangement at the locus of *NSD1* were downgraded to likely benign. This ultimately prevented clinical referral for cases without suspicious personal or family medical history.

Pathogenic and likely pathogenic germline variants are identified in only 27-33% of index cases referred for clinical hereditary cancer testing (LaDuca et al. 2014). This indicates a need for complementary testing strategies in families with phenotypic indications of high-penetrance cancer predisposition syndromes to characterize possible causal non-coding variants and SVs. Long-read sequencing in particular may improve genetic diagnosis in cases with clinical and/or molecular evidence supporting specific candidate genes in tumourigenesis. This was explored in two index cases each with personal history of MMR-deficient CRC and family history suggestive of Lynch syndrome. A candidate SV near the canonical splice acceptor site of exon 15 in *MSH2* that was characterized by nanopore sequencing in Case 15 allowed for the investigation of possible aberrant splicing of *MSH2* as a potential cause of Lynch syndrome in this family. Although Sanger sequencing across the splice junctions of *MSH2* exons 14, 15 and 16 in cDNA did not identify retention of an intronic sequence or skipping of exon 15, suggesting that the intronic insertion did not disrupt the splicing branchpoint or canonical splice acceptor, alternative mechanisms conferring deleterious impacts on gene expression, splicing or function could not be excluded based on this evidence alone.

Complex clinical cases, such as the FPC kindred described here, may similarly benefit from complementary short- and/or long-read GS. Based on the strong association between *BRCA1*, *BRCA2* and *PALB2* carrier status with HR mutational signatures, lack of DNA repair deficiency in the familial tumours suggested that undetected variants in high-penetrance HR genes were unlikely to underlie cancer susceptibility in this family. Using nanopore sequencing,

the possibility of causal germline SVs in known moderate- and high-penetrance cancer predisposition genes, especially for well-characterized pancreatic cancer susceptibility genes *BRCA2*, *ATM*, *PALB2* and *CDKN2A*, was less likely as a differential genetic diagnosis for FPC in this family.

In the absence of pathogenic variants in known disease genes, RNA-seq may provide functional information supporting variant or gene pathogenicity. Further investigation of predicted loss-of-function germline variants identified through short-read GS and evaluation of tissue-specific transcriptional profiles through RNA-seq in the FPC kindred suggested a putative association between the hepatocyte and pancreas-specific gene *PIK3C2G*, insulin resistance, and pancreatic cancer susceptibility. Independent from type II diabetes mellitus, which has been associated with pancreatic cancer both as a risk factor and consequence secondary to cancer onset, serum insulin concentration and insulin resistance have been associated with exocrine pancreatic cancer (Stolzenberg-Solomon et al. 2005; Wolpin et al. 2013). *Pik3c2g*-null mice show several metabolic phenotypes consistent with age- and diet-related insulin resistance, including lower insulin sensitivity, reduced glycogen storage and liver weight, higher circulating triglyceride levels, and reduced glycogen synthase activity (Braccini et al. 2015). Although several studies have reported possible associations between *PIK3C2G* and numerous metabolic phenotypes, such as type II diabetes, body mass index and diabetic nephropathy, further studies will be required to evaluate the role of *PIK3C2G* in aberrant glucose metabolism, insulin resistance, and pancreatic cancer (Daimon et al. 2008; Anderson et al. 2015; Hebbar et al. 2017; Saeed 2018).

Despite the current limitations of long-read sequencing, including the necessity for high molecular weight DNA, higher error rate and increased cost, this technology is particularly

beneficial in the genetic diagnosis of monogenic disorders where NGS has failed to identify a causal variant. Many nonrecurrent SVs result from template switching between homologous repetitive elements, which are inherently difficult to map with short reads. Such variants are inaccurately or incompletely captured by NGS. This was exemplified by two complex rearrangements that could only be resolved through long-read sequencing, and one false-positive inversion that was refractory to accurate interpretation based on short-read sequencing. As clinical GS becomes more widely used for molecular diagnosis in a variety of genetic syndromes, there is a need for standardized guidelines for the identification and validation of SVs using high-throughput sequencing technology. Considering the limitations of NGS, long-read sequencing offers a complementary approach in the diagnostic odyssey of patients and families where standard clinical testing is uninformative.

Chapter 5: Phenotypic characterization of gastric adenocarcinoma and proximal polyposis of the stomach

5.1   Introduction

Although multiple demographic, environmental and genetic factors contribute to gastric cancer risk, familial clustering occurs in around 10-15% of cases (Zanghieri et al. 1990). A strong genetic predisposition may underlie 1-3%, with HDGC accounting for the majority of gastric cancer kindreds. Gastric adenocarcinoma and proximal polyposis of the stomach (GAPPS) is an autosomal dominant hereditary cancer syndrome associated with profuse polyposis in the fundic gland of the stomach and sparing of the gastric antrum (Worthley et al. 2012; Li et al. 2016). Familial clustering of intestinal-type gastric cancer is observed in GAPPS and familial intestinal gastric cancer (FIGC) (Caldas et al. 1999). While the genes involved in FIGC have not been well defined, causal variants in the promoter 1B of *APC* have been identified in individuals with GAPPS and in rare families with FAP (Rohlin et al. 2011; Snow et al. 2015). These findings suggest that coding and non-coding variants in known disease genes may be associated with distinct clinical manifestations.

In spite of the uncertain penetrance of GAPPS, analysis of the *APC* promoter 1B has become routine in clinical genetic testing of hereditary cancer families. Furthermore, with few GAPPS families reported in the literature, the spectrum of clinical manifestations in carriers and the prevalence of pathogenic germline variants in the *APC* promoter are unknown. Here we present preliminary data from an international study of clinical phenotypes in GAPPS and describe three previously unreported GAPPS families. While the pathogenesis of GAPPS is still

not well understood, our findings may ultimately help inform clinical management guidelines for individuals affected by this rare cancer syndrome.

5.2 Materials and Methods

5.2.1 Ascertainment of GAPPS families

This study has been approved by the University of British Columbia Clinical Research Ethics Board (H17-01449). Ethics approval for data collection, sharing and publication was obtained independently by all collaborating institutions, including the National Institutes of Health National Cancer Institute (United States), Masaryk Memorial Cancer Institute (Czech Republic), Kumamoto University (Japan), and King Edward Memorial Hospital (Australia). Participants were ascertained by local investigators, including clinicians, genetic counselors, research nurses or other members of the study team, on the basis of a known clinical and/or molecular diagnosis of GAPPS. Individuals were eligible for the study if they met one of the following criteria:

1. the individual has a clinical diagnosis of GAPPS

2. the individual has an affected first-degree relative and is at-risk for GAPPS; or

3. the individual has a known mutation in the promoter 1B of *APC* but has not presented with fundic gland polyposis.

A clinical diagnosis of GAPPS was made according to original diagnostic criteria described by Worthley *et al.* (2012):

1. gastric polyps restricted to the body and fundus;

2. more than 100 polyps carpeting the proximal stomach in the index case or more than 30 polyps in a first-degree relative of an individual with GAPPS;

3. gastric polyps are predominantly fundic gland polyps (FGPs), some having regions of

   dysplasia (or a family member with either dysplastic FGPs or gastric adenocarcinoma);

4. an autosomal dominant pattern of inheritance; and

5. no evidence of colorectal or duodenal polyposis.

5.2.2   Questionnaire design and data collection

Survey content and design were guided by recommendations made from collaborating

clinicians, geneticists, genetic counselors and scientists with advanced knowledge of GAPPS and

hereditary cancer syndromes. Using a retrospective patient-reported survey design, we sought to

collect information about gastrointestinal symptoms and comorbidities, relevant history of

medical procedures, medication use, and diet and lifestyle factors (Appendix C). Patient-reported

questionnaires were completed either in person or by phone with a member of the research team.

Canadian participants were also eligible to complete an online version of the questionnaire.

Information regarding patient procedures and outcomes were also obtained through a review of

relevant medical records by investigators from respective collaborating institutions. All

questionnaires and pedigrees were de-identified and encoded by a unique study number.

5.2.3   *APC* promoter sequencing

This study has been approved by the BC Cancer Research Ethics Board, and written

informed consent was provided by patients or next-of-kin. Blood and tissue samples for index

cases ascertained on the basis of multiple possible eligibility criteria were received between

March 2002 and June 2013 (Caldas et al. 1999; Brooks-Wilson et al. 2004; Suriano et al. 2005;

Fitzgerald et al. 2010). Pedigrees and medical records were sent by referring centres and

reviewed centrally. Genomic DNA extracted from peripheral blood or saliva was analyzed by bidirectional Sanger sequencing across coding regions of *CDH1* as previously described. All index cases were subsequently sequenced by NGS across the exons and flanking regions of *APC*, *ATM*, *BRCA2*, *CDH1*, *CTNNA1*, *MAP3K6*, *MLH1*, *MSH2*, *MSH3*, *MSH6*, *MSR1*, *MTUS1*, *PALB2*, *PRSS1*, *PTEN*, *RAD21*, *SDHA*, *SDHB*, and *STK11*. NGS libraries were prepared using custom Illumina TruSeq or Nextera assays, and paired-end sequencing and data analysis was performed on the Illumina MiSeq. The *APC* promoter 1B was analyzed by NGS or by bidirectional Sanger sequencing using the following primer sequences: 5'-GCCAGTAAGTGCTGCAACTG-3' (F) and 5'-GGAGAGGGTGAGACATGGAG-3' (R).

5.3   Results

5.3.1   Preliminary findings from an international collaboration for GAPPS

Although several GAPPS families have been reported in the literature, the spectrum of clinical phenotypes associated with pathogenic germline variants in the *APC* promoter 1B have not been methodically characterized. To address the need for evidence-based clinical management guidelines in GAPPS, several clinicians and scientists with expertise in gastrointestinal cancer syndromes developed a patient-reported questionnaire to describe gastrointestinal symptoms, cancer spectrum and screening strategies, medication use, and potential influence of smoking history and alcohol consumption in individuals with GAPPS (Appendix C). Preliminary clinical data collected for 29 individuals from 8 families previously reported in the literature is summarized in Table 5.1 (Worthley et al. 2012; Li et al. 2016; Foretova et al. 2019).

Table 5.1. Cohort characteristics and summary of preliminary data from the GAPPS Clinical Study

	N	Age, Median (Range)	FGPs, N	Stomach and/or Abdominal Pain, N
GAPPS	**29**	**34 (12-66)**	**25 (86 %)**	**14 (48 %)**
Gastrectomy	20	38 (18-66)	20	10
H. pylori infection	5	41 (23-68)	4	1
Gastric cancer	2	43 (29-57)	2	1
Gastrointestinal symptoms				
Heartburn	14	32.5 (12-64)	12	10
Nausea/vomiting	11	30 (15-56)	9	7
Gastric reflux	8	40 (23-64)	8	6
FAP-associated phenotypes				
Colorectal polyps[a]	16	42 (19-68)	15	9
Extracolonic features[b]	4	-	4	4
Biological sex				
Female	20	-	17	9
Male	7	-	6	3
Not reported	2	-	2	2
Lifestyle factors[c]				
Regular alcohol consumption	13	-	13	6
Smoking history	6	-	4	3

[a]Reported colorectal polyp pathology includes polyps NOS, hyperplastic polyps, tubular adenomas with low-grade dysplasia, and sessile polyps.

[b]Reported FAP-associated extracolonic phenotypes include osteomas ($n = 2$), desmoid tumour ($n = 1$), enchondroma ($n = 1$) and supernumerary teeth ($n = 1$).

[c]Regular alcohol consumption is defined as more than two standard drinks per week. Smoking history is defined as daily smoking for at least six months.

Individuals were ascertained to the study based on a clinical diagnosis of GAPPS, defined by florid fundic gland polyposis with antral sparing according to guidelines by Worthley *et al.* (2011), or presence of a pathogenic variant in the *APC* promoter 1B. Among 29 individuals, fundic gland polyposis was observed in 86% ($n = 25$). The median age at which multiple fundic gland polyps were first reported by endoscopy was 34 years (range 12-66).

Gastrectomy had been performed in 20 individuals presenting with fundic gland polyposis, two of whom had a personal history of gastric cancer. Prior to a genetic diagnosis of GAPPS in one individual, prophylactic gastrectomy was performed several years after a diagnosis of Barrett's esophagus. Biannual endoscopic surveillance ultimately led to the identification of multiple fundic gland polyps with regions of dysplasia. In two unrelated individuals for whom information was available, findings of between 20-30 polyps were initially reported with massive fundic gland polyposis observed five years following initial presentation in one case. Regular surveillance by endoscopy is thus currently one of the most important strategies in reducing the risk of malignancy in families with GAPPS. When an informative endoscopy or endoscopic ultrasound is not possible, abdominal computed tomography examination can be used to detect a thickening of the gastric lining suggestive of malignancy (Bhandari et al. 2004; Akbas et al. 2019). Consequently, this procedure may be an effective approach for minimally-invasive cancer screening in individuals who are at-risk for GAPPS.

Among 14 participants (48%) reporting at least occasional stomach and/or abdominal pain prior to gastrectomy, including the two individuals with gastric cancer, many experienced co-occurring gastrointestinal symptoms, including heartburn ($n = 10$), nausea and/or vomiting ($n = 7$) and gastric reflux ($n = 6$). Among gastrointestinal conditions assessed in the current study, gastric reflux was the most common, reported in eight individuals overall. No individuals reported a history of gastric ulcers, irritable bowel syndrome, or irritable bowel disease. *H. pylori* infection was confirmed in five individuals, none of whom had been diagnosed with gastric cancer.

Due to the positive association between cigarette smoking and increased incidence of colorectal polyps, the influence of known modifiers of gastrointestinal polyposis and cancer risk

should also be evaluated in the context of hereditary cancer syndromes (Martínez et al. 1995; Shrubsole et al. 2008). Across all participants in this study, six (21%) reported a history of daily cigarette smoking for a minimum of six months and 13 (45%) reported regular alcohol consumption (Table 5.2). Preliminary investigation of the associations between smoking history, alcohol consumption and fundic gland polyposis do not show a significant increase in the incidence of fundic gland polyps, nor an earlier age at diagnosis. More than half of patient-reported questionnaires (59%) were received from members of a large Australian kindred first reported by Worthley *et al*. (2011). As such, these findings may also represent family-specific factors, such as shared genetic or environmental factors. These results should be interpreted cautiously given the small size of the cohort reported here, and ongoing efforts to describe clinical features and influence of demographic and lifestyle factors on disease presentation are required.

Table 5.2. Preliminary analysis of the influence of smoking and alcohol consumption on the presentation of gastrointestinal polyps in GAPPS

		Fundic Gland Polyps (FGPs)				Colonoscopy Findings		
	N	FGPs, N (%)	P	Age, Median (Range)	P	Procedures, N	Polyps, N (%)	P
All Participants ($N = 29$)								
Smoking History								
Regular	6	4 (67)	0.1798	29.5 (23-56)	0.5523	5	5 (100)	0.1304
Occasional or non-smoker	23	21 (91)		34 (12-66)		19	11 (58)	
Alcohol Consumption								
Regular	13	13 (100)	0.1067	33 (12-66)	0.4136	11	6 (55)	0.3905
Monthly or less	16	12 (75)		41.5 (15-65)		13	10 (77)	
Australian Kindred ($N = 17$)								
Smoking History								
Regular	3	1 (33)	0.1206	23	0.5925	2	2 (100)	0.4872
Occasional or non-smoker	14	12 (86)		32.5 (12-52)		11	6 (55)	
Alcohol Consumption								
Regular	6	6 (100)	0.2374	24 (12-52)	0.5197	5	2 (40)	0.2929
Monthly or less	11	7 (64)		41 (15-47)		8	6 (75)	
Non-Australian Participants ($N = 12$)								
Smoking History								
Regular	3	3 (100)	1	32 (27-56)	0.2639	3	3 (100)	0.4909
Occasional or non-smoker	9	9 (100)		34 (29-66)		8	5 (63)	
Alcohol Consumption								
Regular	7	7 (100)	1	34 (29-66)	0.6237	6	4 (67)	0.6084
Monthly or less	5	5 (100)		51 (27-65)		5	4 (80)	

Fisher's exact test was used to compare categorical variables, and the Wilcoxon rank-sum test was used to compare distributions of age at FGP diagnosis. Regular smokers are defined here as individuals with a history of daily smoking for at least six months. Individuals with regular alcohol consumption are defined as individuals reporting a consumption of two or more standard drinks per week.

A number of factors should be considered regarding the clinical management of GAPPS, including the limitations of endoscopic surveillance, patient-specific risk of prophylactic gastrectomy and family-specific risk of gastric cancer (Oliveira et al. 2015). The neoplastic potential of FGPs in GAPPS was demonstrated by the rapid development of gastric adenocarcinoma in an affected individual despite regular endoscopic surveillance (Repak et al. 2016). Pharmacological treatments have also been suggested as an alternative prophylactic intervention in FAP (Vasen et al. 2008). For example, non-steroidal anti-inflammatory drugs (NSAIDs) have been shown to reduce the number of colorectal adenomas in FAP (Asano and McLeod 2004). However, significant gastrointestinal and cardiovascular side effects have prevented their widespread use. The influence of NSAIDs on the progression of polyposis in FAP suggests that pharmacological treatments may act as modifiers of disease progression in GAPPS as well. In the future, exploring environmental influences on disease expression may identify non-invasive interventions that could be considered in the prophylactic management of GAPPS.

Across GAPPS families known to date, including two previously unreported kindreds, the median age of gastric cancer diagnosis was 50 years (range 24-75, $n = 35$) (Figure 5.1). The earliest known gastric cancer occurrence associated with GAPPS is 24 years. Among 76 reported individuals with confirmed fundic gland polyposis, 26 individuals (34%) had a personal history of gastric cancer. These may be independent of other known risk factors, as the two affected individuals in the current study did not have prior *H. pylori* infection or other primary cancer diagnoses. Ascertainment bias for families with strong cancer history indicates a need for ongoing ascertainment of affected families to estimate the true cancer penetrance in GAPPS. Given the potential for early onset and uncertain lifetime risk of gastric adenocarcinoma in

GAPPS patients, a GAPPS diagnosis in individuals who are at risk can also be confirmed by genetic testing and inform clinical management.
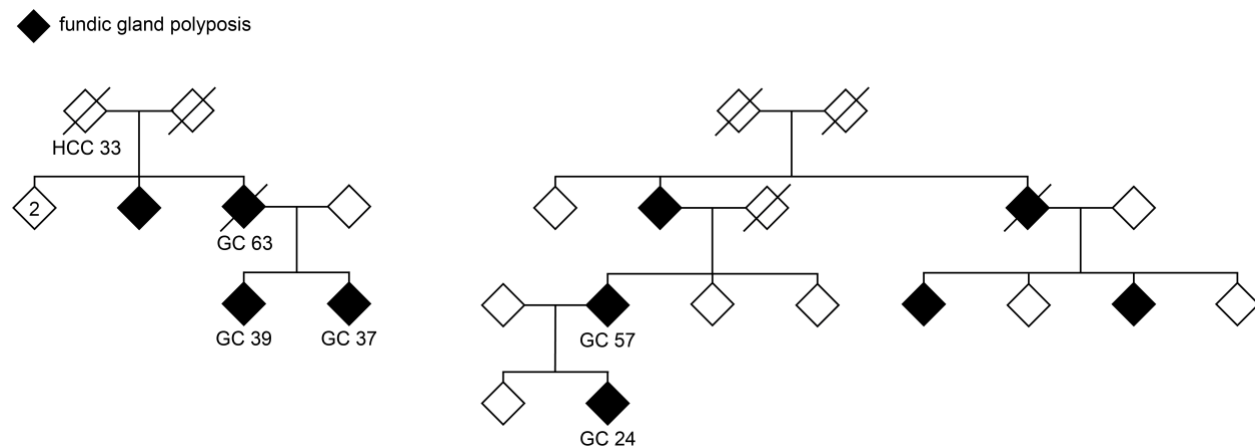
Figure 5.1. Two GAPPS families not previously reported in the literature. Individuals with reported fundic gland polyps are shaded in black and known cancer diagnoses are indicated below. GC, gastric cancer; HCC, hepatocellular carcinoma.

In families with FAP, it is recommended that carriers have periodic examinations of the rectosigmoid from their early teens and of the upper gastrointestinal tract from their late twenties (Vasen et al. 2008). Screening by sigmoidoscopy is recommended every 2 years from age 10, and if an adenoma is detected, annual colonoscopies should be performed until a colectomy is planned. The average age of onset of colorectal cancer in classic FAP is 39 years, and 55 years in attenuated FAP (AFAP). Among published reports of GAPPS families, the earliest polyp phenotype was observed at 10 years and the earliest gastric cancer reported at 31 years. Similar recommendations may therefore be considered in the surveillance of GAPPS-associated FGPs. Our preliminary findings demonstrate some of the challenges in the identification of GAPPS families, with later diagnosis in individuals without a personal history of gastric cancer. This

may result from the absence of specific symptoms associated with fundic gland polyposis or ascertainment bias for families with a strong history of gastric cancer. Given the earliest known gastric cancer diagnosis at 24 years of age, carrier testing and endoscopic screening in confirmed GAPPS families may also be warranted in the teenage years.

