A molecular genetic investigation of *FANCI* as a new candidate ovarian cancer predisposing gene

Caitlin Fierheller

Department of Human Genetics

Faculty of Medicine and Health Sciences

McGill University, Montreal

December 2022

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of the degree of Doctor of Philosophy

Table of Contents

Abstract

Heterozygous carriers of loss-of-function (pathogenic) variants in *BRCA1* or *BRCA2* have an estimated absolute risk of 13 to 58% for ovarian cancer, especially of high grade serous ovarian carcinoma, the most common subtype of ovarian cancer. As not all familial cases of ovarian cancer are attributed to germline pathogenic variants in the known predisposing genes, we performed whole exome sequencing analysis of an unusual *BRCA1* and *BRCA2* pathogenic variant negative French Canadian ovarian cancer family to identify new candidate alleles. A rare missense variant *FANCI* c.1813C>T; p.L605F was identified in the heterozygous state. The aim of this study was to investigate *FANCI* as a candidate ovarian cancer predisposing gene using a molecular genetic approach involving genetic analyses of: the frequency of *FANCI* c.1813C>T in ovarian cancer cases and controls; the germline and somatic exomic landscape of *FANCI* c.1813C>T ovarian cancer carriers; and germline and somatic *FANCI* variants across different cancer types. I identified *FANCI* c.1813C>T more commonly in French Canadian ovarian cancer families compared to cancer-free controls, suggesting its role in risk. This, with in cellulo data, genetic analyses in ovarian cancer cases from other populations, and protein expression analyses, supports the role *FANCI* may play in ovarian cancer risk. Further investigation of the discovery *FANCI* c.1813C>T ovarian cancer family of French Canadian ancestry revealed 66 candidate variants most likely to affect protein function. None of these variants were identified in other *FANCI* c.1813C>T OC cases of French Canadian ancestry. Loss or allelic imbalance of wild type *FANCI* c.1813C>T allele detected in some tumours suggests that abrogation of *FANCI* protein function occurs at the cellular level in ovarian

tumourigenesis. The somatic variant profile of tumour cells from *FANCI* c.1813C>T ovarian cancer carrier cases exhibited features consistent with the known molecular genetic characteristics shared among ovarian cancer tumours. Germline *FANCI* c.1813C>T and other somatic variants were identified in diverse cancer cases, suggesting a possible role in tumourigenesis in other cancer types. This study is the first to propose and provide evidence in support of *FANCI* as a new candidate ovarian cancer predisposing gene, which may account for some of the hereditary ovarian cancer cases that are not due to *BRCA1*, *BRCA2,* and other ovarian cancer predisposing genes.

Résumé

Les porteuses hétérozygotes de variants induisant une perte de fonction (pathogéniques) de *BRCA1* ou *BRCA2* ont un risque absolu de cancer de l'ovaire estimé entre 13 et 58 %, particulièrement pour le carcinome séreux de l'ovaire de haut grade, le sous-type le plus courant. Puisque tous les cas familiaux de cancer de l'ovaire ne sont pas attribuables à des variants germinaux pathogéniques de gènes de prédisposition connus, nous avons effectué une analyse par séquençage de l'exome entier d'une famille de cancers de l'ovaire d'ascendance canadienne-française, négative pour les variants pathogéniques communs de *BRCA1* et *BRCA2* afin d'identifier de nouveaux candidats alléliques. Un variant rare de type faux-sens *FANCI* c.1813C>T; p.L605F a été identifié à l'état hétérozygote. Le but de cette étude était d'étudier *FANCI* en tant que gène candidat de prédisposition au cancer de l'ovaire à l'aide d'une approche de génétique moléculaire impliquant des analyses génétiques de : la fréquence de *FANCI* c.1813C>T dans les cas de cancer de l'ovaire et les contrôles ; le portrait exomique germinal et somatique des patientes atteintes de cancer de l'ovaire porteuses de *FANCI* c.1813C>T; la présence de variants germinaux et somatiques de *FANCI* dans différents types de cancer. J'ai identifié *FANCI* c.1813C>T plus fréquemment dans les cas de cancer de l'ovaire chez les familles d'ascendance canadienne-française par rapport au groupe témoin sans cancer, ce qui suggère un rôle de *FANCI* dans l'augmentation du risque. Ceci, ainsi que des données in cellulo, des analyses génétiques de cas de cancer de l'ovaire dans d'autres populations et des analyses d'expression protéique, soutiennent le rôle que *FANCI* peut jouer dans le risque de développer un cancer de l'ovaire. Une enquête plus approfondie de la famille

de cancer de l’ovaire d’ascendance canadienne-française FANCI c.1813C>T a révélé 66 candidats variants susceptibles d'affecter la fonction de la protéine. Aucun de ces variants n'a été identifié dans d'autres cas de cancer de l’ovaire *FANCI* c.1813C>T d'ascendance canadienne-française. La perte ou le déséquilibre allélique de l’allèle *FANCI* c.1813C>T de type sauvage détectée dans certaines tumeurs suggère que l'abrogation de la fonction protéique de *FANCI* se produit au niveau cellulaire dans la tumorigenèse ovarienne. Le profil de variant somatique des cellules tumorales dérivées des cas de cancer de l’ovaire porteurs de *FANCI* c.1813C>T présentait des caractéristiques compatibles avec les caractéristiques génétiques moléculaires connues partagées par les tumeurs du cancer de l'ovaire. Le variant germinal *FANCI* c.1813C>T et d'autres variants somatiques ont été identifiés dans divers cancers, suggérant un rôle tumorigénique possible dans d'autres types de cancer. Cette étude est la première à proposer et à fournir des évidences supportant le rôle de *FANCI* en tant que nouveau gène de prédisposition au cancer de l'ovaire, ce qui pourrait expliquer certains cas de cancer de l’ovaire héréditaires qui ne sont pas dus à *BRCA1*, *BRCA2* ou à d'autres gènes de prédisposition au cancer de l'ovaire.

Acknowledgements

I didn't always know that I wanted to be a scientist, mainly because I never knew any scientists. I hope I can inspire future scientists to pursue their education like my many mentors inspired me. I'd therefore first like to thank my high school science teachers who encouraged me to learn more about the subject I was so passionate about, my undergraduate professors who expanded the world of science so I could learn about genetics and biochemistry, and my honors thesis supervisor, Dr. Kathleen Hill, for allowing me to work on the most interesting organism (in my opinion), the naked mole rat.

My undergraduate studies on naked mole rat genetics brought me to McGill. I'd like to thank the Department of Human Genetics, especially Ross MacKay and Rimi Joshi, for guidance and instruction in all department matters. Thank you to the Cancer Research Program administrative team, Dr. Marie-Claude Gingras, Veronica Atehortua, and Zehra Khoja. I am grateful for the funding support from the Faculty of Medicine and Health Sciences and the Research Institute of the McGill University Health Centre. Travel funding support from Ovarian Cancer Canada allowed me to travel to the laboratory of Dr. Jean-Yves Masson at Laval University to learn about the biological techniques being used to assess the function of FANCI. I thank all Dr. Masson's lab members for making me feel welcome and allowing me to shadow experiments, but I am especially grateful to Dr. Laure Guitton-Sert for imparting her unparalleled knowledge on me. I appreciate the input and guidance of my supervisory committee members, Drs. Jacques Lapointe, Rima Slim, and Celia Greenwood. To all of the collaborators worldwide who provided genetic data (Drs. Ian Campbell, Trevor Pugh,

Marc Tischkowitz, Luigi Bouchard, and Simon Gravel), performed biological experiments (Dr. Jean-Yves Masson), provided samples/clinical data (Drs. Anne-Marie Mes-Masson, Will Foulkes, Zaki El Haffaf, and Diane Provencher), or performed genetic analyses (Dr. Jiannis Ragoussis) I am appreciative for you and your employees/student's willingness to collaborate on this project. The genetic sequencing of samples by Dr. Jiannis Ragoussis' lab was integral to this thesis and I am very thankful for bioinformatic analysis of, troubleshooting, and re-sequencing when needed. I'm thankful for the bioinformatics knowledge imparted by Dr. Setor Amuzu and for the many insightful discussions we had about ovarian cancer. I appreciate the prior analyses done by Suzanna Arcand, Dr. Karine Bédard, and Dr. Jacek Majewski on the discovery *FANCI* family. Nancy Hamel and Celine Domecq provided invaluable lab advice and assistance. I appreciate all the thoughtful feedback on manuscripts and this thesis provided by Prof Mary Fugiwara. Thank you to all the wonderful students I had the opportunity to teach over the years for making me a better teacher/mentor. To anyone who has had a hand in my progress so far, I am sincerely grateful.

To the Tonin lab: there are no words to describe how important this team is to me and the invaluable support I was provided over the last 6 years. I appreciate the unwavering encouragement of Corinne Serruya both in and outside the lab. Dr. Wejdan Alenezi was the best lab mate I could have asked for, without which this experience would have been wholly different. I truly could not have asked for a better supervisor and mentor for my graduate education, and for that I will never be able to thank Dr. Patricia Tonin enough. Your unyielding support, encouragement, and courage have inspired me in science and in life.

Finally, to my family and friends, who provided the support system to allow me to persevere: my partner (Daryl) who loves science as much as I do and has never wavered in his certainty in me; my parents (Greg and Theresa) who have been confident in me in everything; my brothers (Stuart and Garrett) who keep me grounded; my sisters (Kaitlyn and Jenna) who fiercely back me; my grandparents (Doug and Dyane, Frank and Mary Lou, and Charlene and Neil) who have never doubted me; and my extended family for supporting me every step of the way.

Contribution to original knowledge

Since the reports of the first cancer predisposing genes (CPGs) in the 1980s[1–5] considerable effort has been made to determine predisposition genes underlying cancers. Over 100 CPGs have since been identified[6], with many more being reported each year[7], though they have not all been validated. My strategy, using highly selective inclusion criteria for the discovery of a CPG in a genetically unique population exhibiting genetic drift, allowed me to identify a candidate variant that is uncommon in the general population. I initially used a candidate gene approach to identify the most plausible candidate for further investigation, which was supported by my investigation of the germline landscape of my discovery family F1528. We used a multidisciplinary approach to provide biological evidence of my candidate variant's effect on protein function to complement genetic data suggesting an involvement in ovarian cancer risk. Collectively, this approach provided evidence for the identification of a likely pathogenic variant *FANCI*, a previously unknown CPG.

This project provides evidence in support of *FANCI* as a new candidate ovarian CPG, which may account for some of the hereditary ovarian cancer cases that are not due to *BRCA1*, *BRCA2*, and other ovarian CPGs. In the future, my research findings, once validated, may be used to develop a genetic biomarker for identifying those at risk for ovarian cancer due to *FANCI*, and thus help in management of their risk as has been established for carriers of pathogenic variants in known ovarian cancer risk genes in medical genetic settings.

Contribution of authors

Chapter 2

C.T.F. performed the genetic experiments and wrote the initial draft of the manuscript. L.G.-S and Y.G. performed the in cellulo and in vitro experiments; J.R. developed the sequencing data analysis pipelines; C.T.F., W.M.A., J.M., J.N., R.B., J.P.B., and T.R. performed bioinformatic analyses; C.T.F. and K.K.O. performed statistical analyses; and L.M. and H.F. performed the immunohistochemistry of tissue microarray and associated statistical analyses. Study samples and clinical data were provided by Z.E., E.F., W.D.F., A.-M.M.-M., D.P., and D.N.S. Clinical and genotype data were provided by I.G.C., Z.E., E.F., P.A.J., D.P., M.T., and D.N.S; and C.T.F. and C.S. collated the study samples and clinical data. Supervision of bioinformatic analyses was provided by T.J.P., C.M.T.G., J.R., and P.N.T.; whole exome sequencing by J.M. and J.R.; genetic analysis by C.T.F., S.B., K.B., S.L.A., P.N.T.; and in cellulo and in vitro experiments by J.-Y.M. P.N.T. conceived, designed, and oversaw all aspects of the project. Critical input was provided by I.G.C., Z.E., E.F., J.-Y.M., W.D.F., P.A.J., A.-M.M.-M., D.P., M.T., and D.N.S. All authors contributed to and reviewed the manuscript. All authors read and approved the final manuscript.

Chapter 3

C.T.F. performed all genetic variant analysis and wrote the initial draft of the manuscript. J.R. developed the sequencing data analysis pipelines. C.T.F., W.M.A., S.A., and T.R. performed bioinformatic analyses. K.B. performed initial genetic analyses of family F1528. Study samples and clinical data were provided by Z.E., W.D.F., A.-M.M.-M.,

D.P., E.F., M.T., D.N.S., and I.C. C.T.F. and C.S. collated the study samples and clinical data. TCGA PanCancer Atlas germline *FANCI* c.1813C>T carrier data was provided by J.P.B., S.P., and T.J.P. Gen3G data was provided by S.G. and L.B. Statistical analyses was overseen by C.M.T.G. Supervision of bioinformatic analyses was provided by J.R. and P.N.T.; whole exome sequencing by J.R.; and genetic analysis by C.T.F. and P.N.T. P.N.T. conceived, designed, and oversaw all aspects of the project. All authors contributed to and reviewed the manuscript.