5.3.2   Rare *APC* promoter variants in gastric cancer kindreds unselected for polyposis

To investigate the contribution of *APC* promoter variants to gastric cancer predisposition in *CDH1*-negative families, we sequenced the coding regions of several known cancer predisposition genes and *APC* promoter 1B in 259 individuals from 254 families with a personal and/or family history of gastric cancer and who were unselected for gastric polyposis (Table 5.3). This included 174 individuals meeting IGCLC criteria for HDGC and one meeting criteria for FIGC (Fitzgerald et al. 2010; van der Post et al. 2015). The majority (76.8%) of individuals had a personal history of gastric cancer, with 85.4% diffuse gastric cancer and median age of diagnosis of 42 years (range 9-87). An additional 6 individuals were potential obligate carriers for hereditary gastric cancer predisposition syndromes.

Table 5.3. Personal and family gastric cancer history in *CDH1*-negative index cases unselected for fundic gland polyposis

Cancer History	Index Cases	Family History of GC, No. of Index Cases[a]			
		HDGC	FIGC	Any GC	None
Personal history of GC[b]	199	149	1	16	33
Other cancer history[c]					
Obligate carrier	2	2	0	0	0
Non-obligate carrier	38	12	0	26	0
Unaffected					
Obligate carrier	4	4	0	0	0
Non-obligate carrier	16	7	0	9	0
Total	259	174	1	51	33

Abbreviations: GC, gastric cancer; HDGC, hereditary diffuse gastric cancer; FIGC, familial intestinal gastric cancer
[a] Family history of GC in first- and second-degree relatives.
[b] Index case GC subtypes: DGC ($n = 170$), IGC ($n = 10$), mixed ($n = 4$), NOS ($n = 15$).
[c] Other cancer types: breast ($n = 31$), colon ($n = 4$), ovarian ($n = 1$), prostate ($n = 2$), skin ($n = 2$), thymoma ($n = 1$), uterine ($n = 1$). Two index cases were affected by more than one cancer type.

We identified a previously reported pathogenic GAPPS-associated variant (*APC* c.-191T>C) in an individual meeting clinical IGCLC criteria for HDGC (Figure 5.2). Prior genetic screening of *CDH1* was uninformative and no pathogenic variants were identified in other cancer predisposition genes tested. The index case (III-8) was diagnosed with moderately differentiated stage IB prostate cancer at the age of 73, following a diagnosis of gastric cancer in two children. IV-2 initially presented with lower abdominal pain, distension and ascites at 37 years of age. Upper gastrointestinal endoscopy revealed a gastric mass and multiple 3 mm polypoid lesions throughout the stomach and fundus with sparing of the distal half of the gastric antrum. The patient subsequently succumbed to a stage IV diffuse gastric cancer within three weeks of their initial presentation. IV-4 presented with severe abdominal pain, anorexia, and emesis at 39 years

of age and had guaiac-positive stool upon admission to hospital. Tumour metastases of unknown

origin were identified in the liver, but the patient passed away prior to the diagnosis of a primary

intestinal-type gastric cancer identified upon autopsy. Notably, despite diffuse tumour

involvement in the gastric mucosa, coarsely granular to polypoid texture was observed and

suggested the possibility of precancerous gastric polyposis.
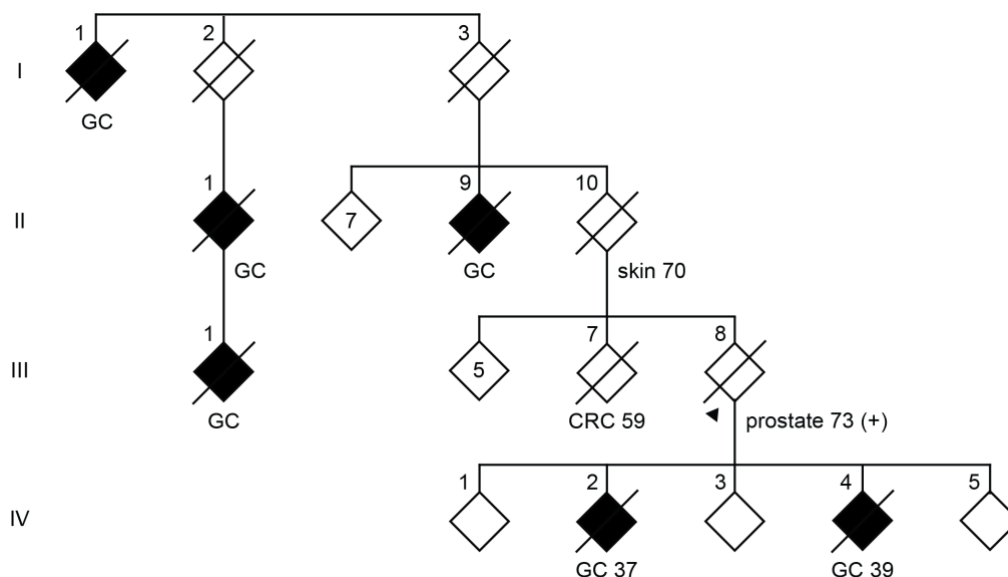
Figure 5.2. Pedigree of an unreported GAPPS family identified retrospectively in a familial gastric cancer cohort unselected for polyposis. GC, gastric cancer; CRC, colorectal cancer.

Unfortunately, we were unable to confirm segregation of the *APC* c.-191T>C variant in

this family, nor were we able to assess the presence of florid gastric polyposis in the index case.

However, fundic gland polyposis with antral sparing identified in one child and possible gastric

polyposis in another suggests an association between the *APC* c.-191T>C variant and

characteristic GAPPS phenotype in this family. Non-gastric cancer types have been occasionally

reported in GAPPS families, including a personal history of thyroid cancer in one affected

individual, in addition to a personal history of prostate cancer in the index case identified in our study (Worthley et al. 2012). It is unclear whether these reflect coincidental sporadic cancer occurrences or if carriers are at an increased risk for developing other gastrointestinal or non-gastric cancer types.

5.4   Discussion

Hereditary cancer susceptibility is commonly observed in unselected cohorts of cancer patients that do not meet current clinical testing guidelines (Schrader et al. 2016; Mandelker et al. 2017). In precision oncology, the utility of exploring inherited genetic variation is achieved through the implementation of cancer prevention and screening strategies and by the use of targeted therapies. However, the clinical significance of germline variants in cancer predisposition genes and accurate estimation of associated lifetime cancer risks across various populations is influenced by individual- and gene-specific factors. Recent variant interpretation guidelines that consider variant impact, gene function, incomplete penetrance, and variable expressivity in specific disease and gene contexts have been shown to improve the accuracy of variant assessment for high-penetrance cancer syndromes (Leroy et al. 2017; Lee et al. 2018). Multidisciplinary and collaborative approaches to genetic diagnosis across rare disease have improved clinical outcomes for patients and their families, indicating that similar initiatives can provide a basis for accurate estimates of cancer penetrance in individuals with rare cancer syndromes (Boycott et al. 2017; Wright et al. 2018).

Causal variants in GAPPS have thus far been restricted to those in the promoter 1B of *APC*, having been identified in most reported families to date. Several SNVs in the *APC* promoter have been associated with reduced transcription factor binding in both GAPPS and

FAP, including GAPPS-associated alleles c.-191T>C, c.-192A>G and c.-195A>C and FAP-associated alleles c.-190G>A and c.-192A>T (Li et al. 2016). However, the clinical heterogeneity and cancer risks associated with *APC* promoter 1B variants have not been fully elucidated. Variability in the expression of the syndrome, particularly the presence of non-gastric phenotypes in affected individuals, indicates the need to further characterize the spectrum of clinical features associated with GAPPS. These include the influence of known risk factors for fundic gland polyps, such as long-term use of proton pump inhibitors, on polyp incidence and age of onset in carriers (Tran-Duy et al. 2016).

While the spectrum of clinical phenotypes and cancer risk associated with GAPPS are not well established, several independent reports of affected families indicate that GAPPS is specifically associated with extensive fundic gland polyposis and incomplete cancer penetrance (Worthley et al. 2012; Yanaru-Fujisawa et al. 2012; Li et al. 2016; Repak et al. 2016; Beer et al. 2017; Anderson et al. 2018; Mitsui et al. 2018; Foretova et al. 2019). As part of ongoing work characterizing the molecular basis of hereditary gastric cancer, we investigated the prevalence of *APC* promoter variants in over 250 individuals with a personal and/or family history of gastric cancer with no known genetic diagnosis. Surprisingly, a single occurrence of a GAPPS-associated variant was found retrospectively in an obligate carrier for autosomal dominant cancer predisposition. Based on current clinical criteria, a history of gastric polyposis and gastric cancer in the kindred described here would indicate genetic assessment for GAPPS. Rare families meeting criteria for HDGC may thus harbour variants in the *APC* promoter 1B, and retrospective assessment of gastric polyposis may be warranted to determine if these families present with the characteristic GAPPS phenotype. Our findings indicate that GAPPS may account for gastric cancer susceptibility in rare families meeting clinical criteria for other cancer predisposition

syndromes, and thus genetically undiagnosed gastric cancer families with any history of fundic gland polyposis should undergo genetic testing of *APC* to exclude the possibility of GAPPS.

Given a common molecular basis in GAPPS, FAP and AFAP, possible colonic involvement has not been excluded in individuals carrying GAPPS-associated variants given the inconsistent reporting of colonoscopy findings. Among 27 patients described in the current study who were examined by colonoscopy, a small number of colorectal polyps (< 10) was reported in 63%. However, it is unclear whether these reflect unrelated phenotypes given an estimated incidence of 21.1% in the general population and advanced polyps (> 9 mm) occurring in 6-8% (Imperiale et al. 2002; Lieberman et al. 2008). Demographic factors such as age, sex, and ethnicity may also contribute to the incidence of colorectal findings in GAPPS families, and these should be considered in larger cohorts. Continuing global collaborative efforts between clinical and research teams will be necessary to describe clinical phenotypes associated with this rare cancer predisposition syndrome.

Characterizing hereditary genetic variation in high- and moderate-penetrance cancer predisposition genes may have implications for cascade carrier testing, cancer risk-reduction and screening interventions in individuals at risk of having inherited a causal germline variant. Therefore, germline GS has the potential to identify previously unknown carriers for genes with defined estimates of lifetime cancer risk, as well as causal genetic variants in suspected hereditary cancer families with otherwise uninformative clinical genetic testing. As demonstrated by us and others, GS may improve molecular diagnosis of cancer predisposition syndromes in comparison to panel testing through the detection of SVs or genetic variation in novel disease genes. Germline variants identified in the promoters of high-penetrance cancer predisposition genes, such as *APC*, further implicate a role for *cis*-regulatory variation in cancer risk. An

accurate molecular diagnosis is critical in order to evaluate cancer and non-cancer phenotypes in carriers, as well as to inform evidence-based guidelines regarding effective cancer screening and prophylactic interventions.

Chapter 6: Conclusion

6.1   Summary

The landscape of natural genetic variation is complex, consisting of small genetic variants and large genomic rearrangements. During the last two decades, advancements in sequencing technologies have allowed the characterization of genetic variants at high resolution across the human genome. This has improved our understanding of how germline variation contributes to both complex and Mendelian diseases, but our incomplete biological and clinical knowledge has been limiting in the interpretation of rare variants underlying cancer predisposition syndromes. The research presented in this dissertation demonstrates that tumour genome and transcriptome sequencing may improve the interpretation of germline genetic variation. In particular, these allow functional characterization of known and novel cancer predisposition genes, identification and resolution of large or complex genetic variants, and detailed phenotypic characterization for specific disease and gene contexts. Using genome and transcriptome sequencing, I explored the molecular mechanisms associated with high-penetrance genes and showed that individual cancer genomes and tissue-specific *in vitro* models can provide insights into the genetic and molecular aetiology of cancer predisposition syndromes. Although the detection and interpretation of certain types of genetic variation is restricted by current limitations of short-read sequencing, long-read sequencing improves the resolution of germline SVs in suspected hereditary cancer families. Finally, I discussed preliminary findings from individuals with GAPPS, a recently described syndrome accounting for rare gastric cancer families. This work supports ongoing collaborative efforts in the molecular and phenotypic

characterization of high-penetrance cancer predisposition syndromes to improve genetic diagnosis and inform evidence-based clinical guidelines in the era of precision medicine.

6.2   Significance

Molecular tumour profiling may indicate the contribution of classical tumour suppressor genes in disease pathogenesis; therefore, sequencing tumour tissue from individuals with constitutional cancer susceptibility may help identify candidate causal genes. MMR deficiency is an uncommon molecular feature observed across a broad range of cancer types, assessed universally in CRC and EC to identify potential Lynch syndrome families. Loss of MMR proteins MLH1, MSH2, MSH6 or PMS2 assessed by IHC can indicate the presence of a germline variant or result from double somatic alterations, indicating a likely sporadic cancer occurrence. Through targeted tumour-normal NGS, we found that more than half of MMR-deficient Lynch syndrome spectrum tumours reflect a likely sporadic occurrence resulting from hypermethylation of *MLH1* or somatic genetic alterations (Chapter 2). These findings ultimately helped inform the development of a modified framework of provincial genetic testing, integrating tumour sequencing into a universal screening paradigm for Lynch syndrome. This approach may also improve genetic diagnosis of families with strong phenotypic and/or molecular indications of Lynch syndrome and where somatic mutations, copy number alterations or LOH cannot explain MMR deficiency.

Although short-read GS allows the identification of small genetic variants and copy number variants with high sensitivity, it is limited in the accurate resolution of complex SVs and variants in repetitive regions of the genome. Long-read sequencing, such as nanopore sequencing or single molecule real-time sequencing, improves the interpretation of germline variants that

remain elusive to short-read GS, thus allowing appropriate cancer risk stratification of carriers (Chapter 4). Complementary short- and long-read sequencing may also improve differential genetic diagnosis in suspected hereditary cancer families to evaluate potential variants in known cancer predisposition genes. Through the exclusion of known cancer predisposition genes, integrated tumour and germline analysis in a molecularly undiagnosed FPC kindred suggested a putative role for altered glucose metabolism in cancer susceptibility associated with a candidate moderate-penetrance gene.

Given the potential information and ethical implications resulting from germline GS, we recently described a framework for the translation of primary and secondary germline findings in precision oncology that was informed by retrospective germline analysis and several years of clinical, molecular and informatics experience in the POG program (Dixon et al. 2020). Owing to the potential implications for genetic diagnosis, cancer risk reduction and targeted cancer therapy, primary germline findings, defined as genetic variation with known or potential implications for cancer susceptibility, prognosis, or treatment, should be an integral part of patient education in cancer genomics programs. Among advanced cancer patients unselected for personal or family cancer history, germline variants with established clinical actionability for cancer susceptibility were identified in 9.6% (Chapter 2). However, many individuals with germline variants conferring dominant cancer predisposition would not be eligible for provincially-funded genetic testing based on current guidelines, indicating the challenges of clinical phenotype-based guidelines in carrier ascertainment. To contribute to the phenotypic characterization of rare cancer predisposition syndromes, we established an international collaboration for GAPPS that continues to assess disease penetrance and cancer screening strategies in affected families (Chapter 5).

6.3   Limitations

6.3.1   Candidate gene discovery through GS

Variant classification remains diagnostically challenging in the hereditary cancer setting given the frequent lack of phenotypic or functional information supporting variant pathogenicity. In the absence of pathogenic variants in known disease genes, genetic linkage analysis in multigenerational pedigrees is a powerful approach to identify candidate genes underlying high-penetrance cancer syndromes. However, several possible confounding factors may restrict the use of this approach, including reduced disease penetrance, presence of phenocopies, pedigree size or structure, and genotyping or pedigree errors (Ott et al. 2015). Patient ascertainment may also limit the availability of genotype data, a limitation of our analysis in a multigenerational FPC kindred (Chapter 4). Despite genome-wide variant calling using short- and long-read sequencing platforms, segregation of candidate variants could not be evaluated in affected family members with higher degrees of separation from the siblings assessed here. GS allowed the identification of several candidate protein-truncating variants that segregated with disease in multiple siblings with PDAC, but the reproducibility of these findings has not yet been investigated in additional families.

6.3.2   Non-coding variants

As variants in *cis*-regulatory elements may promote or reduce transcript expression, sequence analysis for the *APC* promoter 1B and deletion and duplication analysis for the promoters of *BMPR1A*, *GREM1*, *MLH1*, *MSH2*, *PTEN* and *TP53* has been incorporated into current clinical assays (Shin et al. 2002; Zhou et al. 2003; Calva-Cerqueira et al. 2010; Morak et

al. 2011; Leclerc et al. 2018). Global allele-specific expression contributes to normal human variation, as approximately 88% of protein-coding genes show tissue-specific allelic imbalance across individuals from the general population (Stranger et al. 2017). Allelic imbalance of highly penetrant cancer predisposition genes has also been observed in individuals affected by hereditary cancer syndromes and may further indicate the presence of undetected genetic or epigenetic variation affecting their transcription (Chen et al. 2008). Functional annotations aid in the interpretation of non-coding variants identified through GS that may result in allelic imbalance of candidate causal genes. However, the availability of high-quality tissue specimens, presence of heterozygous coding SNPs, and individual variation in the level of gene expression determine the feasibility of assessing allele-specific expression in individuals with potential cancer predisposition syndromes. These factors limited our assessment of the transcriptional consequences of germline SVs in known and suspected carriers (Chapter 4).

6.3.3 Tissue-specific expression

Accurate expression quantification depends on the availability of samples from the tissue of disease origin, and thus practical considerations may currently limit the use of RNA-seq in molecular diagnosis. For cancer patients, RNA-seq of primary tumour tissue may allow unbiased tissue-specific analysis of gene expression, alternative splicing and allele-specific expression. Incorporating tissue- and disease-specific comparators into the analysis of individual tumours from an undiagnosed kindred ultimately helped characterize a novel possible mechanism underlying moderate-penetrance pancreatic cancer susceptibility (Chapter 4). However, somatic mutations resulting in NMD, somatic copy number alterations, and tumour-specific DNA methylation may all influence gene expression in tumour tissue. Organoid model systems offer

an alternative strategy to studying biological mechanisms of cancer predisposition in natural tissue contexts. This approach identified a potential molecular marker of SRC morphology in HDGC, although broader characterization of gastric cell types involved in neoplastic transformation were limited by inherent developmental and anatomical differences between mouse and human (Chapter 3).

6.3.4 Phenotypic characterization of rare syndromes

Evolving guidelines regarding the analysis, interpretation and protection of germline data will inform future clinical translation of hereditary genetic information. Precision medicine initiatives such as the POG program offer opportunities for carrier ascertainment, allowing clinical referral that may not have been otherwise indicated. Given the advanced nature of disease among cancer patients in the POG program, the therapeutic implications of germline variants in cancer susceptibility genes, in particular if knowledge of a germline variant resulted in the consideration or implementation of a change in treatment, could not be assessed. However, the potential implications of germline variants with established therapeutic associations was recognized in cancer genomes characterized by HR deficiency and MSI associated with HBOC and Lynch syndrome, respectively.

Recent evidence-based guidelines for gene- and disease-specific variant curation support a need for collaborative, transparent, and dynamic clinical research initiatives to improve genetic diagnosis for cancer predisposition syndromes. Accurate assessment of patient and family history is important to identify individuals with inherited cancer susceptibility and may be especially informative for clinical variant interpretation (Murff et al. 2004). Accordingly, larger numbers of affected families and variant segregation in multigenerational pedigrees provide

increasing levels of support for variant pathogenicity (Lee et al. 2018). Given the limited number of GAPPS families identified to date, phenotypic characterization of this rare syndrome remains ongoing.

6.4   Future directions

6.4.1   Beyond high-penetrance cancer predisposition genes

BRCA1 and *BRCA2* together account for around 25% of HBOC families, indicating a need to further characterize the molecular basis and clinical implications of germline variants in other known or candidate cancer susceptibility genes. HR is a DSBR mechanism using extensive regions of sequence homology from homologous or sister chromosomes as a template for effective DNA repair. Several moderate- to high-penetrance cancer predisposition genes are involved at various stages of HR, including *BRCA1*, *BRCA2*, *ATM*, *CHEK2* and *PALB2*, among others. Through its multifunctional role in the DNA damage response, *ATM* is an important regulator of DSBR pathways including HR and NHEJ. Biallelic germline variants in *ATM* cause ataxia-telangiectasia, an autosomal recessive syndrome characterized by ataxia, weakened immune systems and susceptibility to leukemia and lymphoma. Despite their critical roles in the DNA damage response, heterozygous variants in *ATM* and *CHEK2* have only been associated with modest increases in lifetime cancer risk.

Germline variants in moderate-penetrance genes involved in DSBR were not associated with disruption of HR assessed by the presence of mutational signatures in paired tumour tissues (Chapter 2). This finding, previously reported in breast cancer cohorts, suggests that alternative molecular mechanisms or other genetic or non-genetic factors contribute to tumour progression and cancer susceptibility in these cases (Mandelker et al. 2019). Similar to high-penetrance

variants in *BRCA1* and *BRCA2*, pathogenic germline variants in moderate-penetrance cancer predisposition genes may be associated with specific somatic events and molecular phenotypes; however, these require further investigation. Although *PALB2* has been associated with an increased risk for cancers within the HBOC spectrum, including female breast, pancreatic, ovarian and male breast cancers, and was associated with HR deficiency in advanced breast cancers described here, the biological mechanisms differentiating *BRCA1-*, *BRCA2-* and *PALB2-* related cancer susceptibility are still unclear (Yang et al. 2020).

Given the role of *BRCA2* and *PALB2* in both the HR and FA pathways, multiple studies have investigated the prevalence of pathogenic variants in other FA genes in breast cancer families without causal variants in *BRCA1* or *BRCA2* (Seal et al. 2003; Thompson et al. 2012; Litim et al. 2013). However, without clear associations between carrier status for Fanconi anemia genes and cancer susceptibility, the clinical significance of monoallelic variants in these genes for individual cancer risk has not been well defined. Four carriers of pathogenic germline variants in *FANCA* and *FANCC* were identified among unselected patients in the POG program, but none showed specific associations with characterized mutational signatures implicating these genes in tumourigenesis.

6.4.2   Population genetic testing

Given the potential opportunities for clinical intervention in high-penetrance cancer predisposition syndromes, identification and molecular diagnosis of individuals with an increased lifetime risk for cancer may allow cancer prevention, enhanced cancer screening, and improved health outcomes in patients and their families. An estimated one in 400 individuals is a carrier for *BRCA1* or *BRCA2* in the general population (Kast et al. 2016). Although the current semi-

opportunistic model for genetic testing aims to offer genetic testing to individuals with the greatest probability of harbouring actionable germline variants, only 38-45% of carriers for these actionable genes identified through population genetic testing meet clinical genetic testing criteria (Metcalfe et al. 2010; Metcalfe et al. 2013). This may reflect several existing barriers to the access of genetic services for individuals either with or without a family history suggestive of HBOC, including family structure, inaccurate information about family cancer history and lack of awareness or support from primary health care providers (Lieberman et al. 2017). This approach may miss many carriers who would otherwise benefit from early detection and personalized cancer risk management.

Population-based genetic testing for *BRCA1* and *BRCA2* in all women currently meets the World Health Organization's criteria for population screening for genetic predisposition to disease (Wilson et al. 1968; King et al. 2014). Lifetime cancer risks associated with *BRCA1* and *BRCA2* are between 46-87% and 38-84% for female breast cancer and 39-63% and 16.5-27% for ovarian cancer, respectively (Petrucelli et al. 1993). An increased risk for several other cancer types has also been associated with BRCA carrier status, including male breast cancer, pancreatic cancer and prostate cancer. Population screening for three common founder variants in *BRCA1* and *BRCA2* in the AJ population increased the number of carriers identified by 40-63% and was shown to be cost-effective when compared to current delivery models of genetic testing based on personal or family cancer history (Manchanda et al. 2015). Breast and ovarian cancer risks are similar among carriers of pathogenic variants in *BRCA1* and *BRCA2* obtained through population genetic screening or based on family history, indicating that carriers identified through population screening may equally benefit from enhanced cancer prevention and screening strategies (King et al. 2003; Gabai-Kapara et al. 2014). An effective model of delivery for population genetic testing

in cancer predisposition genes is ultimately influenced by the nature of the health care system (Foulkes et al. 2016). Implementation of genetic testing as part of routine population-based cancer screening programs may thus increase the accessibility of genetic testing for common cancer predisposition syndromes, such as HBOC or Lynch syndrome, to improve health outcomes and individual health empowerment.

6.5   Conclusions

Overall, our findings suggest that integrated molecular analysis may improve rates of genetic diagnosis, help characterize the functional significance of genetic variation, and allow opportunities for increased cancer screening and cancer prevention. Complementary approaches to genetic testing can inform the identification and diagnosis of individuals with high-penetrance cancer predisposition syndromes, while transcriptome sequencing provides functional information about the impact of genetic variants at the level of genes, cells or tissues. GS technologies currently offer an unbiased analysis of genetic variation. However, significant ethical considerations are also highlighted by the challenges in large-scale germline analysis, including variant interpretation, clinical translation, and privacy and protection of heritable genetic information. These necessitate the development of evidence-based guidelines for integrating genomics technologies into clinical health delivery models that include genetics education among patients and non-specialist health care providers.

References

Aaltonen LA, Peltomäki P, Leach FS, Sistonen P, Pylkkänen L, Mecklin JP, Järvinen H, Powell SM, Jen J, Hamilton SR, et al. 1993. Clues to the pathogenesis of familial colorectal cancer. Science (80-). 260(5109):812–816.

Aaltonen LA, Salovaara R, Kristo P, Canzian F, Hemminki A, Peltomäki P, Chadwick RB, Kääriäinen H, Eskelinen M, Järvinen H, et al. 1998. Incidence of Hereditary Nonpolyposis Colorectal Cancer and the Feasibility of Molecular Screening for the Disease. N Engl J Med. 338(21):1481–1487.

Akbas A, Bakir H, Dasiran MF, Dagmura H, Ozmen Z, Celtek NY, Daldal E, Demir O, Kefeli A, Okan I. 2019. Significance of Gastric Wall Thickening Detected in Abdominal CT Scan to Predict Gastric Malignancy. J Oncol. 2019.

Alexandrov LB, Kim J, Haradhvala NJ, Huang MN, Tian Ng AW, Wu Y, Boot A, Covington KR, Gordenin DA, Bergstrom EN, et al. 2020. The repertoire of mutational signatures in human cancer. Nature. 578(7793):94–101.

Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SAJR, Behjati S, Biankin A V., Bignell GR, Bolli N, Borg A, Børresen-Dale AL, et al. 2013. Signatures of mutational processes in human cancer. Nature. 500(7463):415–421.

Altshuler DM, Durbin RM, Abecasis GR, Bentley DR, Chakravarti A, Clark AG, Donnelly P, Eichler EE, Flicek P, Gabriel SB, et al. 2012. An integrated map of genetic variation from 1,092 human genomes. Nature. 491(7422):56–65.

Anderson A, Swanson L, Plummer R, Abraham J. 2018. Identifying the GAPPS in Hereditary Gastric Polyposis Syndromes: 2707. Am J Gastroenterol. 113.

Anderson D, Cordell HJ, Fakiola M, Francis RW, Syn G, Scaman ESH, Davis E, Miles SJ, McLeay T, Jamieson SE, et al. 2015. First genome-wide association study in an Australian Aboriginal population provides insights into genetic risk factors for body mass index and type 2 diabetes. PLoS One. 10(3):e0119333.

Angèle S, Falconer A, Edwards SM, Dörk T, Bremer M, Moullan N, Chapot B, Muir K, Houlston R, Norman AR, et al. 2004. ATM polymorphisms as risk factors for prostate cancer development. Br J Cancer. 91(4):783–787.

Antoniou AC, Casadei S, Heikkinen T, Barrowdale D, Pylkäs K, Roberts J, Lee A, Subramanian D, De Leeneer K, Fostira F, et al. 2014. Breast-Cancer Risk in Families with Mutations in *PALB2*. N Engl J Med. 371(6):497–506.

Antoniou AC, Sinilnikova OM, Simard J, Léoné M, Dumont M, Neuhausen SL, Struewing JP, Stoppa-Lyonnet D, Barjhoux L, Hughes DJ, et al. 2007. RAD51 135G→C modifies breast cancer risk among BRCA2 mutation carriers: Results from a combined analysis of 19 studies. Am J Hum Genet. 81(6):1186–1200.

Armanios M, Chen J-L, Chang Y-PC, Brodsky RA, Hawkins A, Griffin CA, Eshleman JR, Cohen AR, Chakravarti A, Hamosh A, et al. 2005. Haploinsufficiency of telomerase reverse transcriptase leads to anticipation in autosomal dominant dyskeratosis congenita.

Asano TK, McLeod RS. 2004. Non steroidal anti-inflammatory drugs (NSAID) and Aspirin for preventing colorectal adenomas and carcinomas. Cochrane database Syst Rev.(2):CD004079.

Atchley DP, Albarracin CT, Lopez A, Valero V, Amos CI, Gonzalez-Angulo AM, Hortobagyi GN, Arun BK. 2008. Clinical and pathologic characteristics of patients with BRCA-positive and BRCA-negative breast cancer. J Clin Oncol. 26(26):4282–4288.

Auton A, Abecasis GR, Altshuler DM, Durbin RM, Bentley DR, Chakravarti A, Clark AG, Donnelly P, Eichler EE, Flicek P, et al. 2015. A global reference for human genetic variation. Nature. 526(7571):68–74.

Bahrenberg G, Brauers A, Joost H-G, Jakse G. 2000. Reduced Expression of PSCA, a Member of the LY-6 Family of Cell Surface Antigens, in Bladder, Esophagus, and Stomach Tumors. Biochem Biophys Res Commun. 275(3):783–788.

Barber ME, Save V, Carneiro F, Dwerryhouse S, Lao-Sirieix P, Hardwick RH, Caldas C, Fitzgerald RC. 2008. Histopathological and molecular analysis of gastrectomy specimens from hereditary diffuse gastric cancer patients has implications for endoscopic surveillance of individuals at risk. J Pathol. 216(3):286–94.

Bartsch DK, Gress TM, Langer P. 2012. Familial pancreatic cancerĝ"current knowledge. Nat Rev Gastroenterol Hepatol. 9(8):445–453.

Bass AJ, Thorsson V, Shmulevich I, Reynolds SM, Miller M, Bernard B, Hinoue T, Laird PW, Curtis C, Shen H, et al. 2014. Comprehensive molecular characterization of gastric adenocarcinoma. Nature. 513(7517):202–209.

Beck AC, Yuan H, Liao J, Imperiale P, Shipley K, Erdahl LM, Sugg SL, Weigel RJ, Lizarraga IM. 2020. Rate of BRCA mutation in patients tested under NCCN genetic testing criteria. In: American Journal of Surgery. Vol. 219. Elsevier Inc. p. 145–149.

Beer A, Streubel B, Asari R, Dejaco C, Oberhuber G. 2017. Gastric adenocarcinoma and proximal polyposis of the stomach (GAPPS) – a rare recently described gastric polyposis syndrome – report of a case. Z Gastroenterol. 55(11):1131–1134.

Bell DW, Varley JM, Szydlo TE, Kang DH, Wahrer DCR, Shannon KE, Lubratovich M,

Auton A, Abecasis GR, Altshuler DM, Durbin RM, Bentley DR, Chakravarti A, Clark AG, Donnelly P, Eichler EE, Flicek P, et al. 2015. A global reference for human genetic variation. Nature. 526(7571):68–74.

Bahrenberg G, Brauers A, Joost H-G, Jakse G. 2000. Reduced Expression of PSCA, a Member of the LY-6 Family of Cell Surface Antigens, in Bladder, Esophagus, and Stomach Tumors. Biochem Biophys Res Commun. 275(3):783–788.

Barber ME, Save V, Carneiro F, Dwerryhouse S, Lao-Sirieix P, Hardwick RH, Caldas C, Fitzgerald RC. 2008. Histopathological and molecular analysis of gastrectomy specimens from hereditary diffuse gastric cancer patients has implications for endoscopic surveillance of individuals at risk. J Pathol. 216(3):286–94.

Bartsch DK, Gress TM, Langer P. 2012. Familial pancreatic cancerĝ"current knowledge. Nat Rev Gastroenterol Hepatol. 9(8):445–453.

Bass AJ, Thorsson V, Shmulevich I, Reynolds SM, Miller M, Bernard B, Hinoue T, Laird PW, Curtis C, Shen H, et al. 2014. Comprehensive molecular characterization of gastric adenocarcinoma. Nature. 513(7517):202–209.

Beck AC, Yuan H, Liao J, Imperiale P, Shipley K, Erdahl LM, Sugg SL, Weigel RJ, Lizarraga IM. 2020. Rate of BRCA mutation in patients tested under NCCN genetic testing criteria. In: American Journal of Surgery. Vol. 219. Elsevier Inc. p. 145–149.

Beer A, Streubel B, Asari R, Dejaco C, Oberhuber G. 2017. Gastric adenocarcinoma and proximal polyposis of the stomach (GAPPS) – a rare recently described gastric polyposis syndrome – report of a case. Z Gastroenterol. 55(11):1131–1134.

Bell DW, Varley JM, Szydlo TE, Kang DH, Wahrer DCR, Shannon KE, Lubratovich M,

Verselis SJ, Isselbacher KJ, Fraumeni JF, et al. 1999. Heterozygous germ line hCHK2 mutations in Li-Fraumeni syndrome. Science (80-). 286(5449):2528–2531.

Bertelsen B, Tuxen IV, Yde CW, Gabrielaite M, Torp MH, Kinalis S, Oestrup O, Rohrberg K, Spangaard I, Santoni-Rugiu E, et al. 2019. High frequency of pathogenic germline variants within homologous recombination repair in patients with advanced cancer. npj Genomic Med. 4(1):1–11.

Bhandari S, Shim CS, Kim JH, Jung IS, Cho JY, Lee JS, Lee MS, Kim BS. 2004. Usefulness of three-dimensional, multidetector row CT (virtual gastroscopy and multiplanar reconstruction) in the evaluation of gastric cancer: A comparison with conventional endoscopy, EUS, and histopathology. Gastrointest Endosc. 59(6):619–626.

Bjerknes M, Cheng H. 2006. Neurogenin 3 and the enteroendocrine cell lineage in the adult mouse small intestinal epithelium. Dev Biol. 300(2):722–735.

Blokzijl F, Janssen R, van Boxtel R, Cuppen E. 2018. MutationalPatterns: comprehensive genome-wide analysis of mutational processes. Genome Med. 10(1):33.

De Bock GH, Schutte M, Krol-Warmerdam EMM, Seynaeve C, Blom J, Brekelmans CTM, Meijers-Heijboer H, Van Asperen CJ, Cornelisse CJ, Devilee P, et al. 2004. Tumour characteristics and prognosis of breast cancer patients carrying the germline CHEK2*1100delC variant. J Med Genet. 41(10):731–735.

Boeva V, Popova T, Bleakley K, Chiche P, Cappo J, Schleiermacher G, Janoueix-Lerosey I, Delattre O, Barillot E. 2012. Control-FREEC: A tool for assessing copy number and allelic content using next-generation sequencing data. Bioinformatics. 28(3):423–425.

Boland CR, Goel A. 2010. Microsatellite Instability in Colorectal Cancer. Gastroenterology.

138(6).

Boland CR, Lynch HT. 2013. The history of Lynch syndrome. Fam Cancer. 12(2):145–157.

Boland CR, Thibodeau SN, Hamilton SR, Sidransky D, Eshleman JR, Burt RW, Meltzer SJ, Rodriguez-Bigas MA, Fodde R, Ranzani GN, et al. 1998. A National Cancer Institute Workshop on Microsatellite Instability for Cancer Detection and Familial Predisposition: Development of International Criteria for the Determination of Microsatellite Instability in Colorectal Cancer.

Bonneville R, Krook MA, Kautto EA, Miya J, Wing MR, Chen H-Z, Reeser JW, Yu L, Roychowdhury S. 2017. Landscape of Microsatellite Instability Across 39 Cancer Types. JCO Precis Oncol. 2017(1):1–15.

Bougen-Zhukov N, Nouri Y, Godwin T, Taylor M, Hakkaart C, Single A, Brew T, Permina E, Chen A, Black MA, et al. 2019. Allosteric AKT inhibitors target synthetic lethal vulnerabilities in E-cadherin-deficient cells. Cancers (Basel). 11(9).

Boycott KM, Rath A, Chong JX, Hartley T, Alkuraya FS, Baynam G, Brookes AJ, Brudno M, Carracedo A, den Dunnen JT, et al. 2017. International Cooperation to Enable the Diagnosis of All Rare Genetic Diseases. Am J Hum Genet. 100(5):695–705.

Braccini L, Ciraolo E, Campa CC, Perino A, Longo DL, Tibolla G, Pregnolato M, Cao Y, Tassone B, Damilano F, et al. 2015. PI3K-C2γ 3 is a Rab5 effector selectively controlling endosomal Akt2 activation downstream of insulin signalling. Nat Commun. 6(1):1–15.

Bragulla HH, Homberger DG. 2009. Structure and functions of keratin proteins in simple, stratified, keratinized and cornified epithelia. In: Journal of Anatomy. Vol. 214. Wiley-Blackwell. p. 516–559.

Broca P. 1866. Trait� des tumeurs. Paris: P. Asselin.

Broeks A, Urbanus JHM, Floore AN, Dahler EC, Klijn JGM, Rutgers EJT, Devilee P, Russell NS, Van Leeuwen FE, Van 't Veer LJ. 2000. ATM-heterozygous germline mutations contribute to breast cancer- susceptibility. Am J Hum Genet. 66(2):494–500.

Bronner CE, Baker SM, Morrison PT, Warren G, Smith LG, Lescoe MK, Kane M, Earabino C, Lipford J, Lindblom A, et al. 1994. Mutation in the DNA mismatch repair gene homologue hMLH 1 is associated with hereditary non-polyposis colon cancer. Nature. 368(6468):258–261.

Brooks-Wilson AR, Kaurah P, Suriano G, Leach S, Senz J, Grehan N, Butterfield YSN, Jeyes J, Schinas J, Bacani J, et al. 2004. Germline E-cadherin mutations in hereditary diffuse gastric cancer: Assessment of 42 new families and review of genetic screening criteria. J Med Genet. 41(7):508–517.

Calado RT, Regal JA, Hills M, Yewdell WT, Dalmazzo LF, Zago MA, Lansdorp PM, Hogge D, Chanock SJ, Estey EH, et al. 2009. Constitutional hypomorphic telomerase mutations in patients with acute myeloid leukemia. Proc Natl Acad Sci U S A. 106(4):1187–1192.

Caldas C, Carneiro F, Lynch HT, Yokota J, Wiesner GL, Powell SM, Lewis FR, Huntsman DG, Pharoah PDP, Jankowski JA, et al. 1999. Familial gastric cancer: overview and guidelines for management*. J Med Genet. 36:873–880.

Calva-Cerqueira D, Dahdaleh FS, Woodfield G, Chinnathambi S, Nagy PL, Larsen-Haidle J, Weigel RJ, Howe JR. 2010. Discovery of the BMPR1A promoter and germline mutations that cause juvenile polyposis. Hum Mol Genet. 19(23):4654–62.

Campbell CJ, Ghazal P. 2004. Molecular signatures for diagnosis of infection: Application of microarray technology. In: Journal of Applied Microbiology. Vol. 96. p. 18–23.

Carvalho CMB, Lupski JR. 2016. Mechanisms underlying structural variant formation in

genomic disorders. Nat Rev Genet. 17(4):224–238.

Chaisson MJP, Sanders AD, Zhao X, Malhotra A, Porubsky D, Rausch T, Gardner EJ, Rodriguez OL, Guo L, Collins RL, et al. 2019. Multi-platform discovery of haplotype-resolved structural variation in human genomes. Nat Commun. 10(1).

Chen J, Lau BT, Andor N, Grimes SM, Handy C, Wood-Bouwens C, Ji HP. 2019. Single-cell transcriptome analysis identifies distinct cell types and niche signaling in a primary gastric organoid model. Sci Rep. 9(1):4536.

Chen X, Schulz-Trieglaff O, Shaw R, Barnes B, Schlesinger F, Källberg M, Cox AJ, Kruglyak S, Saunders CT. 2016. Manta: Rapid detection of structural variants and indels for germline and cancer sequencing applications. Bioinformatics. 32(8):1220–1222.

Chen X, Weaver J, Bove BA, Vanderveer LA, Weil SC, Miron A, Daly MB, Godwin AK. 2008. Allelic imbalance in BRCA1 and BRCA2 gene expression is associated with an increased breast cancer risk. Hum Mol Genet. 17(9):1336–1348.

Cheng DT, Prasad M, Chekaluk Y, Benayed R, Sadowska J, Zehir A, Syed A, Wang YE, Somar J, Li Y, et al. 2017. Comprehensive detection of germline variants by MSK-IMPACT, a clinical diagnostic platform for solid tumor molecular oncology and concurrent cancer predisposition testing. BMC Med Genomics. 10(1):33.

Chimienti F, Devergnas S, Favier A, Seve M. 2004. Identification and cloning of a β-cell-specific zinc transporter, ZnT-8, localized into insulin secretory granules. Diabetes. 53(9):2330–2337.

Chiu R, Nip KM, Chu J, Birol I. 2018. TAP: a targeted clinical genomics pipeline for detecting transcript variants using RNA-seq data. BMC Med Genomics. 11(1):79.

Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM. 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly (Austin). 6(2):80–92.

Cookson W, Liang L, Abecasis G, Moffatt M, Lathrop M. 2009. Mapping complex disease traits with global gene expression. Nat Rev Genet. 10(3):184–194.

Corso G, Feroce I, Intra M, Toesca A, Magnoni F, Sargenti M, Naninato P, Caldarella P, Pagani G, Vento A, et al. 2018. BRCA1/2 germline missense mutations: A systematic review. Eur J Cancer Prev. 27(3):279–286.

Couch FJ, Hart SN, Sharma P, Toland AE, Wang X, Miron P, Olson JE, Godwin AK, Pankratz VS, Olswold C, et al. 2015. Inherited mutations in 17 breast cancer susceptibility genes among a large triple-negative breast cancer cohort unselected for family history of breast cancer. J Clin Oncol. 33(4):304–311.

Couch FJ, Shimelis H, Hu C, Hart SN, Polley EC, Na J, Hallberg E, Moore R, Thomas A, Lilyquist J, et al. 2017. Associations between cancer predisposition testing panel genes and breast cancer. JAMA Oncol. 3(9):1190–1196.