List of Figures and Tables

List of Abbreviations

ACMG	American College of Medical Genetics and Genomics
ADA	AdaBoost
ARM	Armadillo repeat
AUS	Australian
BC	Breast cancer
BCAC	Breast Cancer Association Consortium
BWA	Burroughs–Wheeler algorithm
CADD	Combined Annotation Dependent Depletion
CDK	Cyclin dependent kinase
CDN	Canadian
CI	Confidence interval
CHX	Cycloheximide
CNA	Copy number alteration
Condel	Consensus Deleteriousness
COEUR	Canadian Ovarian Experimental Unified Resource
COSMIC	Catalog of Somatic Mutations in Cancer
CPG	Cancer predisposing gene
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
dbscSNV	Database Splicing Consensus Single Nucleotide Variant
dbSNP	Single Nucleotide Polymorphism database
DIP	Database of Interacting Proteins
DEB	1,2:3,4-Diepoxybutane

EDGE	consists of Glutamic acid [E] - Aspartic acid [D] - Glycine [G] - Glutamic acid [E] amino acids
ESP	Exome Sequencing Project
EUR	Europe
EV	Empty vector
ExAC	Exome Aggregation Consortium
FA	Fanconi anemia
FACETS	Fraction and Allele specific Copy number Estimate from Tumour-normal Sequencing
FANCI	Fanconi anemia complementation group I gene
FC	French Canadian
FFPE	Formalin-fixed paraffin-embedded
FTE	Fallopian tube epithelium
Gen3G	Genetics of Glucose regulation in Gestation and Growth
GERP++	Genomic Evolutionary Rate Profiling
gnomAD	Genome Aggregation Database
GWAS	Genome Wide Association Study
HBC	Hereditary breast cancer
HBOC	Hereditary breast and ovarian cancer
HGSC	High-grade serous ovarian carcinoma
HR	Homologous recombination
ICL	Interstrand crosslink
ID2	FANCI-FANCD2 binding complex

IGV	Integrative Genomics Viewer
IHC	Immunohistochemistry
KD	Knockdown
KO	Knockout
MAF	Minor allele frequency
MANTIS	Microsatellite analysis for normal tumour instability
MaxEntScan	Maximum Entropy Modeling of Short Sequence Motifs
MetaLR	Meta-analytic Logistic Regression
MetaSVM	Meta- analytic Support Vector Machine
MGC	McGill Genome Centre
MINT	Molecular Interaction Database
MMC	Mitomycin C
MSI	Microsatellite instability
MUHC	McGill University Health Centre
NCCN®	National Comprehensive Cancer Network®
NHLBI	National Heart, Lung, and Blood Institute
NLS	Nuclear localization signal
OC	Ovarian cancer
OCAC	Ovarian Cancer Association Consortium
OR	Odds ratio
PARPi	Poly (ADP-ribose) polymerase inhibitor
PBL	Peripheral blood lymphocyte
PBWT	Positional Burrows-Wheeler transform

PhastCons	Phylogenetic Analysis with Space/Time models Conservation
PhyloP	Phylogenetic P values
PICKLE	Protein Interaction Knowledgebase
PINOT	Protein Interaction Network Online Tool
PolyPhen2	Polymorphism Phenotyping v2
PROVEAN	Protein Variant Effect Analyzer
REVEL	Rare Exome Variant Ensemble Learner
RF	Random Forest
RRCancer	Réseau de recherche sur le cancer
SBS	Single Base Substitution
SE	Standard error
SEM	Standard error of the mean
SIFT	Sorting Intolerant from Tolerant
SIGNOR	Signalling Network Open Resource
SiPhy	Site-specific Phylogenetic analysis
SLD2	SUMO-like domain 2
SLIM	SUMO-like domain-interacting motif
SNV	Single nucleotide variant
SSPS	Statistical Package for the Social Sciences
TCGA	The Cancer Genome Atlas
TMA	Tissue microarray

VACTERL-H	Vertebral abnormalities, Anal atresia, Cardiac defects, Tracheoesophageal fistula, Esophageal atreasia, Renal and radial abnormalities, Limb abnormalities with Hydrocephalus
VAF	Variant allele frequency
VEST	Variant Effect Scoring Test
WT	Wild type
WES	Whole exome sequencing
WGS	Whole genome sequencing

1.0 Chapter 1: Introduction

1.1 Ovarian cancer genetics

Ovarian cancer (OC) is the most lethal gynecologic malignancy and has a lifetime risk of 1.3% in the North American population[8]. OC is the tenth most common cancer diagnosed in females in Canada[9] with an estimated 3,000 new cases diagnosed in 2022[10]. OC has been estimated, from twin studies, to have a heritable component between 22 and 39%[11,12] and individuals with a first-degree relative diagnosed with OC have a 3- to 7-fold increase in risk[13]. Though the heritability of OC is estimated to be high, the search for inherited cancer predisposition genes (CPGs) began with more prevalent cancers, such as colorectal and breast. By 1990, a region of chromosome 17 was associated with breast cancer (BC) families and suspected of harbouring a major gene that predisposes to BC[14]. The observation that many of these BC families also had relatives diagnosed with OC suggested that a risk gene may not be specific to only one tissue. Further studies of these families eventually led to the identification of the breast and ovarian cancer risk gene, *BRCA1*, in 1994[15,16]. There was still a large proportion of families unaccounted for by *BRCA1*, which prompted investigation into other possible risk loci, which led to the identification of *BRCA2* on chromosome 13 in 1995[17,18]. Other OC predisposing genes have now been identified and are discussed below and high and moderate penetrance genes (Figure 1.1).

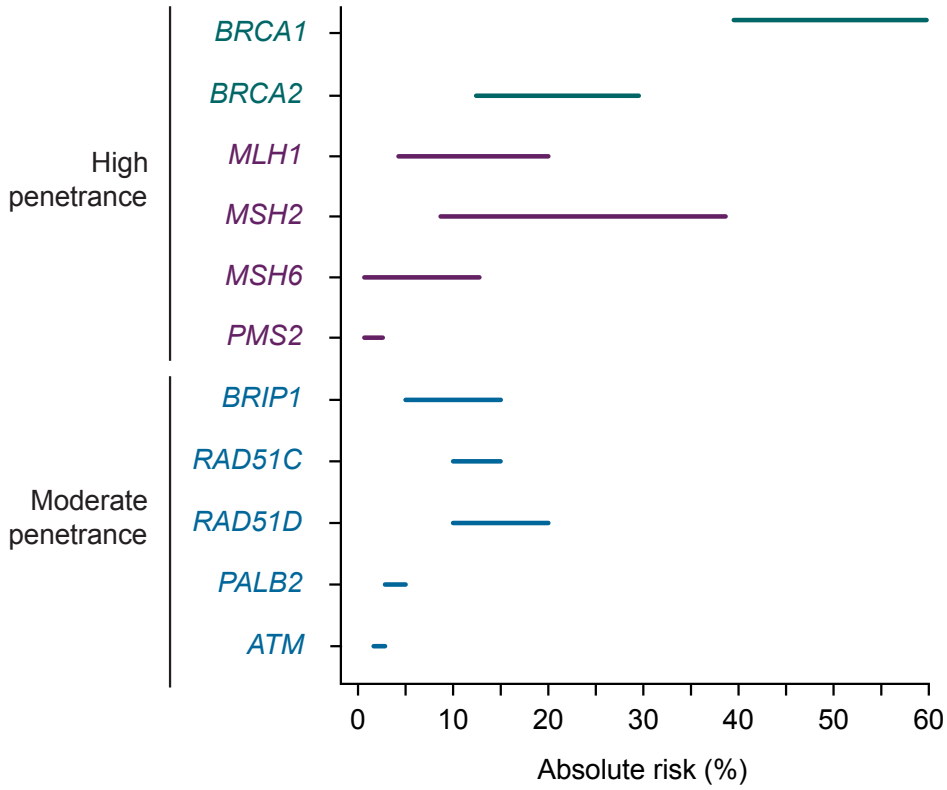


Figure 1.1. Absolute risk associated with moderate and high penetrance OC predisposing genes. The major genes, *BRCA1* and *BRCA2*, are shown in teal, mismatch repair genes are shown in purple, and all moderate penetrance genes are shown in blue.