Couvelard A, Cauvin JM, Goldfain D, Rotenberg A, Robaszkiewicz M, Fléjou JF, Croué A, Volant A, Diebold MD, Vissuzaine C, et al. 2001. Cytokeratin immunoreactivity of intestinal metaplasia at normal oesophagogastric junction indicates its aetiology. Gut. 49(6):761–766.

Craig Venter J, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA, et al. 2001. The sequence of the human genome. Science (80-). 291(5507):1304–1351.

Cristescu R, Lee J, Nebozhyn M, Kim KM, Ting JC, Wong SS, Liu J, Yue YG, Wang J, Yu K, et al. 2015. Molecular analysis of gastric cancer identifies subtypes associated with distinct clinical outcomes. Nat Med. 21(5):449–456.

Cummings BB, Marshall JL, Tukiainen T, Lek M, Donkervoort S, Foley AR, Bolduc V, Waddell LB, Sandaradura SA, O 'grady GL, et al. 2017. Improving genetic diagnosis in Mendelian disease with transcriptome sequencing Genotype-Tissue Expression Consortium.

Cybulski C, Górski B, Huzarski T, Masojć B, Mierzejewski M, Dębniak T, Teodorczyk U, Byrski T, Gronwald J, Matyjasik J, et al. 2004. CHEK2 is a multiorgan cancer susceptibility gene. Am J Hum Genet. 75(6):1131–1135.

Daimon M, Sato H, Oizumi T, Toriyama S, Saito T, Karasawa S, Jimbu Y, Wada K, Kameda W, Susa S, et al. 2008. Association of the PIK3C2G gene polymorphisms with type 2 DM in a Japanese population. Biochem Biophys Res Commun. 365(3):466–471.

Daly MB, Pilarski R, Berry M, Buys SS, Farmer M, Friedman S, Garber JE, Kauff ND, Khan S, Klein C, et al. 2017. Genetic/familial high-risk assessment: Breast and ovarian, version 2.2017: Featured updates to the NCCN guidelines. JNCCN J Natl Compr Cancer Netw. 15(1):9–20.

Daniely Y, Liao G, Dixon D, Linnoila RI, Lori A, Randell SH, Oren M, Jetten AM. 2004. Critical role of p63 in the development of a normal esophageal and tracheobronchial epithelium. Am J Physiol - Cell Physiol. 287(1 56-1):C171-81.

Davies H, Glodzik D, Morganella S, Yates LR, Staaf J, Zou X, Ramakrishna M, Martin S, Boyault S, Sieuwerts AM, et al. 2017. HRDetect is a predictor of BRCA1 and BRCA2 deficiency based on mutational signatures. Nat Med. 23(4):517–525.

Dempfle A, Scherag A, Hein R, Beckmann L, Chang-Claude J, Schäfer H. 2008. Gene-

environment interactions for complex traits: Definitions, methodological requirements and challenges. Eur J Hum Genet. 16(10):1164–1172.

Dent R, Trudeau M, Pritchard KI, Hanna WM, Kahn HK, Sawka CA, Lickley LA, Rawlinson E, Sun P, Narod SA. 2007. Triple-negative breast cancer: Clinical features and patterns of recurrence. Clin Cancer Res. 13(15):4429–4434.

Desmond A, Kurian AW, Gabree M, Mills MA, Anderson MJ, Kobayashi Y, Horick N, Yang S, Shannon KM, Tung N, et al. 2015. Clinical actionability of multigene panel testing for hereditary breast and ovarian cancer risk assessment. JAMA Oncol. 1(7):943–951.

Diaz-Castro B, Gangwani MR, Yu X, Coppola G, Khakh BS. 2019. Astrocyte molecular signatures in Huntington's disease. Sci Transl Med. 11(514).

Ding J, Bashashati A, Roth A, Oloumi A, Tse K, Zeng T, Haffari G, Hirst M, Marra MA, Condon A, et al. 2012. Feature-based classifiers for somatic mutation detection in tumour-normal paired sequencing data. Bioinformatics. 28(2):167–175.

Dunham MA, Neumann AA, Fasching CL, Reddel RR. 2000. Telomere maintenance by recombination in human cells. Nat Genet. 26(4):447–450.

Dunnwald LK, Rossing MA, Li CI. 2007. Hormone receptor status, tumor characteristics, and prognosis: A prospective cohort of breast cancer patients. Breast Cancer Res. 9(1):1–10.

Dupont WD, Page DL. 1985. Risk Factors for Breast Cancer in Women with Proliferative Breast Disease. N Engl J Med. 312(3):146–151.

Duzkale H, Shen J, Mclaughlin H, Alfares A, Kelly M, Pugh T, Funke B, Rehm H, Lebo M. 2013. A systematic approach to assessing the clinical significance of genetic variants. Clin Genet. 84(5):453–463.

Easton DF, Pooley KA, Dunning AM, Pharoah PDP, Thompson D, Ballinger DG, Struewing JP, Morrison J, Field H, Luben R, et al. 2007. Genome-wide association study identifies novel breast cancer susceptibility loci. Nature. 447(7148):1087–1093.

Eichler EE, Flint J, Gibson G, Kong A, Leal SM, Moore JH, Nadeau JH. 2010. Missing heritability and strategies for finding the underlying causes of complex disease. Nat Rev Genet. 11(6):446–450.

Evans MJ, Van Winkle LS, Fanucchi M V., Plopper CG. 2001. Cellular and molecular characteristics of basal cells in airway epithelium. Exp Lung Res. 27(5):401–415.

Fedchenko N, Reifenrath J. 2014. Different approaches for interpretation and reporting of immunohistochemistry analysis results in the bone tissue - a review. Diagn Pathol. 9:221.

Feuk L, Carson AR, Scherer SW. 2006. Structural variation in the human genome. Nat Rev Genet. 7(2):85–97.

Finak G, McDavid A, Yajima M, Deng J, Gersuk V, Shalek AK, Slichter CK, Miller HW, McElrath MJ, Prlic M, et al. 2015. MAST: A flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. Genome Biol. 16(1).

Fishel R, Lescoe MK, Rao MRS, Copeland NG, Jenkins NA, Garber J, Kane M, Kolodner R. 1993. The human mutator gene homolog MSH2 and its association with hereditary nonpolyposis colon cancer. Cell. 75(5):1027–1038.

Fisher B, Costantino J, Redmond C, Poisson R, Bowman D, Couture J, Dimitrov N V., Wolmark N, Wickerham DL, Fisher ER, et al. 1989. A Randomized Clinical Trial Evaluating Tamoxifen in the Treatment of Patients with Node-Negative Breast Cancer Who Have Estrogen-Receptor–

Positive Tumors. N Engl J Med. 320(8):479–484.

Fitzgerald RC, Hardwick R, Huntsman D, Carneiro F, Guilford P, Blair V, Chung DC, Norton J, Ragunath K, Van Krieken JH, et al. 2010. Hereditary diffuse gastric cancer: updated consensus guidelines for clinical management and directions for future research. J Med Genet. 47(7):436–44.

Foretova L, Navratilova M, Svoboda M, Grell P, Nemec L, Sirotek L, Obermannova R, Novotny I, Sachlova M, Fabian P, et al. 2019. GAPPS – gastric adenocarcinoma and proximal polyposis of the stomach syndrome in 8 families tested at Masaryk memorial cancer institute – prevention and prophylactic gastrectomies. Klin Onkol. 32(Supplementum2):2S109-2S117.

Foulkes WD, Knoppers BM, Turnbull C. 2016. Population genetic testing for cancer susceptibility: founder mutations to genomes. Nat Rev Clin Oncol. 13(1):41–54.

Foulkes WD, Stefansson IM, Chappuis PO, Bégin LR, Goffin JR, Wong N, Trudel M, Akslen LA. 2003. Germline BRCA1 mutations and a basal epithelial phenotype in breast cancer. J Natl Cancer Inst. 95(19):1482–5.

Frayling IM, Beck NE, Ilyas M, Dove-Edwin I, Goodman P, Pack K, Bell JA, Williams CB, Hodgson S V., Thomas HJW, et al. 1998. The APC variants I1307K and E1317Q are associated with colorectal tumors, but not always with a family history. Proc Natl Acad Sci U S A. 95(18):10722–10727.

Friebel TM, Domchek SM, Rebbeck TR. 2014. Modifiers of cancer risk in BRCA1 and BRCA2 mutation carriers: systematic review and meta-analysis. J Natl Cancer Inst. 106(6):dju091.

Friend SH, Bernards R, Rogelj S, Weinberg RA, Rapaport JM, Albert DM, Dryja TP. 1986. A human DNA segment with properties of the gene that predisposes to retinoblastoma and

osteosarcoma. Nature. 323(6089):643–646.

Gabai-Kapara E, Lahad A, Kaufman B, Friedman E, Segev S, Renbaum P, Beeri R, Gal M, Grinshpun-Cohen J, Djemal K, et al. 2014. Population-based screening for breast and ovarian cancer risk due to BRCA1 and BRCA2. Proc Natl Acad Sci U S A. 111(39):14205–10.

Gallagher MD, Chen-Plotkin AS. 2018. The Post-GWAS Era: From Association to Function. Am J Hum Genet. 102(5):717–730.

Green RC, Berg JS, Grody WW, Kalia SS, Korf BR, Martin CL, McGuire AL, Nussbaum RL, O'Daniel JM, Ormond KE, et al. 2013. ACMG recommendations for reporting of incidental findings in clinical exome and genome sequencing. Genet Med. 15(7):565–574.

Greenberg MVC, Bourc'his D. 2019. The diverse roles of DNA methylation in mammalian development and disease. Nat Rev Mol Cell Biol. 20(10):590–607.

Gryfe R, Di Nicola N, Geeta L, Gallinger S, Redston M. 1999. Inherited colorectal polyposis and cancer risk of the APC I1307K polymorphism. Am J Hum Genet. 64(2):378–384.

Guilford P, Hopkins J, Harraway J, McLeod M, McLeod N, Harawira P, Taite H, Scoular R, Miller A, Reeve AE. 1998. E-cadherin germline mutations in familial gastric cancer. Nature. 392(6674):402–405.

Ha G, Roth A, Lai D, Bashashati A, Ding J, Goya R, Giuliany R, Rosner J, Oloumi A, Shumansky K, et al. 2012. Integrative analysis of genome-wide loss of heterozygosity and monoallelic expression at nucleotide resolution reveals disrupted pathways in triple-negative breast cancer. Genome Res. 22(10):1995–2007.

Haghverdi L, Lun ATL, Morgan MD, Marioni JC. 2018. Batch effects in single-cell RNA-sequencing data are corrected by matching mutual nearest neighbors. Nat Biotechnol. 36(5):421–

427.

Hall JM, Lee MK, Newman B, Morrow JE, Anderson LA, Huey B, King MC. 1990. Linkage of early-onset familial breast cancer to chromosome 17q21. Science (80-). 250(4988):1684–1689.

Hall MJ, Larson K, Bernhisel R, Hughes E, Rosenthal E, Singh N, Lancaster JM, Kurian AW. 2020. Abstract P5-03-02: Cancer risks associated with pathogenic variants in the ataxia telangiectasia mutated (*ATM*) gene. In: Poster Session Abstracts. Vol. 80. American Association for Cancer Research. p. P5-03-02-P5-03–02.

Hampel H, Frankel WL, Martin E, Arnold M, Khanduja K, Kuebler P, Clendenning M, Sotamaa K, Prior T, Westman JA, et al. 2008. Feasibility of Screening for Lynch Syndrome Among Patients With Colorectal Cancer. J Clin Oncol. 26(35):5783–5788.

Hampel H, Frankel WL, Martin E, Arnold M, Khanduja K, Kuebler P, Nakagawa H, Sotamaa K, Prior TW, Westman J, et al. 2005a. Screening for the Lynch Syndrome (Hereditary Nonpolyposis Colorectal Cancer). N Engl J Med. 352(18):1851–1860.

Hampel H, Frankel WL, Martin E, Arnold M, Khanduja K, Kuebler P, Nakagawa H, Sotamaa K, Prior TW, Westman J, et al. 2005b. Screening for the Lynch Syndrome (Hereditary Nonpolyposis Colorectal Cancer). N Engl J Med. 352(18):1851–1860.

Hampel H, Pearlman R, Beightol M, Zhao W, Jones D, Frankel WL, Goodfellow PJ, Yilmaz A, Miller K, Bacher J, et al. 2018. Assessment of tumor sequencing as a replacement for lynch syndrome screening and current molecular tests for patients with colorectal cancer. JAMA Oncol. 4(6):806–813.

Hanahan D, Weinberg RA. 2000. The hallmarks of cancer. Cell. 100(1):57–70.

Hanahan D, Weinberg RA. 2011. Hallmarks of cancer: The next generation. Cell. 144(5):646–

674.

Hansford S, Kaurah P, Li-Chang H, Woo M, Senz J, Pinheiro H, Schrader KA, Schaeffer DF, Shumansky K, Zogopoulos G, et al. 2015. Hereditary Diffuse Gastric Cancer Syndrome. JAMA Oncol. 1(1):23.

Haradhvala NJ, Kim J, Maruvka YE, Polak P, Rosebrock D, Livitz D, Hess JM, Leshchiner I, Kamburov A, Mouw KW, et al. 2018. Distinct mutational signatures characterize concurrent loss of polymerase proofreading and mismatch repair. Nat Commun. 9(1):1–9.

Hartge P, Struewing JP, Wacholder S, Brody LC, Tucker MA. 1999. The prevalence of common BRCA 1 and BRCA2 mutations among Ashkenazi Jews. Am J Hum Genet. 64(4):963–970.

Hartman AR, Kaldate RR, Sailer LM, Painter L, Grier CE, Endsley RR, Griffin M, Hamilton SA, Frye CA, Silberman MA, et al. 2012. Prevalence of BRCA mutations in an unselected population of triple-negative breast cancer. Cancer. 118(11):2787–2795.

Hebbar P, Elkum N, Alkayal F, John SE, Thanaraj TA, Alsmadi O. 2017. Genetic risk variants for metabolic traits in Arab populations. Sci Rep. 7.

Helgason H, Rafnar T, Olafsdottir HS, Jonasson JG, Sigurdsson A, Stacey SN, Jonasdottir A, Tryggvadottir L, Alexiusdottir K, Haraldsson A, et al. 2015. Loss-of-function variants in ATM confer risk of gastric cancer. Nat Genet. 47(8):906–910.

Helleday T, Eshtad S, Nik-Zainal S. 2014. Mechanisms underlying mutational signatures in human cancers. Nat Rev Genet. 15(9):585–598.

Van Hoeck A, Tjoonk NH, Van Boxtel R, Cuppen E. 2019. Portrait of a cancer: Mutational signature analyses for cancer diagnostics. BMC Cancer. 19(1).

Houlston RS, Webb E, Broderick P, Pittman AM, Di Bernardo MC, Lubbe S, Chandler I,

Vijayakrishnan J, Sullivan K, Penegar S, et al. 2008. Meta-analysis of genome-wide association data identifies four new susceptibility loci for colorectal cancer. Nat Genet. 40(12):1426–1435.

Huang K-L, Mashl RJ, Plon SE, Chen F, Ding L. 2018. Pathogenic Germline Variants in 10,389 Adult Cancers. Cell. 173:355–370.

Huang K, Mashl RJ, Wu Y, Ritter DI, Wang J, Oh C, Paczkowska M, Reynolds S, Wyczalkowski MA, Oak N, et al. 2018. Pathogenic Germline Variants in 10,389 Adult Cancers. Cell. 173(2):355-370.e14.

Humar B, Blair V, Charlton A, More H, Martin I, Guilford P. 2009. E-cadherin deficiency initiates gastric signet-ring cell carcinoma in mice and man. Cancer Res. 69(5):2050–2056.

Hutton JC, O'Brien RM. 2009. Glucose-6-phosphatase catalytic subunit gene family. J Biol Chem. 284(43):29241–29245.

Imperiale TF, Wagner DR, Lin CY, Larkin GN, Rogge JD, Ransohoff DF. 2002. Results of screening colonoscopy among persons 40 to 49 years of age. N Engl J Med. 346(23):1781–1785.

Ionov Y, Peinado MA, Malkhosyan S, Shibata D, Perucho M. 1993. Ubiquitous somatic mutations in simple repeated sequences reveal a new mechanism for colonic carcinogenesis. Nature. 363(6429):558–561.

Jaeger E, Leedham S, Lewis A, Segditsas S, Becker M, Cuadrado PR, Davis H, Kaur K, Heinimann K, Howarth K, et al. 2012. Hereditary mixed polyposis syndrome is caused by a 40-kb upstream duplication that leads to increased and ectopic expression of the BMP antagonist GREM1. Nat Genet. 44(6):699–703.

Jager M, Blokzijl F, Kuijk E, Bertl J, Vougioukalaki M, Janssen R, Besselink N, Boymans S, De Ligt J, Pedersen JS, et al. 2019. Deficiency of nucleotide excision repair is associated with

mutational signature observed in cancer. Genome Res. 29(7):1067–1077.

Jain M, Koren S, Miga KH, Quick J, Rand AC, Sasani TA, Tyson JR, Beggs AD, Dilthey AT, Fiddes IT, et al. 2018. Nanopore sequencing and assembly of a human genome with ultra-long reads. Nat Biotechnol. 36(4):338–345.

Jasperson KW, Tuohy TM, Neklason DW, Burt RW. 2010. Hereditary and Familial Colon Cancer. Gastroenterology. 138(6):2044–2058.

Jiang M, Li H, Zhang Y, Yang Y, Lu R, Liu K, Lin S, Lan X, Wang H, Wu H, et al. 2017. Transitional basal cells at the squamous-columnar junction generate Barrett's oesophagus. Nature. 550(7677):529–533.

Jones SJ, Laskin J, Li YY, Griffith OL, An J, Bilenky M, Butterfield YS, Cezard T, Chuah E, Corbett R, et al. 2010. Evolution of an adenocarcinoma in response to selection by targeted kinase inhibitors. Genome Biol. 11(8).

Jovanovic I, Tzardi M, Mouzas IA, Micev M, Pesko P, Milosavljevic T, Zois M, Sganzos M, Delides G, Kanavaros P. 2002. Changing pattern of cytokeratin 7 and 20 expression from normal epithelium to intestinal metaplasia of the gastric mucosa and gastroesophageal junction. Histol Histopathol. 17(2):445–454.

Jovov B, Van Itallie CM, Shaheen NJ, Carson JL, Gambling TM, Anderson JM, Orlando RC. 2007. Claudin-18: a dominant tight junction protein in Barrett's esophagus and likely contributor to its acid resistance. Am J Physiol Liver Physiol. 293(6):G1106–G1113.

van der Kaaij RT, van Kessel JP, van Dieren JM, Snaebjornsson P, Balagué O, van Coevorden F, van der Kolk LE, Sikorska K, Cats A, van Sandick JW. 2018. Outcomes after prophylactic gastrectomy for hereditary diffuse gastric cancer. Br J Surg. 105(2):e176–e182.

Kalia SS, Adelman K, Bale SJ, Chung WK, Eng C, Evans JP, Herman GE, Hufnagel SB, Klein TE, Korf BR, et al. 2017. Recommendations for reporting of secondary findings in clinical exome and genome sequencing, 2016 update (ACMG SF v2.0): a policy statement of the American College of Medical Genetics and Genomics. Genet Med. 19(2):249–255.

Kamps R, Brandão RD, Bosch BJ van den, Paulussen ADC, Xanthoulea S, Blok MJ, Romano A. 2017. Next-Generation Sequencing in Oncology: Genetic Diagnosis, Risk Prediction and Cancer Classification. Int J Mol Sci. 18(2).

Karolchik D. 2004. The UCSC Table Browser data retrieval tool. Nucleic Acids Res. 32(90001):493D – 496.

Kast K, Rhiem K, Wappenschmidt B, Hahnen E, Hauke J, Bluemcke B, Zarghooni V, Herold N, Ditsch N, Kiechle M, et al. 2016. Prevalence of BRCA1/2 germline mutations in 21 401 families with breast and ovarian cancer. J Med Genet. 53(7):465–71.

Kauff ND, Perez-Segura P, Robson ME, Scheuer L, Siegel B, Schluger A, Rapaport B, Frank TS, Nafa K, Ellis NA, et al. 2002. Incidence of non-founder BRCA1 and BRCA2 mutations in high risk Ashkenazi breast and ovarian cancer families. J Med Genet. 39(8):611–614.

Kaufman B, Shapira-Frommer R, Schmutzler RK, Audeh MW, Friedlander M, Balmaña J, Mitchell G, Fried G, Stemmer SM, Hubert A, et al. 2015. Olaparib monotherapy in patients with advanced cancer and a germline BRCA1/2 mutation. J Clin Oncol. 33(3):244–250.

King M-C, Levy-Lahad E, Lahad A. 2014. Population-Based Screening for *BRCA1* and *BRCA2*. JAMA. 312(11):1091.

King M-C, Marks JH, Mandell JB, New York Breast Cancer Study Group. 2003. Breast and ovarian cancer risks due to inherited mutations in BRCA1 and BRCA2. Science. 302(5645):643–

6.

Knudson AG. 1971. Mutation and Cancer: Statistical Study of Retinoblastoma.

Kohlmann W, Gruber SB. 2004 Apr 12. Lynch Syndrome. 2004 Feb 5 [Updated 2018 Apr 12].
In: Adam MP, Ardinger HH, Pagon RA, et al., editors. GeneReviews® [Internet]. Seattle (WA):
University of Washington, Seattle; 1993-2019. GeneReviews(®).