Absolute risks reported by National Comprehensive Cancer Network® Clinical Practice Guidelines in Oncology – Genetic/Familial High Risk Assessment: Breast, Ovarian, and Pancreatic Version 1.2023 – September 7, 2022[19].

1.1.1 High penetrance genes

Both *BRCA1* and *BRCA2* are high penetrance OC predisposing genes. Pathogenic variants in these genes are associated with an autosomal dominant pattern of inheritance and are identified in 50-85% of hereditary breast and ovarian cancer (HBOC) syndrome families and in 10-15% of all epithelial OC cases[20–22]. As pathogenic variants in *BRCA1* and *BRCA2* are identified in a large proportion of OC cases, they are referred to as the major OC risk genes. Pathogenic variants in *BRCA1* and *BRCA2* confer absolute risks for OC of 39-58% and 13-29% (Figure 1.1), respectively, reported by the National Comprehensive Cancer Network® (NCCN®) Clinical Practice Guidelines in Oncology – Genetic/Familial High Risk Assessment: Breast, Ovarian, and Pancreatic Version 1.2023 – September 7, 2022[19].

Other high penetrance OC genes include those associated with Lynch syndrome, also known as hereditary non-polyposis colorectal cancer, which predominantly features colorectal and endometrial cancers[23]. Lynch syndrome is characterized by autosomal dominant inheritance of germline pathogenic variants in the mismatch DNA repair pathway genes, *MLH1*[24,25], *MSH2*[26], *MSH6*[27], and *PMS2*[28]. Most Lynch syndrome families are accounted for by *MLH1* (15-40%) and *MSH2* (20-40%), and have an absolute risk for OC of 4-20% and 8-38%, respectively[19,29] (Figure 1.1). Pathogenic variants are less commonly identified in Lynch syndrome families in *MSH6* (12-35%) and *PMS2* (5-25%) and confer absolute risks for OC of ≤1-13% and 1.3-3%, respectively[19] (Figure 1.1). The frequency of all pathogenic variants in these four mismatch DNA repair genes is <1% in sporadic OC cases[30].

1.1.2 Moderate penetrance genes

As germline pathogenic variants in the high penetrance OC predisposing genes only account for approximately 55% of multi-case OC families[31], a candidate gene approach, focusing on genes in the same DNA repair pathway as BRCA1 and BRCA2 which will be discussed below, was employed to identify other OC risk genes. Three genes, *BRIP1*, *RAD51C*, and *RAD51D*, which function as members of the same DNA repair pathway that intersects the function of BRCA1 and BRCA2, were identified in OC families and are now established OC predisposing genes[32–37]. The absolute risk for carriers of pathogenic variants in *BRIP1*, *RAD51C*, and *RAD51D* is 5-15%, 10-15%, and 10-20%, respectively[19] (Figure 1.1), and therefore these genes are classified as moderate penetrance genes for OC.

PALB2 (partner and localizer of BRCA2), which interacts with BRCA2 in DNA repair, was first identified as a BC predisposing gene[38,39] in 2007 and was thought to exhibit little-to-no risk to OC. In the largest study to date of 524 *PALB2* pathogenic variant carrier families, the relative risk for OC was estimated to be 2.91 (95% confidence interval (CI) 2.84-9.65)[40] and the absolute risk is estimated at 3-5% for OC by NCCN® Guidelines[19] (Figure 1.1).

Individuals harbouring heterozygous pathogenic variants in *ATM* were initially associated with increased BC risk[41], though results were conflicting across other studies[42–44]. More recent studies have shown an association of variants in *ATM* with OC[45–48], though the absolute risk is estimated to be 2-3%[19] (Figure 1.1). The initial identification of variants in *PALB2* and *ATM* associated with BC led to the inclusion of

these genes on multi-gene panels for the testing of BC and OC cases. This likely aided in the subsequent establishment of an association with OC.

All identified moderate penetrance OC predisposing genes have been associated with an autosomal dominant pattern of inheritance. Pathogenic variants in moderate penetrance genes account for approximately another 5% of multi-case OC families[31].

1.1.3 Associated markers

Genome Wide Association Studies (GWAS) have identified more than 30 markers that are associated with OC risk[49,50]. These markers have odds ratios (OR) ranging from 1.09 to 2.19[49], which would be considered as low risk for disease and none have been validated as OC risk variants. As the main goal of GWAS is to identify associations across the genome, these variants have been correlated with individuals with disease, though they may not be causal. In addition, the markers identified may not be the drivers of the association as many are in intergenic or intronic regions of the genome and may be in linkage disequilibrium with a causal variant.

Polygenic risk scores have been developed for other cancers such as breast and colorectal, and more recently this has been applied to OC[51,52]. It is currently unclear how useful these scores will be in the clinic as even individuals in the top fifth percentile of scores only had an absolute risk for epithelial OC by age 80 of 2.9%[51], which is similar to other emerging OC risk genes which have not been applied clinically due to lack of information[19]. Thus, for this thesis I have chosen to focus on rare, protein coding variants in the genome that could be directly associated with epithelial OC risk.

1.1.4 Cancer phenotypes associated with OC predisposing genes

The pathological subtypes of OC have distinct etiological, genetic, and clinical attributes. Epithelial OC accounts for over 90% of all malignant OC tumours with the remainder being evenly distributed between tumours arising from stromal and germ cells[53]. Nearly all benign and malignant OC tumours arise from these three cell origins. Most often, pathogenic variants in OC risk genes are identified in epithelial ovarian cancers, rather than those of stromal or germ cell origin. Pathogenic variants in *BRCA1*, *BRCA2*, *BRIP1*, *RAD51C*, and *RAD51D* are mainly identified in high-grade serous ovarian carcinoma (HGSC), the most common epithelial OC histopathological subtype that accounts for 70% of epithelial OCs[54–56]. Pathogenic variants in the mismatch repair genes (*MLH1*, *MSH2*, *MSH6*, and *PMS2*) are identified mostly in endometrioid (1.65%) and clear cell (1.87%) subtypes of epithelial OC, compared to HGSC (0.35%)[30,57]. Pathogenic variants in any of these genes may also be identified in other epithelial OC subtypes, such as mucinous, though it is much less common (0-0.58% depending on the gene mutated)[30]. Due to the prevalence of HGSC and the fact that known CPGs have been identified in only 60% of multi-case OC families[31], I will mainly focus on this histopathological subtype of OC.

The median age at OC diagnosis in North America is 63 years[8]. This age at diagnosis has been estimated to vary if an individual harbours a pathogenic variant in a high risk OC predisposing gene, for example the median age at diagnosis for *BRCA1* pathogenic variant carriers is 53 years and 65 years for *BRIP1* pathogenic variant carriers[58]. As there is currently no effective screening method for OC[59], individuals harbouring pathogenic variants in *BRCA1*, *BRCA2*, *BRIP1*, *RAD51C*, or *RAD51D* are

offered risk reducing salpingo-oophorectomy[19], removal of the fallopian tubes and ovaries, which decreases risk significantly (up to 80%)[60,61]. For *PALB2* and *ATM,* there is insufficient evidence to recommend risk reducing surgery and carriers should be counseled based on family history[19].

Following a diagnosis of OC, the first line of treatment is primary debulking surgery followed by platinum-based chemotherapy, a clinical course which has been the standard of care since the 1980s[62]. There has been some evidence for use of neoadjuvant chemotherapy followed by interval debulking surgery, though the choice between this method and the standard of care is still debated[62]. First-line treatment with carboplatin or cisplatin and paclitaxel is the standard chemotherapy used following primary debulking surgery. These guidelines for first-line care in OC are largely based on HGSC cases as they encompass the largest subgroup of cases.

The two major OC predisposing genes, *BRCA1* and *BRCA2,* have been associated with an increased risk of other cancers such as the aforementioned BC (>60% absolute risk), pancreatic cancer (≤5% and 5-10%, respectively), prostate cancer (7-26% and 19-81%, respectively), and melanoma (elevated but unknown risk for *BRCA2* carriers only)[19]. This tissue tropism is also exhibited by other CPGs associated with cancer syndromes, such as the mismatch repair genes (hereditary non-polyposis colorectal cancer syndrome) that increase risk for cancers including colorectal, endometrioid, stomach, hepatobiliary, upper urologic tract, and ovarian[23]. It is not well understood why this tissue specificity exists but it has been posited that cell of origin and differentiation, tumour suppressor barriers that initiate cell death or senescence, chromatin organization and regulatory elements, single catastrophic genetic events, or

tolerance of DNA damage could influence this tissue specificity[63]. It is therefore possible that newly identified CPGs may also exhibit tissue tropism with an increased risk for more than one cancer type.

1.1.5 Somatic genetics of HGSC tumours

Approximately 40% of known CPGs have been shown to be mutated in tumours and play a role as drivers of tumourigenesis[64]. One such example is *TP53*, the most commonly mutated gene in human tumours, suggesting an integral role of aberration of this pathway in tumourigenesis[65]. Common alterations in cancer genomes may glean insight to molecular mechanisms driving cancer phenotypes and possible therapeutic targets[66,67].

The most common somatic alteration in HGSC tumours is pathogenic variants identified in *TP53*, which is observed in >95% of patients[68]. Other less prevalent variants have been identified in *CSMD3* (6%), *NF1* (4%), *BRCA1* (3%), *BRCA2* (3%), *CDK12* (3%), and *RB1* (2%) in HGSC tumours[68]. There is a characteristic 'long tail' of other variants that are found at very low frequencies and can be specific to each individual tumour[68,69]. Pathway analysis has shown that the FA-HR pathway is altered in over 50% of HGSC tumours[68]. HGSC is characterized by genome-wide copy number alterations (CNAs) with one of the most common focal amplifications observed in *CCNE1*[68]. Overall survival has been reported to be lower for patients with *CCNE1* amplification compared to patients without *CCNE1* amplification[70]. However, when stratifying patients based on *BRCA1* or *BRCA2* pathogenic variant status (germline or somatic) no survival disadvantage was identified for patients with *CCNE1* amplified tumours[68].

1.2 Why search for new OC predisposing genes?

Since the discovery of *BRCA1* and *BRCA2* in the mid-90's, it has been evident that not all OC families are accounted for by pathogenic variants in these two genes[71]. It was reported that 27% of families with two cases of OC and 54% of families with three or more cases of OC had an identifiable pathogenic variant in *BRCA1* or *BRCA2*[71]. In the largest study of OC families with two or more cases in first- or second-degree relatives the authors identified a pathogenic variant in *BRCA1* or *BRCA2* in 51% of cases[31]. Other established high and moderate penetrance OC predisposing genes (mismatch repair genes, *ATM*, *BRIP1*, *PALB2*, *RAD51C*, and *RAD51D*) accounted for another 9% of these OC families[31] (Figure 1.2). Other genes have been associated with OC risk (including variants identified from GWAS), though this risk has yet to be established, and each new gene is expected to account for a small proportion of OC cases. Thus, there remains a proportion of OC families not attributed to genetic factors that have been previously established, suggesting other OC predisposing genes are yet to be discovered.

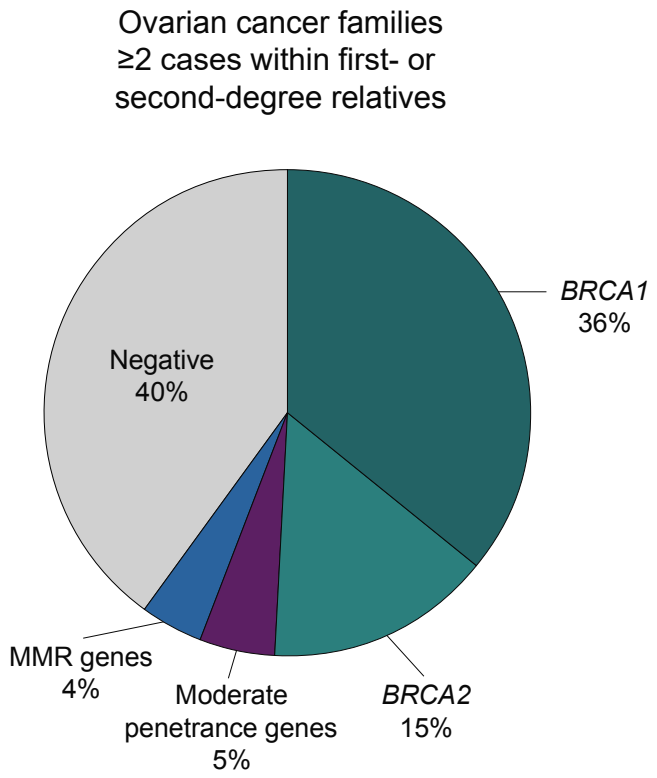


Figure 1.2. Frequency of pathogenic variants in moderate and high penetrance OC predisposing genes identified in OC families with ≥2 epithelial OC cases within first- or second-degree relatives (n=229 families).