Kremer LS, Bader DM, Mertes C, Kopajtich R, Pichler G, Iuso A, Haack TB, Graf E,
Schwarzmayr T, Terrile C, et al. 2017. ARTICLE Genetic diagnosis of Mendelian disorders via
RNA sequencing.

Kurian AW, Hare EE, Mills MA, Kingham KE, McPherson L, Whittemore AS, McGuire V,
Ladabaum U, Kobayashi Y, Lincoln SE, et al. 2014. Clinical evaluation of a multiple-gene
sequencing panel for hereditary cancer risk assessment. J Clin Oncol. 32(19):2001–2009.

LaDuca H, Stuenkel AJ, Dolinsky JS, Keiles S, Tandy S, Pesaran T, Chen E, Gau C-L, Palmaer
E, Shoaepour K, et al. 2014. Utilization of multigene panels in hereditary cancer predisposition
testing: analysis of more than 2,000 patients. Genet Med. 16(11):830–837.

Lage K, Hansena NT, Karlberg EO, Eklund AC, Roque FS, Donahoe PK, Szallasi Z, Jensen TS,
Brunak S. 2008. A large-scale analysis of tissue-specific pathology and gene expression of
human disease genes and complexes. Proc Natl Acad Sci U S A. 105(52):20870–20875.

Laken SJ, Petersen GM, Gruber SB, Oddoux C, Ostrer H, Giardiello FM, Hamilton SR, Hampel
H, Markowitz A, Klimstra D, et al. 1997. Familial colorectal cancer in Ashkenazim due to a
hypermutable tract in APC. Nat Genet. 17(1):79–83.

Lamlum H, Tassan N Al, Jaeger E, Frayling I, Sieber O, Bin Reza F, Eckert M, Rowan A,
Barclay E, Atkin W, et al. 2000. Germline APC variants in patients with multiple colorectal

adenomas, with evidence for the particular importance of E1317Q. Hum Mol Genet. 9(15):2215–2221.

Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, Fitzhugh W, et al. 2001. Initial sequencing and analysis of the human genome. Nature. 409(6822):860–921.

Landrum MJ, Lee JM, Benson M, Brown GR, Chao C, Chitipiralla S, Gu B, Hart J, Hoffman D, Jang W, et al. 2018. ClinVar: improving access to variant interpretations and supporting evidence. Nucleic Acids Res. 46(D1):D1062–D1067.

Lang H, Piso P, Stukenborg C, Raab R, Jähne J. 2000. Management and results of proximal anastomotic leaks in a series of 1114 total gastrectomies for gastric carcinoma. Eur J Surg Oncol. 26(2):168–171.

Latham A, Srinivasan P, Kemel Y, Shia J, Bandlamudi C, Mandelker D, Middha S, Hechtman J, Zehir A, Dubard-Gault M, et al. 2019. Microsatellite instability is associated with the presence of Lynch syndrome pan-cancer. J Clin Oncol. 37(4):286–295.

Le DT, Durham JN, Smith KN, Wang H, Bartlett BR, Aulakh LK, Lu S, Kemberling H, Wilt C, Luber BS, et al. 2017. Mismatch repair deficiency predicts response of solid tumors to PD-1 blockade. Science (80-). 357(6349):409–413.

Le DT, Uram JN, Wang H, Bartlett BR, Kemberling H, Eyring AD, Skora AD, Luber BS, Azad NS, Laheru D, et al. 2015. PD-1 Blockade in Tumors with Mismatch-Repair Deficiency. N Engl J Med. 372(26):2509–2520.

Leclerc J, Flament C, Lovecchio T, Delattre L, Ait Yahya E, Baert-Desurmont S, Burnichon N, Bronner M, Cabaret O, Lejeune S, et al. 2018 Apr 12. Diversity of genetic events associated with

MLH1 promoter methylation in Lynch syndrome families with heritable constitutional epimutation. Genet Med.

Lee EYHP, Muller WJ. 2010. Oncogenes and tumor suppressor genes. Cold Spring Harb Perspect Biol. 2(10).

Lee K, Krempely K, Roberts ME, Anderson MJ, Carneiro F, Chao E, Dixon K, Figueiredo J, Ghosh R, Huntsman D, et al. 2018. Specifications of the ACMG/AMP variant curation guidelines for the analysis of germline CDH1 sequence variants. Hum Mutat. 39(11):1553–1568.

Lennerz JKM, Kim SH, Oates EL, Huh WJ, Doherty JM, Tian X, Bredemeyer AJ, Goldenring JR, Lauwers GY, Shin YK, et al. 2010. The transcription factor MIST1 is a novel human gastric chief cell marker whose expression is lost in metaplasia, dysplasia, and carcinoma. Am J Pathol. 177(3):1514–1533.

Lenoir GM, Lynch H, Watson P, Conway T, Lynch J, Narod S, Feunteun J. 1991. Familial breast-ovarian cancer locus on chromosome 17q12-q23. Lancet. 338(8759):82–83.

Leroy B, Ballinger ML, Baran-Marszak F, Bond GL, Braithwaite A, Concin N, Donehower LA, El-Deiry WS, Fenaux P, Gaidano G, et al. 2017. Recommended guidelines for validation, quality control, and reporting of TP53 variants in clinical practice. Cancer Res. 77(6):1250–1260.

Levy-Lahad E, Catane R, Eisenberg S, Kaufman B, Hornreich G, Lishinsky E, Shohat M, Weber BL, Beller U, Lahad A, et al. 1997. Founder BRCA1 and BRCA2 mutations in Ashkenazi Jews in Israel: Frequency and differential penetrance in ovarian cancer and in breast- ovarian cancer families. Am J Hum Genet. 60(5):1059–1067.

Levy-Lahad E, Lahad A, Eisenberg S, Dagan E, Paperna T, Kasinetz L, Catane R, Kaufman B, Beller U, Renbaum P, et al. 2001. A single nucleotide polymorphism in the RAD51 gene

modifies cancer risk in BRCA2 but not BRCA1 carriers. Proc Natl Acad Sci U S A. 98(6):3232–3236.

Li H. 2018. Minimap2: Pairwise alignment for nucleotide sequences. Bioinformatics. 34(18):3094–3100.

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 25(14):1754–1760.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 25(16):2078–2079.

Li J, Woods SL, Healey S, Beesley J, Chen X, Lee JS, Sivakumaran H, Wayte N, Nones K, Waterfall JJ, et al. 2016. Point Mutations in Exon 1B of APC Reveal Gastric Adenocarcinoma and Proximal Polyposis of the Stomach as a Familial Adenomatous Polyposis Variant. Am J Hum Genet. 98(5):830–842.

Liang J, Lin C, Hu F, Wang F, Zhu L, Yao X, Wang Y, Zhao Y. 2013. APC polymorphisms and the risk of colorectal neoplasia: a HuGE review and meta-analysis. Am J Epidemiol. 177(11):1169–79.

Lieberman DA, Holub JL, Moravec MD, Eisen GM, Peters D, Morris CD. 2008. Prevalence of colon polyps detected by colonoscopy screening in asymptomatic black and white patients. JAMA - J Am Med Assoc. 300(12):1417–1422.

Lieberman S, Lahad A, Tomer A, Cohen C, Levy-Lahad E, Raz A. 2017. Population screening for BRCA1/BRCA2 mutations: lessons from qualitative analysis of the screening experience. Genet Med. 19(6):628–634.

Liedtke C, Mazouni C, Hess K, Andre F. 2008. Response to Neoadjuvant Therapy and Long-

Term Survival in Patients With Triple-Negative Breast Cancer Definition of treatment strategies for Breast Cancer View project Methodology Papers View project. Artic J Clin Oncol.

Lincoln SE, Kobayashi Y, Anderson MJ, Yang S, Desmond AJ, Mills MA, Nilsen GB, Jacobs KB, Monzon FA, Kurian AW, et al. 2015. A systematic comparison of traditional and multigene panel testing for hereditary breast and ovarian cancer genes in more than 1000 patients. J Mol Diagnostics. 17(5):533–544.

Lindblom A, Tannergård P, Werelius B, Nordenskjöld M. 1993. Genetic mapping of a second locus predisposing to hereditary non–polyposis colon cancer. Nat Genet. 5(3):279–282.

Litim N, Labrie Y, Desjardins S, Ouellette G, Plourde K, Belleau P, Durocher F. 2013. Polymorphic variations in the FANCA gene in high-risk non-BRCA1/2 breast cancer individuals from the French Canadian population. Mol Oncol. 7(1):85–100.

Loboda A, Nebozhyn M V., Watters JW, Buser CA, Shaw PM, Huang PS, Van'T Veer L, Tollenaar RA, Jackson DB, Agrawal D, et al. 2011. EMT is the dominant program in human colon cancer. BMC Med Genomics. 4.

Lu C, Xie M, Wendl MC, Wang J, McLellan MD, Leiserson MDM, Huang K, Wyczalkowski MA, Jayasinghe R, Banerjee T, et al. 2015. Patterns and functional implications of rare germline variants across 12 cancer types. Nat Commun. 6(1):10086.

Lun ATL, McCarthy DJ, Marioni JC. 2016. A step-by-step workflow for low-level analysis of single-cell RNA-seq data with Bioconductor. F1000Research. 5(4):2122.

Lykke-Andersen S, Heick Jensen T. 2015. Nonsense-mediated mRNA decay: an intricate machinery that shapes transcriptomes.

Lynch HT, Krush AJ. 1967. Heredity and adenocarcinoma of the colon. Gastroenterology.

53(4):517–27.

Lynch HT, Krush AJ. 1971. Cancer family "G" revisited: 1895-1970. Cancer. 27(6):1505–1511.

Lynch HT, Lynch PM, Lanspa SJ, Snyder CL, Lynch JF, Boland CR. 2009. Review of the Lynch syndrome: History, molecular genetics, screening, differential diagnosis, and medicolegal ramifications. Clin Genet. 76(1):1–18.

Lynch HT, Shaw MW, Magnuson CW, Larsen AL, Krush AJ. 1966. Hereditary Factors in Cancer: Study of Two Large Midwestern Kindreds. Arch Intern Med. 117(2):206–212.

Mai PL, Best AF, Peters JA, DeCastro RM, Khincha PP, Loud JT, Bremer RC, Rosenberg PS, Savage SA. 2016. Risks of first and subsequent cancers among TP53 mutation carriers in the National Cancer Institute Li-Fraumeni syndrome cohort. Cancer. 122(23):3673–3681.

Manchanda R, Legood R, Burnell M, McGuire A, Raikou M, Loggenberg K, Wardle J, Sanderson S, Gessler S, Side L, et al. 2015. Cost-effectiveness of Population Screening for BRCA Mutations in Ashkenazi Jewish Women Compared With Family History–Based Testing. JNCI J Natl Cancer Inst. 107(1).

Manchanda R, Loggenberg K, Sanderson S, Burnell M, Wardle J, Gessler S, Side L, Balogun N, Desai R, Kumar A, et al. 2015. Population Testing for Cancer Predisposing BRCA1/BRCA2 Mutations in the Ashkenazi-Jewish Community: A Randomized Controlled Trial. JNCI J Natl Cancer Inst. 107(1).

Mandelker D, Kumar R, Pei X, Selenica P, Setton J, Arunachalam S, Ceyhan-Birsoy O, Brown DN, Norton L, Robson ME, et al. 2019. The Landscape of Somatic Genetic Alterations in Breast Cancers from CHEK2 Germline Mutation Carriers. JNCI Cancer Spectr. 3(2).

Mandelker D, Zhang L, Kemel Y, Stadler ZK, Joseph V, Zehir A, Pradhan N, Arnold A, Walsh

MF, Li Y, et al. 2017. Mutation Detection in Patients With Advanced Cancer by Universal Sequencing of Cancer-Related Genes in Tumor and Normal DNA vs Guideline-Based Germline Testing. JAMA. 318(9):825.

Marshall CJ. 1991. Tumor suppressor genes. Cell. 64(2):313–326.

Martínez ME, McPherson RS, Annegers JF, Levin B. 1995. Cigarette smoking and alcohol consumption as risk factors for colorectal adenomatous polyps. J Natl Cancer Inst. 87(4):274–279.

Martini M, De Santis MC, Braccini L, Gulluni F, Hirsch E. 2014. PI3K/AKT signaling pathway and cancer: An updated review. Ann Med. 46(6):372–383.

McCarthy DJ, Campbell KR, Lun ATL, Wills QF. 2017. Scater: Pre-processing, quality control, normalization and visualization of single-cell RNA-seq data in R. Bioinformatics. 33(8):1179–1186.

Meaburn EL, Schalkwyk LC, Mill J. 2010. Allele-specific methylation in the human genome: Implications for genetic studies of complex disease. Epigenetics. 5(7):578–582.

Meijers-Heijboer H, Van den Ouweland A, Klijn J, Wasielewski M, De Shoo A, Oldenburg R, Hollestelle A, Houben M, Crepin E, Van Veghel-Plandsoen M, et al. 2002. Low-penetrance susceptibility to breast cancer due to CHEK2*1100delC in noncarriers of BRCA1 or BRCA2 mutations: The CHEK2-breast cancer consortium. Nat Genet. 31(1):55–59.

Merker JD, Wenger AM, Sneddon T, Grove M, Zappala Z, Fresard L, Waggott D, Utiramerur S, Hou Y, Smith KS, et al. 2018. Long-read genome sequencing identifies causal structural variation in a Mendelian disease. Genet Med. 20(1):159–163.

Metcalfe KA, Poll A, Royer R, Llacuachaqui M, Tulman A, Sun P, Narod SA. 2010. Screening

for founder mutations in BRCA1 and BRCA2 in unselected Jewish women. J Clin Oncol. 28(3):387–91.

Metcalfe KA, Poll A, Royer R, Nanda S, Llacuachaqui M, Sun P, Narod SA. 2013. A comparison of the detection of BRCA mutation carriers through the provision of Jewish population-based genetic testing compared with clinic-based genetic testing. Br J Cancer. 109(3):777–779.

Milne RL, Antoniou AC. 2016. Modifiers of breast and ovarian cancer risks for BRCA1 and BRCA2 mutation carriers. Endocr Relat Cancer. 23(10):T69–T84.

Mitsui Y, Yokoyama R, Fujimoto S, Kagemoto K, Kitamura S, Okamoto K, Muguruma N, Bando Y, Eguchi H, Okazaki Y, et al. 2018. First report of an Asian family with gastric adenocarcinoma and proximal polyposis of the stomach (GAPPS) revealed with the germline mutation of the APC exon 1B promoter region. Gastric Cancer. 21(6):1058–1063.

Mohammed IA, Streutker CJ, Riddell RH. 2002. Utilization of cytokeratins 7 and 20 does not differentiate between Barrett's esophagus and gastric cardiac intestinal metaplasia. Mod Pathol. 15(6):611–616.

Morak M, Koehler U, Schackert HK, Steinke V, Royer-Pokora B, Schulmann K, Kloor M, Höchter W, Weingart J, Keiling C, et al. 2011. Biallelic MLH1 SNP cDNA expression or constitutional promoter methylation can hide genomic rearrangements causing Lynch syndrome. J Med Genet. 48(8):513–519.

Muraro MJ, Dharmadhikari G, Grün D, Groen N, Dielen T, Jansen E, van Gurp L, Engelse MA, Carlotti F, de Koning EJP, et al. 2016. A Single-Cell Transcriptome Atlas of the Human Pancreas. Cell Syst. 3(4):385-394.e3.

Murff HJ, Byrne D, Syngal S. 2004. Cancer risk assessment: Quality and impact of the family history interview. Am J Prev Med. 27(3):239–245.

Nagarsheth N, Wicha MS, Zou W. 2017. Chemokines in the cancer microenvironment and their relevance in cancer immunotherapy. Nat Rev Immunol. 17(9):559–572.

Nagasaka T, Rhees J, Kloor M, Gebert J, Naomoto Y, Boland CR, Goel A. 2010. Somatic hypermethylation of MSH2 is a frequent event in Lynch syndrome colorectal cancers. Cancer Res. 70(8):3098–3108.

Nagy R, Sweet K, Eng C. 2004. Highly penetrant hereditary cancer syndromes. Oncogene. 23(38):6445–6470.

Nakano T, Morishita S, Katafuchi A, Matsubara M, Horikawa Y, Terato H, Salem AMH, Izumi S, Pack SP, Makino K, et al. 2007. Nucleotide Excision Repair and Homologous Recombination Systems Commit Differentially to the Repair of DNA-Protein Crosslinks. Mol Cell. 28(1):147–158.

Needleman SB, Wunsch CD. 1970. A general method applicable to the search for similarities in the amino acid sequence of two proteins. J Mol Biol. 48(3):443–453.

Nevanlinna H, Bartek J. 2006. The CHEK2 gene and inherited breast cancer susceptibility. Oncogene. 25(43):5912–5919.

Niimi T, Nagashima K, Ward JM, Minoo P, Zimonjic DB, Popescu NC, Kimura S. 2001. claudin-18, a Novel Downstream Target Gene for the T/EBP/NKX2.1 Homeodomain Transcription Factor, Encodes Lung- and Stomach-Specific Isoforms through Alternative Splicing. Mol Cell Biol. 21(21):7380–7390.

Nik-Zainal S, Morganella S. 2017. Mutational signatures in breast cancer: The problem at the

DNA level. Clin Cancer Res. 23(11):2617–2629.

Nouri Y. 2019. The Establishment and Characterisation of Gastric Organoids as a Model for Hereditary Diffuse Gastric Cancer. University of Otago.

Oliveira C, Pinheiro H, Figueiredo J, Seruca R, Carneiro F. 2015. Familial gastric cancer: Genetic susceptibility, pathology, and implications for management. Lancet Oncol. 16(2):e60–e70.

Ott J, Wang J, Leal SM. 2015. Genetic linkage analysis in the age of whole-genome sequencing. Nat Rev Genet. 16(5):275–84.

Park SY, Kim HS, Hong EK, Kim WH. 2002. Expression of cytokeratins 7 and 20 in primary carcinomas of the stomach and colorectum and their value in the differential diagnosis of metastatic carcinomas to the ovary. Hum Pathol. 33(11):1078–1085.

Peltomäki P, Aaltonen LA, Sistonen P, Pylkkänen L, Mecklin JP, Järvinen H, Green JS, Jass JR, Weber JL, Leach FS, et al. 1993. Genetic mapping of a locus predisposing to human colorectal cancer. Science (80-). 260(5109):810–812.

Peri S, Caretti E, Tricarico R, Devarajan K, Cheung M, Sementino E, Menges CW, Nicolas E, Vanderveer LA, Howard S, et al. 2017. Haploinsufficiency in tumor predisposition syndromes: Altered genomic transcription in morphologically normal cells heterozygous for VHL or TSC mutation. Oncotarget. 8(11):17628–17642.

Pernot S, Voron T, Perkins G, Lagorce-Pages C, Berger A, Taieb J. 2015. Signet-ring cell carcinoma of the stomach: Impact on prognosis and specific therapeutic challenge. World J Gastroenterol. 21(40):11428–11438.

Petrucelli N, Daly MB, Pal T. 1993. BRCA1- and BRCA2-Associated Hereditary Breast and

Ovarian Cancer. University of Washington, Seattle.

Pharoah PDP, Day NE, Duffy S, Easton DF, Ponder BAJ. 1997. Family history and the risk of breast cancer: A systematic review and meta-analysis. Int J Cancer. 71(5):800–809.

Phillips SM, Banerjea A, Feakins R, Li SR, Bustin SA, Dorudi S. 2004. Tumour-infiltrating lymphocytes in colorectal cancer with microsatellite instability are activated and cytotoxic. Br J Surg. 91(4):469–475.

Pilati C, Shinde J, Alexandrov LB, Assié G, André T, Hélias-Rodzewicz Z, Ducoudray R, Le Corre D, Zucman-Rossi J, Emile J-F, et al. 2017. Mutational signature analysis identifies *MUTYH* deficiency in colorectal cancers and adrenocortical carcinomas. J Pathol. 242(1):10–15.

Pinto D, Pinto C, Guerra J, Pinheiro M, Santos R, Vedeld HM, Yohannes Z, Peixoto A, Santos C, Pinto P, et al. 2018. Contribution of *MLH1* constitutional methylation for Lynch syndrome diagnosis in patients with tumor MLH1 downregulation. Cancer Med. 7(2):433–444.

Pleasance E, Titmuss E, Williamson L, Kwan H, Culibrk L, Zhao EY, Dixon K, Fan K, Bowlby R, Jones MR, et al. 2020. Pan-cancer analysis of advanced patient tumors reveals interactions between therapy and genomic landscapes. Nat Cancer. 1(4):452–468.

Plon SE, Eccles DM, Easton D, Foulkes WD, Genuardi M, Greenblatt MS, Hogervorst FBL, Hoogerbrugge N, Spurdle AB, Tavtigian S V. 2008. Sequence variant classification and reporting: recommendations for improving the interpretation of cancer susceptibility genetic test results. Hum Mutat. 29(11):1282–1291.

Polak P, Kim J, Braunstein LZ, Karlic R, Haradhavala NJ, Tiao G, Rosebrock D, Livitz D, Kübler K, Mouw KW, et al. 2017. A mutational signature reveals alterations underlying deficient homologous recombination repair in breast cancer. Nat Genet. 49(10):1476–1486.

Portela-Gomes GM, Lukinius A, Ljungberg O, Efendic S, Ahrén B, Abdel-Halim SM. 2003. PACAP is expressed in secretory granules of insulin and glucagon cells in human and rodent pancreas. Evidence for generation of cAMP compartments uncoupled from hormone release in diabetic islets. Regul Pept. 113(1–3):31–9.

van der Post RS, Vogelaar IP, Carneiro F, Guilford P, Huntsman D, Hoogerbrugge N, Caldas C, Schreiber KEC, Hardwick RH, Ausems MGEM, et al. 2015. Hereditary diffuse gastric cancer: updated clinical guidelines with an emphasis on germline CDH1 mutation carriers. J Med Genet. 52(6):361–74.

Poynter JN, Siegmund KD, Weisenberger DJ, Long TI, Thibodeau SN, Lindor N, Young J, Jenkins MA, Hopper JL, Baron JA, et al. 2008. Molecular characterization of MSI-H colorectal cancer by MLHI promoter methylation, immunohistochemistry, and mismatch repair germline mutation screening. Cancer Epidemiol Biomarkers Prev. 17(11):3208–3215.

Rahman N. 2014. Realizing the promise of cancer predisposition genes. Nature. 505(7483):302–308.

Ramaekers F, Huysmans A, Schaart G, Moesker O, Vooijs P. 1987. Tissue distribution of keratin 7 as monitored by a monoclonal antibody. Exp Cell Res. 170(1):235–249.

Rashid M, Fischer A, Wilson CH, Tiffen J, Rust AG, Stevens P, Idziaszczyk S, Maynard J, Williams GT, Mustonen V, et al. 2016. Adenoma development in familial adenomatous polyposis and MUTYH -associated polyposis: Somatic landscape and driver genes. J Pathol. 238(1):98–108.

Rausch T, Zichner T, Schlattl A, Stütz AM, Benes V, Korbel JO. 2012. DELLY: Structural variant discovery by integrated paired-end and split-read analysis. Bioinformatics. 28(18).

Reisle C, Mungall KL, Choo C, Paulino D, Bleile DW, Muhammadzadeh A, Mungall AJ, Moore RA, Shlafman I, Coope R, et al. 2019. MAVIS: Merging, annotation, validation, and illustration of structural variants. Bioinformatics. 35(3):515–517.

Remmele W, Stegner HE. 1987. [Recommendation for uniform definition of an immunoreactive score (IRS) for immunohistochemical estrogen receptor detection (ER-ICA) in breast cancer tissue]. Pathologe. 8(3):138–40.

Renwick A, Thompson D, Seal S, Kelly P, Chagtai T, Ahmed M, North B, Jayatilake H, Barfoot R, Spanova K, et al. 2006. ATM mutations that cause ataxia-telangiectasia are breast cancer susceptibility alleles. Nat Genet. 38(8):873–875.

Repak R, Kohoutova D, Podhola M, Rejchrt S, Minarik M, Benesova L, Lesko M, Bures J. 2016. The first European family with gastric adenocarcinoma and proximal polyposis of the stomach: case report and review of the literature. Gastrointest Endosc. 84(4):718–725.

Rhees J, Arnold M, Boland CR. 2014. Inversion of exons 1–7 of the MSH2 gene is a frequent cause of unexplained Lynch syndrome in one local population. Fam Cancer. 13(2):219–225.

Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, Grody WW, Hegde M, Lyon E, Spector E, et al. 2015. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med. 17(5):405–423.

Roberts NJ, Jiao Y, Yu J, Kopelovich L, Petersen GM, Bondy ML, Gallinger S, Schwartz AG, Syngal S, Cote ML, et al. 2012. ATM mutations in patients with hereditary pancreatic cancer. Cancer Discov. 2(1):41–46.

Roberts NJ, Norris AL, Petersen GM, Bondy ML, Brand R, Gallinger S, Kurtz RC, Olson SH,

Rustgi AK, Schwartz AG, et al. 2016. Whole Genome Sequencing Defines the Genetic Heterogeneity of Familial Pancreatic Cancer. Cancer Discov. 6(2):166–75.

Robertson G, Schein J, Chiu R, Corbett R, Field M, Jackman SD, Mungall K, Lee S, Okada HM, Qian JQ, et al. 2010. De novo assembly and analysis of RNA-seq data. Nat Methods. 7(11):909–912.

Robinson DR, Wu YM, Lonigro RJ, Vats P, Cobain E, Everett J, Cao X, Rabban E, Kumar-Sinha C, Raymond V, et al. 2017. Integrative clinical genomics of metastatic cancer. Nature. 548(7667):297–303.

Rohlin A, Engwall Y, Fritzell K, Göransson K, Bergsten A, Einbeigi Z, Nilbert M, Karlsson P, Björk J, Nordling M. 2011. Inactivation of promoter 1B of APC causes partial gene silencing: evidence for a significant role of the promoter in regulation and causative of familial adenomatous polyposis. Oncogene. 30(50):4977–4989.

Rosenbluth JM, Schackmann RCJ, Gray GK, Selfors LM, Li CMC, Boedicker M, Kuiken HJ, Richardson A, Brock J, Garber J, et al. 2020. Organoid cultures from normal and cancer-prone human breast tissues preserve complex epithelial lineages. Nat Commun. 11(1):1–14.

Saeed M. 2018. Locus and gene-based GWAS meta-analysis identifies new diabetic nephropathy genes. Immunogenetics. 70(6):347–353.

Saeki K, Zhu M, Kubosaki A, Xie J, Lan MS, Notkins AL. 2002. Targeted disruption of the protein tyrosine phosphatase-like molecule IA-2 results in alterations in glucose tolerance tests and insulin secretion. Diabetes. 51(6):1842–1850.

Sakamoto H, Yoshimura K, Saeki N, Katai H, Shimoda T, Matsuno Y, Saito D, Sugimura H, Tanioka F, Kato S, et al. 2008. Genetic variation in PSCA is associated with susceptibility to

diffuse-type gastric cancer. Nat Genet. 40(6):730–740.

Salvador MU, Truelson MRF, Mason C, Souders B, LaDuca H, Dougall B, Black MH, Fulk K, Profato J, Gutierrez S, et al. 2019. Comprehensive paired tumor/germline testing for Lynch syndrome: Bringing resolution to the diagnostic process. In: Journal of Clinical Oncology. Vol. 37. American Society of Clinical Oncology. p. 647–657.

Sanchis-Juan A, Stephens J, French CE, Gleadall N, Mégy K, Penkett C, Shamardina O, Stirrups K, Delon I, Dewhurst E, et al. 2018. Complex structural variants in Mendelian disorders: identification and breakpoint resolution using short- and long-read genome sequencing. Genome Med. 10(1).

Saunders CT, Wong WSW, Swamy S, Becq J, Murray LJ, Cheetham RK. 2012. Strelka: Accurate somatic small-variant calling from sequenced tumor-normal sample pairs. Bioinformatics. 28(14):1811–1817.

Savage SA, Alter BP. 2009. Dyskeratosis Congenita. Hematol Oncol Clin North Am. 23(2):215–231.

Schmidt MK, Hogervorst F, Van Hien R, Cornelissen S, Broeks A, Adank MA, Meijers H, Waisfisz Q, Hollestelle A, Schutte M, et al. 2016. Age-And tumor subtype-specific breast cancer risk estimates for CHEK2∗1100delC Carriers. J Clin Oncol. 34(23):2750–2760.

Schneider G, Schmidt-Supprian M, Rad R, Saur D. 2017. Tissue-specific tumorigenesis: Context matters. Nat Rev Cancer. 17(4):239–253.

Schrader KA, Cheng DT, Joseph V, Prasad M, Walsh M, Zehir A, Ni A, Thomas T, Benayed R, Ashraf A, et al. 2016. Germline Variants in Targeted Tumor Sequencing Using Matched Normal DNA. JAMA Oncol. 2(1):104–11.

Schroeder C, Faust U, Sturm M, Hackmann K, Grundmann K, Harmuth F, Bosse K, Kehrer M, Benkert T, Klink B, et al. 2015. HBOC multi-gene panel testing: comparison of two sequencing centers. Breast Cancer Res Treat. 152(1):129–136.

Schwarze K, Buchanan J, Taylor JC, Wordsworth S. 2018. Are whole-exome and whole-genome sequencing approaches cost-effective? A systematic review of the literature. Genet Med. 20(10):1122–1130.

Scully R, Chen J, Plug A, Xiao Y, Weaver D, Feunteun J, Ashley T, Livingston DM. 1997. Association of BRCA1 with Rad51 in mitotic and meiotic cells. Cell. 88(2):265–275.

Seal S, Barfoot R, Jayatilake H, Smith P, Renwick A, Bascombe L, McGuffog L, Evans DG, Eccles D, Easton DF, et al. 2003. Evaluation of Fanconi Anemia genes in familial breast cancer predisposition. Cancer Res. 63(24):8596–9.

Sedlazeck FJ, Rescheneder P, Smolka M, Fang H, Nattestad M, Von Haeseler A, Schatz MC. 2018. Accurate detection of complex structural variations using single-molecule sequencing. Nat Methods. 15(6):461–468.

Segerstolpe Å, Palasantza A, Eliasson P, Andersson EM, Andréasson AC, Sun X, Picelli S, Sabirsh A, Clausen M, Bjursell MK, et al. 2016. Single-Cell Transcriptome Profiling of Human Pancreatic Islets in Health and Type 2 Diabetes. Cell Metab. 24(4):593–607.

Shanik MH, Xu Y, Skrha J, Dankner R, Zick Y, Roth J. 2008. Insulin resistance and hyperinsulinemia: is hyperinsulinemia the cart or the horse? Diabetes Care. 31 Suppl 2.

Sharan SK, Morimatsu M, Albrecht U, Lim DS, Regel E, Dinh C, Sands A, Eichele G, Hasty P, Bradley A. 1997. Embryonic lethality and radiation hypersensitivity mediated by Rad51 in mice lacking Brca2. Nature. 386(6627):804–810.

Shen B, Ormsby AH, Shen C, Dumot JA, Shao YW, Bevins CL, Gramlich TL. 2002. Cytokeratin expression patterns in noncardia, intestinal metaplasia-associated gastric adenocarcinoma: Implication for the evaluation of intestinal metaplasia and tumors at the esophagogastric junction. Cancer. 94(3):820–831.

Shimelis H, Mesman RLS, Von Nicolai C, Ehlen A, Guidugli L, Martin C, Calléja FMGR, Meeks H, Hallberg E, Hinton J, et al. 2017. BRCA2 hypomorphic missense variants confer moderate risks of breast cancer. Cancer Res. 77(11):2789–2799.

Shin K-H, Shin J-H, Kim J-H, Park J-G. 2002. Mutational analysis of promoters of mismatch repair genes hMSH2 and hMLH1 in hereditary nonpolyposis colorectal cancer and early onset colorectal cancer patients: identification of three novel germ-line mutations in promoter of the hMSH2 gene. Cancer Res. 62(1):38–42.

Shinde J, Bayard Q, Imbeaud S, Hirsch TZ, Liu F, Renault V, Zucman-Rossi J, Letouzé E. 2018. Palimpsest: an R package for studying mutational and structural variant signatures along clonal evolution in cancer. Bioinformatics. 34(19):3380–3381.

Shirts BH, Konnick EQ, Upham S, Walsh T, Ranola JMO, Jacobson AL, King M-C, Pearlman R, Hampel H, Pritchard CC. 2018. Using Somatic Mutations from Tumors to Classify Variants in Mismatch Repair Genes. Am J Hum Genet. 103(1):19–29.

Shlien A, Campbell BB, De Borja R, Alexandrov LB, Merico D, Wedge D, Van Loo P, Tarpey PS, Coupland P, Behjati S, et al. 2015. Combined hereditary and somatic mutations of replication error repair genes result in rapid onset of ultra-hypermutated cancers. Nat Genet. 47(3):257–262.

Shrubsole MJ, Wu H, Ness RM, Shyr Y, Smalley WE, Zheng W. 2008. Alcohol Drinking,

Cigarette Smoking, and Risk of Colorectal Adenomatous and Hyperplastic Polyps. Am J Epidemiol. 167(9):1050–1058.

De Silva IU, McHugh PJ, Clingen PH, Hartley JA. 2000. Defining the Roles of Nucleotide Excision Repair and Recombination in the Repair of DNA Interstrand Cross-Links in Mammalian Cells. Mol Cell Biol. 20(21):7980–7990.

Simpson JT, Wong K, Jackman SD, Schein JE, Jones SJM, Birol I. 2009. ABySS: A parallel assembler for short read sequence data. Genome Res. 19(6):1117–1123.

Slattery ML, Kerber RA. 1993. A Comprehensive Evaluation of Family History and Breast Cancer Risk: The Utah Population Database. JAMA J Am Med Assoc. 270(13):1563–1568.

Smit AFA, Hubley R, Green P. RepeatMasker Open-3.0.

Smyrk TC, Watson P, Kaul K, Lynch HT. 2001. Tumor-infiltrating lymphocytes are a marker for microsatellite instability in colorectal carcinoma. Cancer. 91(12):2417–2422.

Snow AK, Tuohy TMF, Sargent NR, Smith LJ, Burt RW, Neklason DW. 2015. APC promoter 1B deletion in seven American families with familial adenomatous polyposis. Clin Genet. 88(4):360–5.

Southey MC, Goldgar DE, Winqvist R, Pylkäs K, Couch F, Tischkowitz M, Foulkes WD, Dennis J, Michailidou K, van Rensburg EJ, et al. 2016. PALB2, CHEK2 and ATM rare variants and cancer risk: Data from COGS. J Med Genet. 53(12):800–811.

Stankiewicz P, Lupski JR. 2010. Structural Variation in the Human Genome and its Role in Disease. Annu Rev Med. 61(1):437–455.

Starita LM, Ahituv N, Dunham MJ, Kitzman JO, Roth FP, Seelig G, Shendure J, Fowler DM. 2017. Variant Interpretation: Functional Assays to the Rescue. Am J Hum Genet. 101(3):315–

325.

Stolzenberg-Solomon RZ, Graubard BI, Chari S, Limburg P, Taylor PR, Virtamo J, Albanes D. 2005. Insulin, glucose, insulin resistance, and pancreatic cancer in male smokers. J Am Med Assoc. 294(22):2872–2878.

Stranger BE, Brigham LE, Hasz R, Hunter M, Johns C, Johnson M, Kopen G, Leinweber WF, Lonsdale JT, McDonald A, et al. 2017. Enhancing GTEx by bridging the gaps between genotype, gene expression, and disease.

Strong VE, Gholami S, Shah MA, Tang LH, Janjigian YY, Schattner M, Selby L V., Yoon SS, Salo-Mullen E, Stadler ZK, et al. 2017. Total Gastrectomy for Hereditary Diffuse Gastric Cancer at a Single Center. Ann Surg. 266(6):1006–1012.

Sudmant PH, Rausch T, Gardner EJ, Handsaker RE, Abyzov A, Huddleston J, Zhang Y, Ye K, Jun G, Fritz MHY, et al. 2015. An integrated map of structural variation in 2,504 human genomes. Nature. 526(7571):75–81.

Suriano G, Yew S, Ferreira P, Senz J, Kaurah P, Ford JM, Longacre TA, Norton JA, Chun N, Young S, et al. 2005. Characterization of a recurrent germ line mutation of the E-cadherin gene: Implications for genetic testing and clinical management. Clin Cancer Res. 11(15):5401–5409.

Susswein LR, Marshall ML, Nusbaum R, Vogel Postula KJ, Weissman SM, Yackowski L, Vaccari EM, Bissonnette J, Booker JK, Cremona ML, et al. 2016. Pathogenic and likely pathogenic variant prevalence among the first 10,000 patients referred for next-generation cancer panel testing. Genet Med. 18(8):823–832.

Suter CM, Martin DIK, Ward RL. 2004. Germline epimutation of MLH1 in individuals with multiple cancers. Nat Genet. 36(5):497–501.

Swift M, Reitnauer PJ, Morrell D, Chase CL. 1987. Breast and Other Cancers in Families with Ataxia-Telangiectasia. N Engl J Med. 316(21):1289–1294.

Thibodeau ML, Zhao EY, Reisle C, Ch'ng C, Wong H-L, Shen Y, Jones MR, Lim HJ, Young S, Cremin C, et al. 2019. Base excision repair deficiency signatures implicate germline and somatic MUTYH aberrations in pancreatic ductal adenocarcinoma and breast cancer oncogenesis. Cold Spring Harb Mol case Stud. 5(2):a003681.

Thompson D, Duedal S, Kirner J, McGuffog L, Last J, Reiman A, Byrd P, Taylor M, Easton DF. 2005. Cancer risks and mortality in heterozygous ATM mutation carriers. J Natl Cancer Inst. 97(11):813–22.

Thompson ER, Doyle MA, Ryland GL, Rowley SM, Choong DYH, Tothill RW, Thorne H, Barnes DR, Li J, Ellul J, et al. 2012. Exome Sequencing Identifies Rare Deleterious Mutations in DNA Repair Genes FANCC and BLM as Potential Breast Cancer Susceptibility Alleles. Horwitz MS, editor. PLoS Genet. 8(9):e1002894.

Tischkowitz M, Brunet JS, Bégin LR, Huntsman DG, Cheang MCU, Akslen LA, Nielsen TO, Foulkes WD. 2007. Use of immunohistochemical markers can refine prognosis in triple negative breast cancer. BMC Cancer. 7(1):1–11.

Tran-Duy A, Spaetgens B, Hoes AW, de Wit NJ, Stehouwer CDA. 2016. Use of Proton Pump Inhibitors and Risks of Fundic Gland Polyps and Gastric Cancer: Systematic Review and Meta-analysis. Clin Gastroenterol Hepatol. 14(12):1706-1719.e5.

Tran CP, Lin C, Yamashiro J, Reiter RE. 2002. Prostate stem cell antigen is a marker of late intermediate prostate epithelial cells. Mol Cancer Res. 1(2):113–121.

Tung N, Domchek SM, Stadler Z, Nathanson KL, Couch F, Garber JE, Offit K, Robson ME.

2016. Counselling framework for moderate-penetrance cancer-susceptibility mutations. Nat Rev Clin Oncol. 13(9):581–588.

Tung N, Lin NU, Kidd J, Allen BA, Singh N, Wenstrup RJ, Hartman AR, Winer EP, Garber JE. 2016. Frequency of germline mutations in 25 cancer susceptibility genes in a sequential series of patients with breast cancer. J Clin Oncol. 34(13):1460–1468.

Umar A, Boland CR, Terdiman JP, Syngal S, de la Chapelle A, Rüschoff J, Fishel R, Lindor NM, Burgart LJ, Hamelin R, et al. 2004. Revised Bethesda Guidelines for hereditary nonpolyposis colorectal cancer (Lynch syndrome) and microsatellite instability. J Natl Cancer Inst. 96(4):261–8.

Vahteristo P, Bartkova J, Eerola H, Syrjäkoski K, Ojala S, Kilpivaara O, Tamminen A, Kononen J, Aittomäki K, Heikkilä P, et al. 2002. A CHEK2 genetic variant contributing to a substantial fraction of familial breast cancer. Am J Hum Genet. 71(2):432–438.

Vasen HFA, Mecklin J-P, Meera Khan P, Lynch HT. 1991. The International Collaborative Group on Hereditary Non-Polyposis Colorectal Cancer (ICG-HNPCC). Dis Colon Rectum. 34(5):424–425.

Vasen HFA, Möslein G, Alonso A, Aretz S, Bernstein I, Bertario L, Blanco I, Bülow S, Burn J, Capella G, et al. 2008. Guidelines for the clinical management of familial adenomatous polyposis (FAP). In: Gut. Vol. 57. BMJ Publishing Group. p. 704–713.

Vasen HFA, Watson P, Mecklin JP, Lynch HT. 1999. New clinical criteria for hereditary nonpolyposis colorectal cancer (HNPCC, Lynch syndrome) proposed by the International Collaborative Group on HNPCC. In: Gastroenterology. Vol. 116. W.B. Saunders. p. 1453–1456.

Verdun RE, Karlseder J. 2006. The DNA Damage Machinery and Homologous Recombination

Pathway Act Consecutively to Protect Human Telomeres. Cell. 127(4):709–720.

Verzi MP, Khan AH, Ito S, Shivdasani RA. 2008. Transcription Factor Foxq1 Controls Mucin Gene Expression and Granule Content in Mouse Stomach Surface Mucous Cells. Gastroenterology. 135(2):591–600.

Viel A, Bruselles A, Meccia E, Fornasarig M, Quaia M, Canzonieri V, Policicchio E, Urso ED, Agostini M, Genuardi M, et al. 2017. A Specific Mutational Signature Associated with DNA 8-Oxoguanine Persistence in MUTYH-defective Colorectal Cancer. EBioMedicine. 20:39–49.

Walsh T, Mandell JB, Norquist BM, Casadei S, Gulsuner S, Lee MK, King MC. 2017. Genetic Predisposition to Breast Cancer Due to Mutations Other Than BRCA1 and BRCA2 Founder Alleles Among Ashkenazi Jewish Women. JAMA Oncol. 3(12):1647–1653.

Wang RC, Smogorzewska A, De Lange T. 2004. Homologous recombination generates t-loop-sized deletions at human telomeres. Cell. 119(3):355–368.