Adapted from data presented in Flaum et al.[31]

1.2.1 Strategies to identify new OC predisposing genes

Several factors pose challenges to finding new OC predisposing genes including the low incidence of OC, rarity of pathogenic variants in each OC predisposing gene, and genetic and allelic heterogeneity of OC. Epidemiological factors, such as oral contraceptive pill use and reproduction, may also affect the prevalence of OC. It has been suggested that the projected 27% lower incidence of OC in Canada in 2021 compared to 1984 could be partially due to the increased use of oral contraceptive pills and changes in reproduction (first births at later age)[9].

Classical linkage analysis using informative genetic markers in large families with affected members was used to identify CPGs such as *RB1*, *TP53*, *BRCA1*, and *BRCA2*[64]. Logarithm of odds scores were used to identify candidate regions followed by meiotic recombination mapping to refine the location including the putative CPG[72]. Additional polymorphic markers up and downstream of the putative gene could then be used to map alleles in linkage until one was found that no longer segregated with disease[73]. Once a CPG was identified through this method, they were assessed by germline DNA sequencing from linked families and segregation of the same pathogenic variant with affected members within the same family[73]. This method was used at a time when the Human Genome Project was underway but not yet complete[74], and the biological function of many genes was unknown. Linkage analysis therefore allowed for the identification of associated markers with little knowledge of the human genome and function of the encoding proteins. Caveats to this approach include the need for large families with multiple affected members with available genetic material and informative genetic markers within the region of interest. Although, when large families with multiple

affected members are identified this approach becomes advantageous allowing for segregation analysis of variant alleles with disease to vet candidates[75]. These discoveries of early CPGs, such as *BRCA1* and *RB1*, led to investigations of the biological pathways in which these genes function, informing future studies aiming to identify new CPGs[64].

With the advent of massively parallel sequencing approaches, such as whole exome and whole genome sequencing (WGS), and with increasing biochemical information about protein pathways, candidate gene approaches have been used in the identification of new CPGs[64]. A candidate gene approach relies on understanding the mechanisms underlying the disease[76], such as a specific protein pathway suspected to be involved in tumourigenesis. This approach can also be useful in more common or complex diseases where the risk associated with each candidate is expected to be small[76]. A candidate gene approach led to the identification of the moderate penetrance OC predisposing genes *RAD51C*[34] and *RAD51D*[37]. The genes were investigated as candidates as they're part of the same molecular pathway as *BRCA1* and *BRCA2*[77], as will be detailed later in this chapter.

Whole exome sequencing (WES) involves the capture of regions that cover the human exome followed by high-throughput DNA sequencing of these target-enriched genomic regions[78]. Enriching for the coding regions of the human genome has been desirable as the regions are more interpretable compared to WGS[78], and had allowed for the identification of causal genes for Mendelian diseases[79]. Single nucleotide variants (SNVs) and small insertions and deletions can be determined from the DNA sequencing data from WES and further interpreted following variant annotation with

bioinformatic tools[78]. These bioinformatic tools include those using in silico analyses of variants to determine conservation of the locus or likelihood of affecting the encoded protein function. The tools used widely when I began my PhD project were Genomic Evolutionary Rate Profiling (GERP++)[80] for conservation of a locus and Sorting Intolerant from Tolerant (SIFT)[81] and Polymorphism Phenotyping v2 (PolyPhen2)[82]. Since then, many more tools have been developed to interpret SNVs, particularly missense variants. These tools have been evaluated for those with the ability to best predict pathogenic variants[83–86] and can be applied to assist in the interpretation of variants. These tools in combination with functional analyses to test the biological impact of candidate variants have become important in interpreting missense variants identified in the germline of cancer cases.

The study of families with multiple cases of cancer has been integral in the discovery of many CPGs as the clustering of cancers within families suggests a heritable component to the disease. As the heritable component of common cancers is estimated at 33% from twin studies but only 3% of cancers can be attributed to CPGs, there appears to be a large number of genetic factors yet to be determined[12,64]. Variants in the above-mentioned high and moderate penetrance OC predisposing genes account for approximately 60% of OC families defined as families with two or more OC cases within first- or second-degree relatives[31]. It is thus plausible that additional new risk genes may be identified in families that do not harbour pathogenic variants in any of the known or proposed OC predisposing genes. An obstacle to these discoveries is the rarity of such families, as it has been estimated that in the general population only 5% of individuals report a family history of two first-degree relatives with OC[87]. A combination

of strategies can therefore be applied to increase the possibility of identifying new OC predisposing genes.

1.2.2 French Canadian population of Quebec, Canada

Worldwide, there are populations that have undergone genetic drift after a population bottleneck, such as the Ashkenazi Jewish[88], Finnish[89], Icelandic[90], and French Canadian (FC) of Quebec[91] populations. The unique genetic demography of each of these populations can be attributed to endogamy or isolation by distance[88,92–95].

The FC population of Quebec has been referred to as a founder population due to shared ancestry from an estimated 8,483 founders who contributed to the contemporary population[96,97], though there is no evidence of loss of genetic diversity[91]. Approximately 10,000 settlers arrived from France to colonize Quebec (originally called Nouvelle France) over 400 years ago in 1600[73]. Over many generations there has been a propagation of rare alleles in this population due to large family size and geographic isolation. Examples of genetic diseases that are more prevalent in the FC population due to its complex genetic demography include a type of Leigh syndrome (MIM 220111), Tay-Sachs disease (MIM 272800), oculopharyngeal muscular dystrophy (MIM 164300), and β-thalassemia (MIM 141900)[97]. For *BRCA1* and *BRCA2,* a limited number of variants have been identified in FC OC cases and few frequently occur in the FC population[73] (Figure 1.3), in comparison to the over 7,000 variants that have been identified worldwide in these genes[98]. Similarly, five *RAD51C* and *RAD51D* variants have been identified in OC cases of FC ancestry[99], compared to over 200 variants worldwide[98]. The unique genetic demography of the FC population of Quebec thus provides an opportunity to identify rare disease-causing variants that may be more