Wang X, Ouyang H, Yamamoto Y, Kumar PA, Wei TS, Dagher R, Vincent M, Lu X, Bellizzi AM, Ho KY, et al. 2011. Residual embryonic cells as precursors of a Barrett's-like metaplasia. Cell. 145(7):1023–1035.

Wang X, Park J, Susztak K, Zhang NR, Li M. 2019. Bulk tissue cell type deconvolution with multi-subject single-cell expression reference. Nat Commun. 10(1):1–9.

Warthin AS. 1913. Heredity with reference to carcinoma: As shown by the study of the cases examined in the pathological laboratory of the university of michigan, 1895-1913. Arch Intern Med. XII(5):546–555.

Watson JD, Crick FHC. 1953. Molecular structure of nucleic acids: A structure for deoxyribose nucleic acid. Nature. 171(4356):737–738.

Weischenfeldt J, Symmons O, Spitz F, Korbel JO. 2013. Phenotypic impact of genomic structural variation: insights from and for human disease. Nat Rev Genet. 14.

Weischer M, Bojesen SE, Ellervik C, Tybjærg-Hansen A, Nordestgaard BG. 2008. CHEK2*1100delC genotyping for clinical assessment of breast cancer risk: Meta-analyses of 26,000 patient cases and 27,000 controls. J Clin Oncol. 26(4):542–548.

Weisenberger DJ, Siegmund KD, Campan M, Young J, Long TI, Faasse MA, Kang GH, Widschwendter M, Weener D, Buchanan D, et al. 2006. CpG island methylator phenotype underlies sporadic microsatellite instability and is tightly associated with BRAF mutation in colorectal cancer. Nat Genet. 38(7):787–793.

Wick RR, Judd LM, Holt KE. 2019. Performance of neural network basecalling tools for Oxford Nanopore sequencing. Genome Biol. 20(1):129.

Wilson JMG, Jungner G, Organization WH. 1968. Principles and practice of screening for disease / J. M. G. Wilson, G. Jungner.

de Witte C, Kutzera J, van Hoeck A, Nguyen L, Boere I, Jalving M, Ottevanger P, van Schaik - van de Mheen C, Stevense M, Kloosterman W, et al. 2020. Distinct genomic profiles are associated with treatment response and survival in ovarian cancer.

Wolpin BM, Bao Y, Qian ZR, Wu C, Kraft P, Ogino S, Stampfer MJ, Sato K, Ma J, Buring JE, et al. 2013. Hyperglycemia, insulin resistance, impaired pancreatic beta-cell function, and risk of pancreatic cancer. J Natl Cancer Inst. 105(14):1027–1035.

Wong H-L, Yang KC, Shen Y, Zhao EY, Loree JM, Kennecke HF, Kalloger SE, Karasinska JM, Lim HJ, Mungall AJ, et al. 2018. Molecular characterization of metastatic pancreatic neuroendocrine tumors (PNETs) using whole-genome and transcriptome sequencing. Cold

Spring Harb Mol case Stud. 4(1).

Wooster R, Neuhausen SL, Mangion J, Quirk Y, Ford D, Collins N, Nguyen K, Seal S, Tran T, Averill D, et al. 1994. Localization of a breast cancer susceptibility gene, BRCA2, to chromosome 13q12-13. Science (80-). 265(5181):2088–2090.

Worthley DL, Phillips KD, Wayte N, Schrader KA, Healey S, Kaurah P, Shulkes A, Grimpen F, Clouston A, Moore D, et al. 2012. Gastric adenocarcinoma and proximal polyposis of the stomach (GAPPS): a new autosomal dominant syndrome. Gut. 61(5):774–779.

Wright CF, FitzPatrick DR, Firth H V. 2018. Paediatric genomics: Diagnosing rare disease in children. Nat Rev Genet. 19(5):253–268.

Wu L, Shi W, Long J, Guo X, Michailidou K, Beesley J, Bolla MK, Shu XO, Lu Y, Cai Q, et al. 2018. A transcriptome-wide association study of 229,000 women identifies new candidate susceptibility genes for breast cancer. Nat Genet. 50(7):968–978.

Yanaru-Fujisawa R, Nakamura S, Moriyama T, Esaki M, Tsuchigame T, Gushima M, Hirahashi M, Nagai E, Matsumoto T, Kitazono T. 2012. Familial fundic gland polyposis with gastric cancer. Gut. 61(7):1103–4.

Yang X, Leslie G, Doroszuk A, Schneider S, Allen J, Decker B, Dunning AM, Redman J, Scarth J, Plaskocinska I, et al. 2020. Cancer risks associated with germline PALB2 pathogenic variants: An international study of 524 families. J Clin Oncol. 38(7):674–685.

Yoo J, Lee GD, Kim JH, Lee SN, Chae H, Han E, Kim Y, Kim M. 2020. Clinical validity of next-generation sequencing multi-gene panel testing for detecting pathogenic variants in patients with hereditary breast-ovarian cancer syndrome. Ann Lab Med. 40(2):148–154.

Zanghieri G, Gregorio C Di, Sacchetti C, Fante R, Sassatelli R, Cannizzo G, Carriero A, De

Leon MP. 1990. Familial occurrence of gastric cancer in the 2-year experience of a population-based registry. Cancer. 66(9):2047–2051.

Zhang C, Moriguchi T, Kajihara M, Esaki R, Harada A, Shimohata H, Oishi H, Hamada M, Morito N, Hasegawa K, et al. 2005. MafA Is a Key Regulator of Glucose-Stimulated Insulin Secretion. Mol Cell Biol. 25(12):4969–4976.

Zhou XP, Waite KA, Pilarski R, Hampel H, Fernandez MJ, Bos C, Dasouki M, Feldman GL, Greenberg LA, Ivanovich J, et al. 2003. Germline PTEN promoter mutations and deletions in Cowden/Bannayan-Riley-Ruvalcaba syndrome result in aberrant PTEN protein and dysregulation of the phosphoinositol-3-kinase/Akt pathway. Am J Hum Genet. 73(2):404–411.

Appendices

Appendix A Supplementary Materials for Chapter 2

Supplementary Table 2.1. Cancer predisposition genes evaluated for germline findings as part of the Personalized OncoGenomics program

Gene Symbol	Entrez ID	OMIM ID	Inheritance	Locus
ABRAXAS1 (FAM175A)	84142	611143	AD	4q21.23
AKT1	207	164730	SM	14q32.33
ALK	238	105590	AD	2p23.2
APC	324	611731	AD	5q21-q22
ATM	472	607585	CX	11q22.3
ATR	545	601215	AR	3q22-q24
AXIN2	8313	604025	AD	17q24
BAP1	8314	603089	AD	3p21.1
BARD1	580	601593	AD	2q34-q35
BLM	641	604610	AR	15q26.1
BMPR1A	657	601299	AD	10q22.3
BRCA1	672	113705	AD	17q21
BRCA2	675	600185	CX	13q12.3
BRIP1	83990	605882	AD	17q22
CBL	867	165360	AD	11q23.3
CDC73	79577	607393	AD	1q25-q31
CDH1	999	192090	AD	16q22.1
CDK4	1019	123829	AD	12q14
CDKN1B	1027	600778	AD	12p13
CDKN2A	1029	600160	AD	9p21
CHEK2	11200	604373	AD	22q12.1
DICER1	23405	606241	AD	14q32.13
DKC1	1736	305000	XLR	Xq28
EGFR	1956	131550	AD	7p11.2
EPCAM	4072	185535	CX	2p21
ERCC2	2068	126340	AR	19q13.2-q13.3
ERCC3	2071	133510	AR	2q21
ERCC4	2072	133520	AR	16p13.3-p13.13
ERCC5	2073	133530	AR	13q33

Supplementary Table 2.1. Cancer predisposition genes evaluated for germline findings as part of the Personalized OncoGenomics program (continued from previous page)

ETV6	2120	616216	AD	12p13.2
EZH2	2146	601573	AD	7q36.1
FANCA	2175	607139	AR	16q24.3
FANCC	2176	613899	AR	9q22.3
FH	2271	136850	CX	1q42.1
FLCN	201163	607273	AD	17p11.2
GATA2	2624	137295	AD	3q21.3
GREM1	26585	603054	AD	15q13.3
HNF1A	6927	142410	AD	12q24.31
HRAS	3265	190020	AD	11p15.5
IDH1	3417	147700	SM	2q34
KIT	3815	164920	AD	4q12
MAX	4149	154950	AD	14q23.3
MEN1	4221	613733	AD	11q13.1
MET	4233	164860	AD	7q31.2
MITF	4286	156845	AD	3p14-p13
MLH1	4292	120436	AD	3p22.2
MRE11	4361	600814	CX	11q21
MSH2	4436	609309	AD	2p21
MSH6	2956	600678	AD	2p16.3
MUTYH	4595	604933	AR	1p34.1
NBN	4683	602667	CX	8q21.3
NF1	4763	613113	AD	17q11.2
NF2	4771	607379	AD	22q12.2
NSD1	64324	606681	AD	5q35.2-q35.3
NTHL1	4913	602656	AR	16p13.3
PALB2	79728	610355	CX	16p12.2
PAX5	5079	167414	AD	9p13.2
PDGFRA	5156	173490	AD	4q12
PHOX2B	8929	603851	AD	4p13
PIK3CA	5290	171834	SM	3q26.32
PMS1	5378	600258	AD	2q31-q33
PMS2	5395	600259	AD	7p22.1
POLD1	5424	612591	AD	19q13.33
POLE	5426	174762	AD	12q24.33

Supplementary Table 2.1. Cancer predisposition genes evaluated for germline findings as part of the Personalized OncoGenomics program (continued from previous page)

PRKAR1A	5573	188830	AD	17q24.2
PTCH1	5727	601309	AD	9q22.32
PTEN	5728	601728	AD	10q23.31
PTPN11	5781	176876	AD	12q24.13
RAD50	10111	604040	CX	5q31.1
RAD51	5888	179617	AD	15q15.1
RAD51B	5890	602948	AD	14q24.1
RAD51C	5889	602774	AD	17q22
RAD51D	5892	602954	AD	17q12
RB1	5925	614041	AD	13q14.2
RECQL4	9401	603780	CX	8q24.3
RET	5979	164761	AD	10q11.21
RUNX1	861	151385	AD	21q22.12
SDHA	6389	600857	CX	5p15.33
SDHAF2	54949	613019	AD	11q12.2
SDHB	6390	185470	AD	1p36.13
SDHC	6391	602413	AD	1q23.3
SDHD	6392	602690	AD	11q23.1
SH2D1A	4068	300490	XR	Xq25
SMAD4	4089	600993	AD	18q21.2
SMARCA4	6597	603254	AD	19p13.2
SMARCB1	6598	601607	AD	22q11.23
STK11	6794	602216	AD	19p13.3
SUFU	51684	607035	AD	10q24.32
TERC	7012	127550	AD	3q26.2
TERT	7015	187270	AD	5p15.33
TGFBR1	7046	190181	AD	9q22.33
TINF2	26277	613990	AD	14q12
TMEM127	55654	613403	AD	2q11.2
TP53	7157	191170	AD	17p13.1
TSC1	7248	605284	AD	9q34.13
TSC2	7249	191092	AD	16p13.3
VHL	7428	608537	CX	3p25.3
WRN	7486	277700	AR	8p12
WT1	7490	607102	CX	11p13

Appendix B Supplementary Materials for Chapter 4

Supplementary Materials and Methods

Somatic mutation analysis

Small somatic mutations, somatic copy number alterations, somatic structural variants and loss of heterozygosity were identified as previously described (Pleasance 2020). Tumour purity, or neoplastic cellularity, was modeled from copy ratio data using pathologist-derived estimates. Substitution types and 5' and 3' nucleotide contexts were extracted for somatic SNVs using the R package MutationalPatterns (Blokzjil 2018). The number and percent of mutations in each of 6 substitution groups and 96 possible trinucleotide contexts were calculated, and the global pattern of mutations was compared to 30 mutational signatures characterized in version 2 of the Catalog of Somatic Mutations in Cancer (COSMIC) using a non-negative least squares method described by Blokzijl et al (https://cancer.sanger.ac.uk/cosmic/signatures_v2).

RNA sequencing

Peripheral blood RNA collected in the PAXgene Blood RNA Tube (PreAnalytix) for III-12 was extracted according to manufacturer's instructions using the RNeasy Mini Kit (Qiagen). Paired-end strand-specific RNA-seq libraries prepared from tumour and peripheral blood RNA were sequenced on the Illumina HiSeq to an average depth of 200 million reads. Paired-end reads were aligned to the human reference genome version hg19 using STAR version 2.5.2b and basic two-pass mapping (Dobin 2013). Duplicate reads were marked with Picard tools and gene- and transcript-level quantification was performed using RSEM. Expected counts were normalized within samples using fixed upper quartile normalization, and normalized counts for the TCGA PAAD cohort were obtained using TCGAbiolinks (Colaprico *et al.*, 2015; TCGA,

2017). For mRNA subtype classification and expression percentile calculations, batch correction was first performed using the R package sva to account for known biases in library preparation and sequencing protocols between project sites (Leek 2012). Corrected expression values were estimated by parametric empirical Bayesian adjustment defined by the sva ComBat function on log-transformed and normalized counts.

mRNA subtype classification

Prior to gene expression-based subtype classification, corrected count data was standardized to a mean of 0 and standard deviation of 1. Clustering was performed by non-negative matrix factorization using published gene expression classifiers. To derive comparable estimates of tumour purity across FPC, POG and TCGA cohorts, stromal and immune cell infiltration across the combined expression dataset was modeled using ESTIMATE (Yoshihara 2013). Tumour purity was then evaluated according to the following formula described by Yoshihara *et al*. (2013): cos(0.6049872018+ 0.0001467884 × ESTIMATE score). Across the combined dataset, three- and four-subtype classifications with 62 and 613 contributing genes, respectively, showed a greater dependence on tumour purity than the two-subtype 50-gene classifier described by Moffit *et al.* (2015) (Collisson *et al.*, 2011; Bailey *et al.*, 2016). Therefore, the two-subtype Moffitt classification to use in downstream analysis. A metabolic gene expression classifier was similarly applied and did not show significant contributions from tumour purity in glycolytic or cholesterogenic subtype classification (Karasinka 2019).

Differential expression analysis

Differential expression was evaluated between familial and unrelated POG PDACs using the R package DESeq2 (Love 2014). To account for natural variation in tissue-specific gene expression, tumour biopsy site was incorporated as a covariate while disabling automatic outlier removal using Cooks cut-off and independent filtering. Genes with a Benjamini Hochberg-adjusted P-value ≤ 0.01 and absolute fold-change > 1.5 were selected for pathway analysis using limma (Ritchie 2015). Gene expression percentiles for the FPC kindred were calculated using corrected count data independently between unrelated POG PDACs and TCGA PAADs.

Allele-specific expression analysis

Allelic read depth in tumour GS and RNA-seq was evaluated at biallelic heterozygous SNVs using GATK ASEReadCounter, requiring a minimum mapping quality of 10 and base quality of 20 (McKenna *et al.*, 2010). Loci with a variant allele frequency under 0.3 or above 0.7 in peripheral blood, with a read depth under 10 in RNA-seq or that occurred in regions with an ENCODE 50bp mappability score < 1 were excluded from downstream analysis. Gene-level ASE was computed using MBASED, and significant allelic imbalance was defined in genes with an adjusted p-value ≤ 0.05. Variant-level allelic imbalance was identified at sites with significant differences in DNA and mRNA allele frequency with an FDR-adjusted p-value ≤ 0.05 (Fisher's exact test).
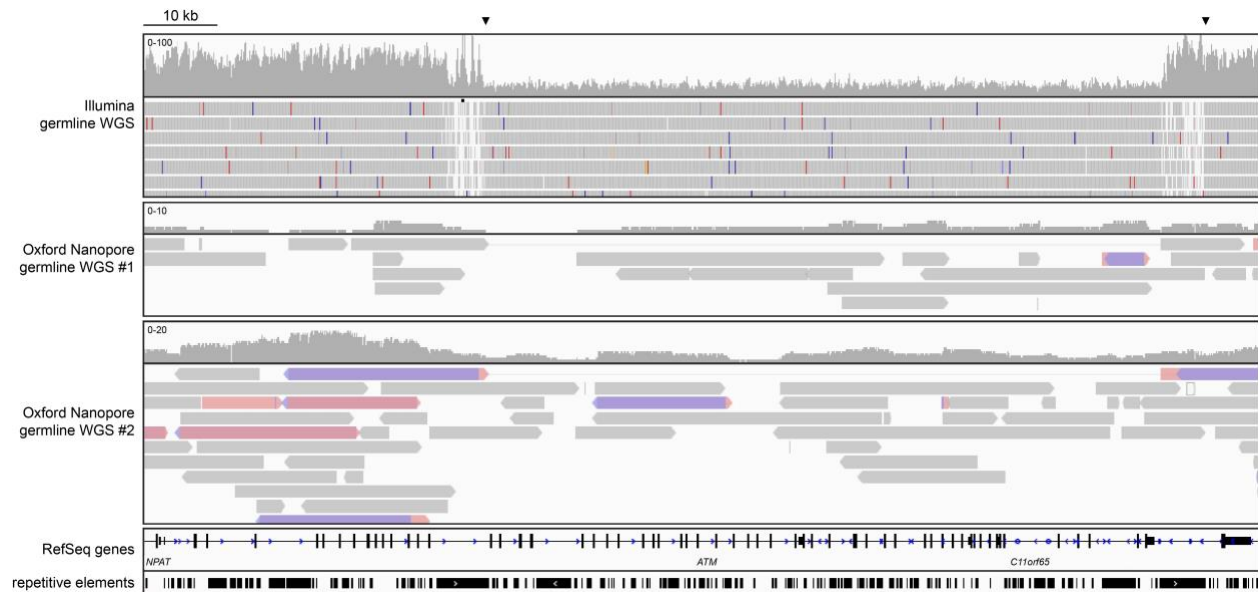
Supplementary Figure 4.1. Illumina and Oxford Nanopore genome sequencing data supporting a 96 kb deletion in *ATM*. Illumina and Oxford Nanopore genome sequencing data for Case 6 visualized using IGV at the locus of ATM. One Nanopore read spanning the breakpoint junction from two independent sequencing runs are shown mapping to flanking regions of the predicted breakpoints (black arrows) and are connected by a thin gray line.
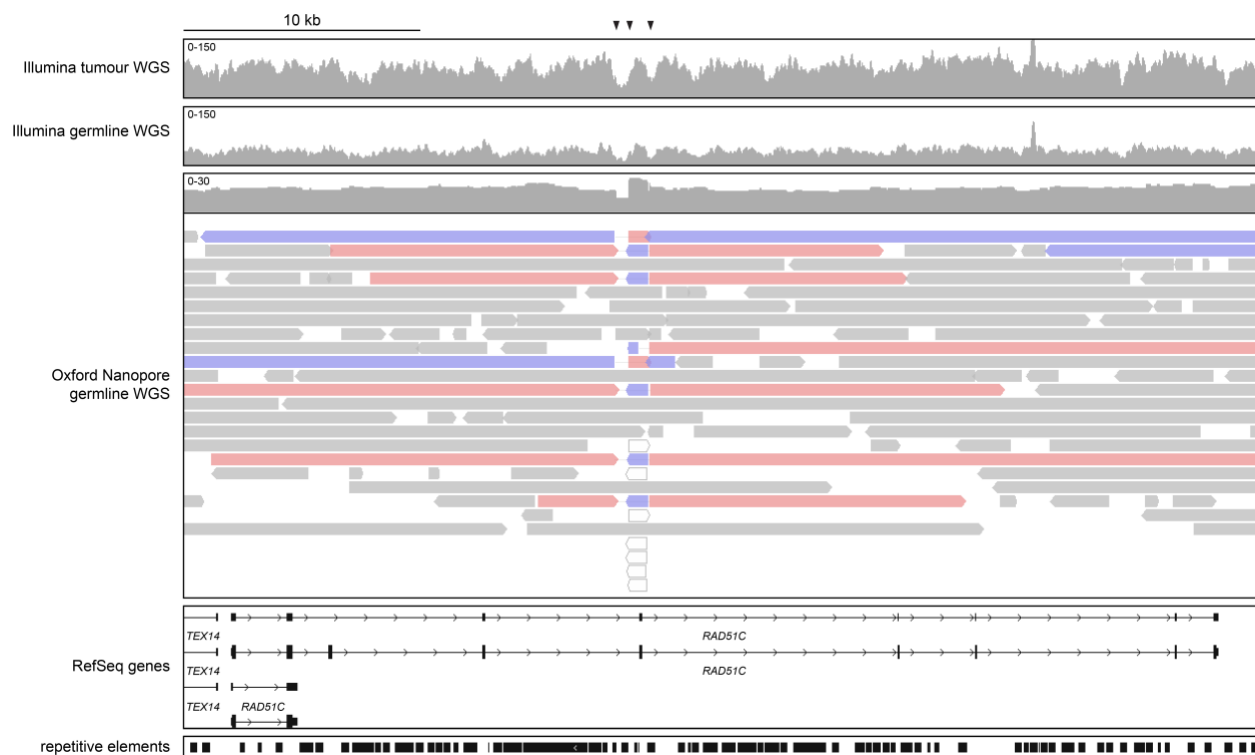
Supplementary Figure 4.2. Illumina and Oxford Nanopore genome sequencing data supporting a single-exon inversion in *RAD51C*. Illumina and Oxford Nanopore genome sequencing data for Case 7 visualized using IGV at the locus of RAD51C. Split Nanopore reads spanning the breakpoint junctions are shown mapping to flanking regions of the predicted breakpoints (black arrows) connected by a thin gray line. Read segments coloured red and blue denote split reads mapping to both plus and minus strands, indicating a probable inversion event.

Supplementary Table 4.1. Illumina and Oxford Nanopore genome sequencing and variant calling information for candidate germline structural variants.

ID	Chromosome	5' breakpoint	3' breakpoint	Type	Call method	5' breakpoint	3' breakpoint	Type (subtype)	Length	Variant reads	Runs
					Illumina GS variant calling	*Oxford Nanopore variant information*					
1	16p13	1,566,535	2,119,866	INV	custom script	1,566,516	1,566,651	INS (AL)	131 bp	7	1
2	16p13	1,566,535	2,119,866	INV	DELLY, Manta	1,566,507	1,566,633	INS (AL)	129 bp	3	1
3	16p13	1,566,535	2,119,866	INV	DELLY, Manta	1,566,499	1,566,631	INS (AL)	132 bp	10	1
4	5q35	176,441,543	176,603,468	INV	DELLY, Manta, Trans-ABySS	176,441,543	176,603,468	INV (SR)	161,925 bp	10	2
						176,409,771$_a$	176,441,549$_a$	INV (SR)	31,778 bp	15	2
5$_b$	16p13	2,126,780	2,214,187	INV	DELLY, Manta	2,126,780	2,214,187	INV (SR)	87,407 bp	1	2
						2,093,920	2,212,350	INV (SR)	118,430 bp	3	2
6	11q22	108,137,586	108,227,717	DEL	Control-FREEC	108,137,370	108,233,694	DEL (SR)	96,324 bp	2	2
7	17q22	56,786,751	56,787,647	INV	Manta	56,786,207	56,786,758$_c$	DEL (SR)	551 bp	5	2
						56,786,751	56,787,655$_c$	INV (SR)	904 bp	8	2
8	11q22	108,118,496	108,121,054	DEL	DELLY, Manta	108,118,507	108,121,041	DEL (AL, SR)	2,534 bp	9	1
9	17q21	41,217,614	41,295,110	DEL	Control-FREEC, DELLY, Manta	41,217,612	41,295,114	DEL (SR)	77,502 bp	4	1
10	17q21	41,235,786	41,250,846	DEL	Control-FREEC	41,236,461	41,250,954	DEL (AL, SR)	14,493 bp	8	2
11	2p21	47,545,553	47,674,137	DEL	Control-FREEC, DELLY, Manta	47,545,553	47,673,900	DEL (SR)	128,347 bp	8	1

Supplementary Table 4.1. Illumina and Oxford Nanopore genome sequencing and variant calling information for candidate germline structural variants. (continued from previous page)

12	16q24	89,844,986	89,869,214	DEL	Control-FREEC, DELLY, Manta, Trans-ABySS	89,844,987	89869211	DEL (SR)	24,224 bp	4	2
13	16p12	23,631,306	23,634,733	DEL	DELLY, Manta	23,631,313	23,634,736	DEL (AL)	3,423 bp	4	1
14	17p13	7,576,941	7,580,192	DEL	DELLY, Manta, Trans-ABySS	NA	NA	NA	NA	NA	NA

[a]Breakpoints resolved through manual curation: 176,409,841-176,441,555 (31,714 bp)
[b]Case 5 was sequenced only on the Oxford Nanopore MinION
[c]Breakpoints resolved through manual curation: 56,786,207-56,786,751 (544 bp)
[c]Breakpoints resolved through manual curation: 56,786,751-56,787,647 (896 bp)
AL, alignment; DEL, deletion; INS, insertion; SR, split reads

GCS
GAPPS Clinical Study

Patient and family member survey

We are studying the clinical features of gastric adenocarcinoma and proximal polyposis of the stomach (GAPPS), a hereditary cancer predisposition characterized by the presence of hundreds of benign growths (called polyps) in the lining of the stomach. Individuals with GAPPS have an increased risk of gastric (stomach) cancer, but the clinical features associated with GAPPS are not completely known.

This survey will allow us to better understand the clinical manifestations of GAPPS, factors that influence disease progression and estimate the risk for other cancers. Data collected from this study may be used to help form clinical guidelines regarding the treatment and management of GAPPS and help inform the health care decisions of affected individuals and their families.

Confidentiality

All personal information collected in this survey will be kept strictly confidential, and your identity will not be included in any reports on the findings of this study.

Instructions

Please answer each question to the best of your knowledge. For questions where a selection of answers is provided, please select the answer that best applies to you. **For family members who have had a gastrectomy, all questions relate to your health BEFORE gastrectomy**.

Some questions will ask for detailed information about symptoms that you may have experienced, medical conditions that you have had, and your use of specific medications. Please answer all of the questions to the best of your knowledge, and please do not hesitate to talk to the research assistant if you have any questions or concerns.

Thank you for your participation!

A. GAPPS

A1. What is your year of birth? .... ☐ ☐ ☐ ☐

A2. Have you ever been diagnosed with GAPPS?

☐ No   ☐ Yes → If yes, when were you diagnosed? .... Year ☐ ☐ ☐ ☐   Age ☐ ☐

→ How were you diagnosed?
☐ upper GI endoscopy   ☐ genetic testing for the GAPPS mutation

A3. Have you had a gastrectomy?

☐ No   ☐ Yes → If yes, when did it happen? .... Year ☐ ☐ ☐ ☐   Age ☐ ☐

B. Procedures

B1. *An upper GI endoscopy is used to visualize the esophagus and upper GI tract by passing a flexible tube carrying a light and camera (an endoscope) through the mouth.*

Have you ever had an upper GI endoscopy?   ☐ Yes   ☐ No → *Continue to section B2.*

→ If yes, when was/were the procedure(s) done?
→ Please describe the results of the procedure(s) in the box provided. *(E.g. gastric or duodenal polyps, cancer, gastric reflux, gastric ulcers, anemia.)*

1st upper GI endoscopy Year ☐☐☐☐ or Age ☐☐	Results →
2nd upper GI endoscopy Year ☐☐☐☐ or Age ☐☐	Results →
3rd upper GI endoscopy Year ☐☐☐☐ or Age ☐☐	Results →
4th upper GI endoscopy Year ☐☐☐☐ or Age ☐☐	Results →
If more than 4, when was your *last* upper GI endoscopy? Year ☐☐☐☐ or Age ☐☐	Results →

→ If known, please provide the name of your specialist(s) and/or the clinic location(s) where the procedure(s) were performed.

Specialist(s) name(s) →

Clinic location(s) →

B2. *A <u>colonoscopy</u> is used to visualize the colon by passing an endoscope through the anus.*

Have you ever had a colonoscopy? .... ☐ Yes   ☐ No → *Continue to section B3.*

→ If yes, when was/were the procedure(s) done?
→ Please describe the results of the procedure(s) in the box provided. *(E.g. colon polyps, desmoid tumours, cancer)*

1st colonoscopy Year ☐☐☐☐ or Age ☐☐	Results →
2nd colonoscopy Year ☐☐☐☐ or Age ☐☐	Results →
3rd colonoscopy Year ☐☐☐☐ or Age ☐☐	Results →
4th colonoscopy Year ☐☐☐☐ or Age ☐☐	Results →
If more than 4, when was your *last* colonoscopy? Year ☐☐☐☐ or Age ☐☐	Results →

→ If known, please provide the name of your specialist(s) and/or the clinic location where the procedure(s) were performed.

Specialist(s) name(s) →

Clinic location(s) →

B3. Please list any other procedures that you have had in the last 10 years related to gastro-intestinal symptoms (e.g. abdominal MRI or CT scan, ultrasound, capsule endoscopy...)

Procedure	Year	Results

C. Health
The following questions relate to your health BEFORE gastrectomy

C1. *At their worst,* **how often did you experience the following symptoms (unrelated to an acute episode of gastro-enteritis)?**

1. Heartburn	☐ Never → go to question 2 ☐ occasional ☐ weekly ☐ daily When did your symptoms start? ....Age ☐☐ Was/Is there anything in particular that triggered your symptoms (e.g. acidic foods, alcohol, caffeine)? *If yes, please specify.* [ ] Was/Is there anything in particular that relieve your symptoms? [ ] Did your symptoms stop or improve? ☐ They stopped. ☐ I still experience these symptoms. When did you last experience these symptoms? ....Age ☐☐
2. Nausea	☐ Never → go to question 3 ☐ occasional ☐ weekly ☐ daily When did your symptoms start? ....Age ☐☐ Was/Is there anything in particular that triggered your symptoms (e.g. acidic foods, alcohol, caffeine)? *If yes, please specify.* [ ] Was/Is there anything in particular that relieve your symptoms? [ ] Did your symptoms stop or improve? ☐ They stopped. ☐ I still experience these symptoms. When did you last experience these symptoms? ....Age ☐☐
3. Vomiting	☐ Never → go to question 4 ☐ occasional ☐ weekly ☐ daily When did your symptoms start? ....Age ☐☐ Was/Is there anything in particular that triggered your symptoms (e.g. acidic foods, alcohol or caffeine)? *If yes, please specify.*

	Was/Is there anything in particular that relieve your symptoms? Did your symptoms stop or improve? ☐ They stopped. ☐ I still experience these symptoms. When did you last experience these symptoms? ....Age ☐☐
4. Stomach pain	☐Never → go to question 5 ☐occasional ☐ weekly ☐ daily When did your symptoms start? ....Age ☐☐ Was/Is there anything in particular that triggered your symptoms (e.g. acidic foods, alcohol, caffeine)? *If yes, please specify.* Was/Is there anything in particular that relieve your symptoms? Did your symptoms stop or improve? ☐ They stopped. ☐ I still experience these symptoms. When did you last experience these symptoms? ....Age ☐☐
5. Regurgitation	☐Never → go to question 6 ☐occasional ☐ weekly ☐ daily When did your symptoms start? ....Age ☐☐ Was/Is there anything in particular that triggered your symptoms (e.g. acidic foods, alcohol, caffeine)? *If yes, please specify.* Was/Is there anything in particular that relieve your symptoms? Did your symptoms stop or improve? ☐ They stopped. ☐ I still experience these symptoms. When did you last experience these symptoms? ....Age ☐☐
6. Change or unexplained loss of appetite	☐Never → go to question 7 ☐occasional ☐ regular When did your symptoms start? ....Age ☐☐

	Was/Is there anything in particular that triggered your symptoms? *If yes, please specify.*
	Was/Is there anything in particular that relieve your symptoms?
	Did your symptoms stop or improve? ☐ They stopped. ☐ I still experience these symptoms. When did you last experience these symptoms? ....Age
7. Abdominal pain/bloating	☐Never → go to question 8 ☐occasional ☐ weekly ☐ daily When did your symptoms start? ....Age Was/Is there anything in particular that triggered your symptoms (e.g. acidic foods, alcohol, caffeine)? *If yes, please specify.* Was/Is there anything in particular that relieve your symptoms? Did your symptoms stop or improve? ☐ They stopped. ☐ I still experience these symptoms. When did you last experience these symptoms? ....Age
8. Rectal bleeding	☐Never → go to next section ☐occasional ☐ weekly ☐ daily When did your symptoms start? ....Age Was/Is there anything in particular that triggered your symptoms (e.g. acidic foods, alcohol, caffeine)? *If yes, please specify.* Was/Is there anything in particular that relieve your symptoms? Did your symptoms stop or improve? ☐ They stopped. ☐ I still experience these symptoms. When did you last experience these symptoms? ....Age

C2. Have you ever been diagnosed with any of the following medical conditions?

Gastric reflux ....☐ Yes ☐ No

If yes, how old were you when you were diagnosed?.. Age ☐☐

After your diagnosis, did you take any medication for this?...........................☐ Yes ☐ No

    If yes, for how long did you take medication? ....

    How often did you take them?.................................

    Please list medications taken →
    (E.g. Nexium, Somac, Zantac, Tagamet,
    Axid, Pepcid, Pariet, Losec, Pantaloc,
    Prevacid, other)

Gastric ulcer... ☐ Yes ☐ No

If yes, how old were you when you were diagnosed?.. Age ☐☐

After your diagnosis, did you take any medication for this?...........................☐ Yes ☐ No

    If yes, for how long did you take medication? ....

    How often did you take them?.................................

    Please list medications taken →
    (E.g. Nexium, Pariet, Somac, Pepcid,
    Tagamet, Zantac, amoxicillin,
    clarithromycin, proton pump inhibitors,
    other)

Anemia... ☐ Yes ☐ No

If yes, how old were you when you were diagnosed?... Age [ ][ ]

What type of anemia were you diagnosed with?

☐ Iron-deficiency anemia   ☐ Aplastic anemia   ☐ I'm not sure

☐ Pernicious anemia   ☐ Hemolytic anemia

Please describe the treatment(s) you received for your anemia →

(E.g. iron supplements, B12 supplements, blood transfusion, erythropoietin therapy, stem cell transplant)

Irritable bowel syndrome... ☐ Yes ☐ No

If yes, how old were you when you were diagnosed?... Age [ ][ ]

After your diagnosis, did you take any medication for this?........................... ☐ Yes ☐ No

If yes, for how long did you take medication? ....

How often did you take them? ....

Please list medications taken →

(E.g. Bentyl, Levsin, antidepressants, laxatives)

Inflammatory bowel disease (Crohn's or ulcerative colitis) .... ☐ Yes ☐ No

If yes, how old were you when you were diagnosed?.. Age ☐☐

After your diagnosis, did you take any medication for this?...........................☐ Yes ☐ No

If yes, for how long did you take medication? ....

How often did you take them? ....

Please list medications taken →

(E.g. Adalimumab, Azathioprine, Mercaptopurine, Infliximab, Methotrexate, other)

After your diagnosis, did you have a colectomy or proctocolectomy? ....☐ Yes ☐ No

Helicobacter pylori infection ....☐ Yes ☐ No

If yes, how old were you when you were diagnosed?.. Age ☐☐

After your diagnosis, did you take any medication for this?...........................☐ Yes ☐ No

If yes, for how long did you take medication? ....

How often did you take them? ....

Please list medications taken →
(E.g. Nexium, Pariet, Somac, Pantoloc, Losec, Prevacid, amoxicillin, clarithromycin, metronidazole, other)

Was the treatment successful? ....☐ Yes ☐ No

Was this confirmed through a follow-up test (e.g. breath test)? ....☐ Yes ☐ No

C3. Have you been diagnosed with any of the following? *Please check all that apply.*

☐ Desmoid tumours

☐ Osteomas (benign bony lumps)

☐ Dental abnormalities (supernumerary or missing teeth)

☐ Benign cutaneous lesions (epidermal cysts, lipomas, fibroma)

☐ Congenital hypertrophy and retinal pigment epithelium (CHRPE)

C4. Have you ever been diagnosed with cancer?

> *If yes, please indicate the type of cancer, the age you were when you were diagnosed and any treatments that you received in the table below.*
>
> *If no, continue to section C5.*

	Age at diagnosis	**Treatment received (e.g. medications, procedures, surgery)**
Gastric cancer		☐ surgery ☐ chemotherapy ☐ radiation ☐ other (specify) → [_____]
Colon cancer		☐ surgery ☐ chemotherapy ☐ radiation ☐ other (specify) → [_____]
Other: *please specify* → [_____]		☐ surgery ☐ chemotherapy ☐ radiation ☐ other (specify) → [_____]
Other: *please specify* → [_____]		☐ surgery ☐ chemotherapy ☐ radiation ☐ other (specify) → [_____]
Other: *please specify* → [_____]		☐ surgery ☐ chemotherapy ☐ radiation ☐ other (specify) → [_____]

C5. *A stool test is often used to screen for early signs of colon cancer. These include the gFOBT (guaiac-based fecal occult blood test) and FIT (fecal immunochemical test).*

Have you ever had a stool test (gFOBT or FIT)? .... ☐ Yes   ☐ No → *Continue to section C6.*

> → If yes, when was your *first* stool test? .... Year [ ][ ][ ][ ] or Age [ ][ ]
> → How many tests have you had? .... [ ]

C6. If you have been diagnosed with any other heart problems or other gastric or abdominal problems, please list them below.

Condition	Age at diagnosis	Treatment received (e.g. medications, procedures, surgery)

D. Medication Use

D1. Please indicate if you have taken any of the following medications, including when they were taken and the dose (in mg or µg/day) if known.

→ **Medication for reflux heartburn** *(e.g. Pariet, Pantoloc, Losec, Prevacid, Nexium, Dexilant, Zantac, Tagamet, Axid, Pepcid)*

Name	Dose	How often do you take it	When did you start taking it (age or year)	How long did you take it for
		☐ Daily ☐ Occasionally ☐ Rarely (less than every 6 months)		
		☐ Daily ☐ Occasionally ☐ Rarely		
		☐ Daily ☐ Occasionally ☐ Rarely		
		☐ Daily ☐ Occasionally ☐ Rarely		
		☐ Daily ☐ Occasionally ☐ Rarely		
		☐ Daily ☐ Occasionally ☐ Rarely		

→ **Heart or blood pressure medication** (e.g. Adesan, Candesartan, Asartan, Atacand, Covazan, Karvea, Avapro, Cozaar, Micardis, Atacand, Teveten, Olmetec, Diovan, other)

Name	Dose	How often do you take it	When did you start taking it (age or year)	How long did you take it for
		☐ Daily ☐ Occasionally ☐ Rarely		
		☐ Daily ☐ Occasionally ☐ Rarely		
		☐ Daily ☐ Occasionally ☐ Rarely		
		☐ Daily ☐ Occasionally ☐ Rarely		
		☐ Daily ☐ Occasionally ☐ Rarely		
		☐ Daily ☐ Occasionally ☐ Rarely		
		☐ Daily ☐ Occasionally ☐ Rarely		

→ **Aspirin**

Name	Dose	How often do you take it	When did you start taking it (age or year)	How long did you take it for
		☐ Daily ☐ Occasionally ☐ Rarely		
		☐ Daily ☐ Occasionally ☐ Rarely		

		☐ Daily ☐ Occasionally ☐ Rarely		

→ **Other anti-inflammatory medications** *(e.g. Celebrex, Voltaren, Advil, Anaprox, Aleve)*

Name	Dose	How often do you take it	When did you start taking it (age or year)	How long did you take it for
		☐ Daily ☐ Occasionally ☐ Rarely		
		☐ Daily ☐ Occasionally ☐ Rarely		

→ **Other medications**

Name	Dose	How often do you take it	When did you start taking it (age or year)	How long did you take it for
		☐ Daily ☐ Occasionally ☐ Rarely		
		☐ Daily ☐ Occasionally ☐ Rarely		
		☐ Daily ☐ Occasionally ☐ Rarely		
		☐ Daily ☐ Occasionally ☐ Rarely		
		☐ Daily ☐ Occasionally ☐ Rarely		

E. Diet and Lifestyle

E1. Do you have any dietary intolerances (e.g. lactose, gluten)?

☐ Yes → If yes, please specify: [   ]

☐ No

E2. Do you follow a specific diet (e.g. vegetarian, vegan, gluten-free)?

☐ Yes → If yes, please specify: [   ]

☐ No

E3. On average, how many times *per week* do you eat unprocessed red meat (e.g. steak, lamb)?

☐ Never   ☐ 1-3   ☐ 4-7   ☐ If more often, please specify: [   ] times per week

E4. On average, how many times *per week* do you eat processed red meat (e.g. bacon, sausage)?

☐ Never   ☐ 1-3   ☐ 4-7   ☐ If more often, please specify: [   ] times per week

E5. Do you take any vitamins or supplements?

☐ Yes → If yes, please specify: [   ]

☐ No

E6. Have you ever been a regular smoker (daily for at least 6 months)?............ ☐ Yes   ☐ No

If yes, how old were you when you started? .... [   ]

Do you still smoke regularly?   ☐ Yes   ☐ No .... *If no, age stopped:* [   ]

On average, how many cigarettes do/ did you smoke *per day*? .... [   ]

E7. Do you drink alcohol?... ☐ Yes   ☐ No

If yes, how many standard drinks of alcohol do you drink per week on average? .... [   ]

How often do you have a drink containing alcohol?
☐ Never   ☐ Monthly or less   ☐ 2-4 times/month   ☐ 2-3 times/week   ☐ 4+ times/week

How many standard drinks* of alcohol do you drink on a typical day when you are drinking?
☐ 1-2   ☐ 3-4   ☐ 5-6   ☐ 7-9   ☐ 10+

How often do you have 5 or more drinks on one occasion?
☐ Never   ☐ Less than monthly   ☐ Monthly   ☐ Weekly   ☐ Daily or almost daily

*__Australian standard drinks:__ the Australian standard drink measure contains 10grams of alcohol (equivalent to 12.5mls of pure alcohol). For example:
- 100ml glass of red wine at 13% alc vol = 1 standard drink.
- 100ml glass of white wine at 11.5% alc vol = 0.9 of a standard drink.
- 375ml bottle or can of full strength beer at 4.8% alc vol = 1.4 standard drinks.
- 30ml nip of high strength spirit at 40% alc vol = 1 standard drink.
- 330ml bottle of full strength ready-to-drink 5% acl vol = 1.2 standard drinks.

F. Additional Information

Is there anything else that you would like to tell us about your health or lifestyle that you feel might be relevant to this questionnaire? If so, please tell us in the space provided.

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

Thank you!