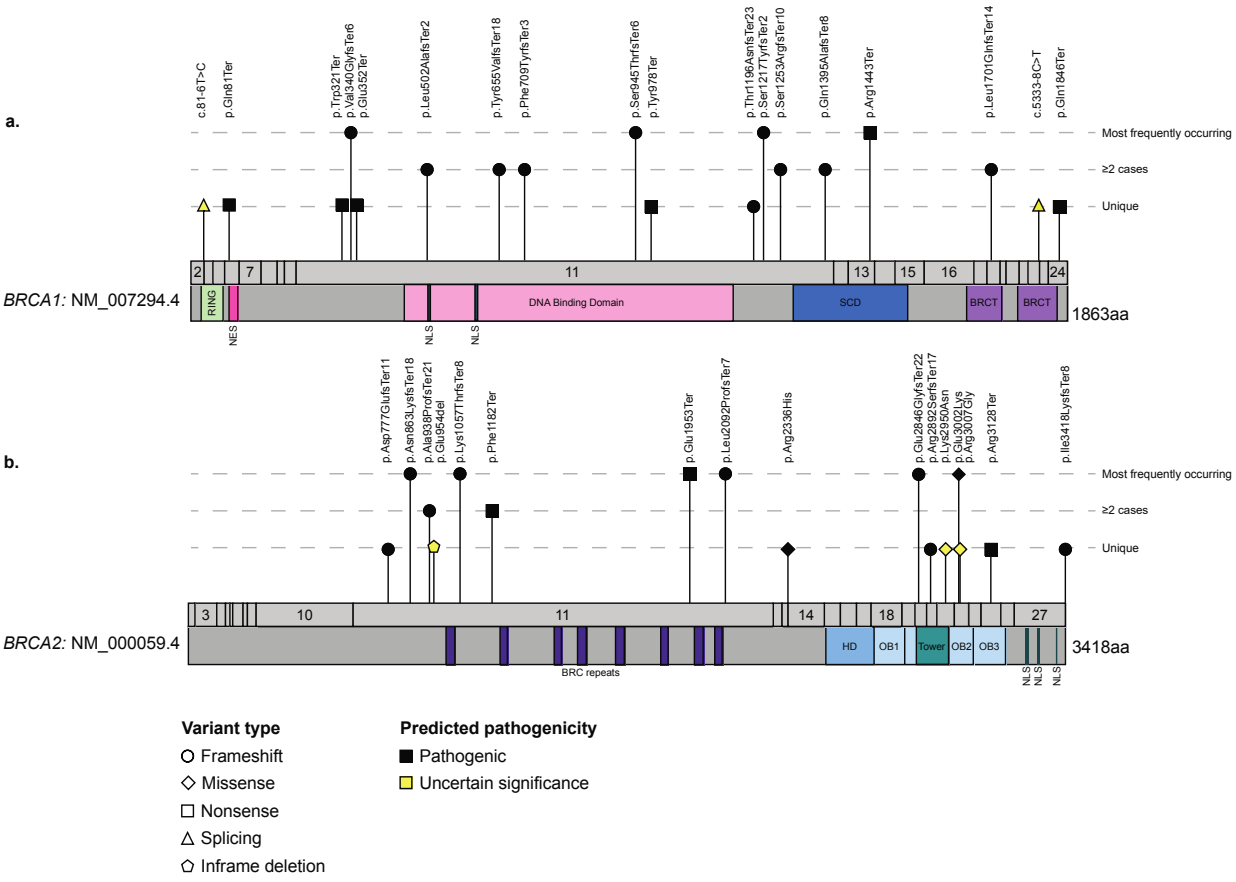


Figure 1.3. Pathogenic variants and variants of uncertain significance reported in French Canadians of Quebec mapped to full length *BRCA1* (a) or *BRCA2* (b) transcripts. Variants are predicted to be pathogenic or have uncertain significance based on ClinVar and/or American College of Medical Genetics and Genomics (ACMG) guidelines. RING = Really Interesting New Gene domain; NES = Nuclear export signal; NLS = Nuclear localization signal (*BRCA1*[100] and *BRCA2*[101]); SCD = Serine cluster domain[102]; BRCT = BRCA1 C Terminus domain; BRC repeats = BRCA2 repeats; HD = Helical domain; OB = Oligonucleotide binding; Tower = Domain essential for DNA binding[103]. BRCA1 GenBank: AAC37594.1[104], BRCA2 GenBank: AAB07223.1[105], DNA binding domain[106].

Adapted from Fierheller et al.[73]

prevalent in this population. The above-mentioned strategies employing a candidate gene approach in highly selected families from the FC population of Quebec has been successful in facilitating the characterization of OC predisposing genes such as *BRCA1*, *BRCA2*, *RAD51C*, and *RAD51D*[73,99].

1.3 Biology of OC predisposing genes

BRCA1 and BRCA2 proteins function in the homologous recombination (HR) DNA repair pathway that repairs double-stranded DNA breaks[107]. This pathway is largely active during S and G2 phases of the cell cycle[108] and relies on the sister chromatid for high fidelity repair[109]. The moderate penetrance OC predisposing genes, *BRIP1*, *PALB2*, *RAD51C*, and *RAD51D*, also function in this HR DNA repair pathway. The mismatch DNA repair genes (*MLH1*, *MSH2*, *MSH6*, and *PMS2*) function in the repair of spontaneously mispaired bases and small insertion-deletion loops generated during replication via the mismatch DNA repair pathway[110]. Similarly, other genes implicated in OC risk, such as *ATM* discussed above, are involved in DNA repair pathways and/or genome integrity/stability. It has also been shown that approximately 20% of HGSC cases have identifiable germline variants in genes that encode proteins involved in the HR pathway and in the associated upstream pathway, the Fanconi anemia (FA) pathway[69].

Abrogation of the FA-HR pathway exhibits synthetic lethality with poly(ADP)-ribose polymerases (PARP) which function in the choice between HR and the non-homologous end-joining pathway, the other main pathway in the repair of DNA double-stranded breaks[109]. PARP inhibitors (PARPi) are now clinically used to treat individuals diagnosed with HGSC who harbour *BRCA1* or *BRCA2* pathogenic germline or somatic

variants. This targeted therapy has been a breakthrough for the treatment of HGSC for which first-line chemotherapy, platinum and taxol, has not changed in over 20 years. It has been shown that some individuals who do not have identifiable *BRCA1* or *BRCA2* pathogenic germline or somatic variants also respond to PARPi[111]. There has been considerable effort in the research community to determine why, and it is suggested that any abrogation of the FA-HR pathway could lead to synthetic lethality upon treatment with PARPi[111]. Combinations of therapies targeting multiple pathways, such as PARPi with immunomodulators or anti-angiogenesis therapies, may help elucidate mechanisms underlying response to PARPi in individuals with and without pathogenic variants in *BRCA1* or *BRCA2*[112].

1.4 Fanconi anemia pathway

Interstrand crosslinks (ICLs), which are highly toxic DNA lesions that prevent transcription and replication, are repaired by the FA pathway[113,114]. ICLs may be caused by endogenous (e.g., aldehydes) or exogenous (e.g., mitomycin C) agents, and occur in each cell approximately 10 times a day[113–115]. There are three key groups that make up the FA pathway: (1) the FA core complex; (2) the ID2 complex; and (3) downstream proteins involved in HR (Table 1.1). The FA core complex consists of FANCA, -B, -C, -E, -F, -G, -L, -M, and -T and the FA associated genes FAAP20, FAAP24, and FAAP100. The ID2 complex is a heterodimer of FANCI and FANCD2, which are both monoubiquitinated by the FA core complex. Downstream proteins, including those encoding OC risk genes (*FANCS [BRCA1]*, *-D1 [BRCA2]*, *-J [BRIP1]*, *-N* [*PALB2*], and *-O [RAD51C]*), are involved in the resolution of the double-stranded DNA break that occurs after excision of the ICL.

1.5 Fanconi anemia

In 1927, the Swiss physician-scientist Guido Fanconi first described FA in a family with severe bone marrow failure[116]. FA is a rare (approximately 1 in 100,000) heterogenous disease associated with bone marrow failure, congenital abnormalities, and cancer risk[117]. There have been 22 genes associated with FA with approximately 5% of FA patients remaining unexplained by pathogenic variants in these genes (Table 1.1). This gene family is characterized by participation in the same pathway, the FA-HR DNA repair pathway[117]. In contrast, other gene families are grouped based on the amino acid similarity of the encoding proteins, homologous genes with similar function[118]. Complementation of patient-derived cells were used to determine if FA patients had the same or different gene involved. These genes are designated as Fanconi anemia Complementation Group A-W and given the gene symbol *FANCA-W*, with the exclusion of H (as it was later determined to be the same as complementation group A) and K (as *FANCA* and *FANCK* were deemed to sound too similar). Of the 22 genes associated with FA, 20 are associated with an autosomal recessive mode of inheritance[117]. *FANCB* is located on the X-chromosome and is therefore associated with X-linked inheritance; a dominant-negative de novo pathogenic variant was identified in the only reported *FANCR* (*RAD51*) patient[119]. Most (approximately 80%) of FA patients have a pathogenic variant in *FANCA*, -C, and -G; there is clinical heterogeneity depending on the gene involved[117]. Approximately one third of FA patients do not present with any visible phenotype and therefore it is recommended that all individuals with persistent pancytopenia be tested for FA[117,120]. In clinic, a diagnosis of FA is confirmed by using a chromosome breakage test or identification of causal variant(s) in

Table 1.1 Genes associated with Fanconi anemia and molecular function in the FA-HR pathway.

Gene	Alias	Molecular function	Frequency in FA	Genomic location
FANCA	–	FA core complex	60-70%	16q24.3
FANCB[1]	–	FA core complex	2%	Xp22.2
FANCC	–	FA core complex	14%	9q22.23
FANCD1	*BRCA2*	DNA repair control and effector recruitment	2%	13q13.1
FANCD2	–	ID2 complex	3%	3p25.3
FANCE	–	FA core complex	3%	6p21.31
FANCF	–	FA core complex	2%	11p14.3
FANCG	*XRCC9*	FA core complex	10%	9p13.3
FANCI	–	ID2 complex	1%	15q26.1
FANCJ	*BRIP1*	DNA helicase	2%	17q23.2
FANCL	–	E3 ubiquitin ligase	<1%	2p16.1
FANCM	–	FA core complex	<1%	14p21.2
FANCN	*PALB2*	Regulates BRCA1	<1%	16p12.2
FANCO	*RAD51C*	DNA repair via HR	<1%	17q22
FANCP	*SLX4*	Interacts with several nucleases	<1%	16p13.3
FANCQ	*ERCC4/XPF*	DNA repair nuclease	<1%	16p13.12
FANCR[2]	*RAD51*	DNA repair via HR	2 reported	15q15.1
FANCS	*BRCA1*	DNA repair via HR	<1%	17q21.31
FANCT	*UBE2T*	E2 ubiquitin-conjugating enzyme	<1%	1q32.1
FANCU	*XRCC2*	DNA repair via HR	1 reported	7q36.1
FANCV	*REV7/MAD2L2*	Translesion DNA synthesis	1 reported	1p36.22
FANCW	*RFWD3*	E3 ubiquitin ligase	1 reported	16q23.1
Unknown	-	-	<5%	-

[1]Associated with X-linked pattern of inheritance
[2]Dominant-negative
FA: Fanconi anemia; HR: homologous recombination; ID2: FANCI and FANCD2 heterodimeric complex

the known FA genes by DNA sequencing[117]. Some of the FA genes are associated with VACTERL-H (Vertebral abnormalities, Anal atresia, Cardiac defects, Tracheoesophageal fistula, Esophageal atreasia, Renal and radial abnormalities, Limb abnormalities with Hydrocephalus)[121], a rare genetic disorder characterized by evidence of at least three of the eight anatomical components (MIM 276950).

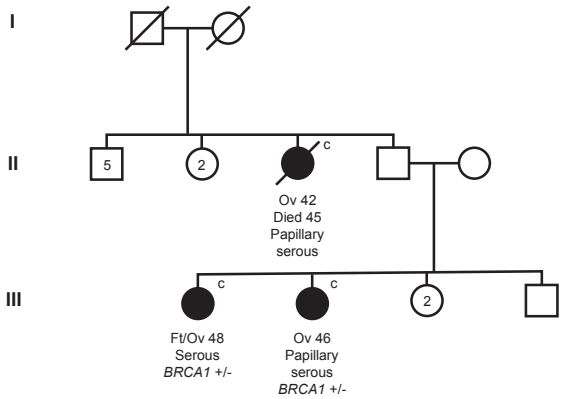
In the context of FA, OC predisposition genes *FANCD1 (BRCA2)* and *FANCN* (*PALB2*) are associated with specific cancer phenotypes[122]. FA patients with pathogenic variants in these two genes were reported to have acute leukemia as is common in FA, however, these patients were almost exclusively diagnosed with Wilms tumour, brain tumour, or neuroblastoma[122].

1.6 Preliminary analyses suggesting *FANCI* as an OC risk candidate

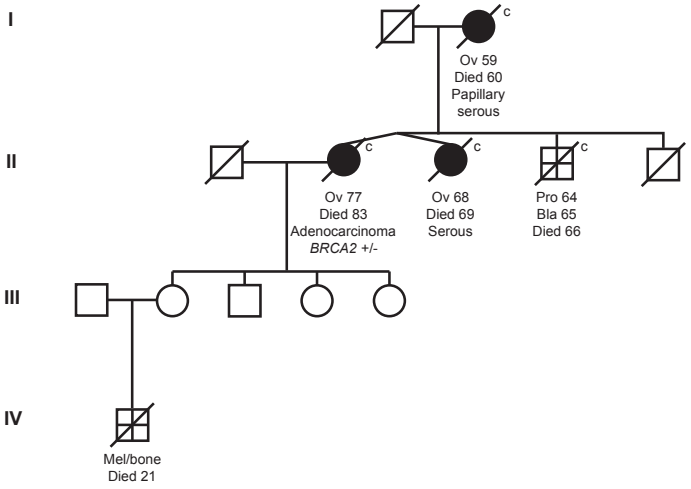
In 2007, three families of FC ancestry with three or more cases of OC in first-, second-, or third-degree relatives were reported by Dr. Patricia Tonin's group[123]. One family had a pathogenic *BRCA1* variant, one family had a pathogenic *BRCA2* variant, and the third family was negative for pathogenic variants in both *BRCA1* and *BRCA2* (Figure 1.4). When WES became available, this family (F1528) was sequenced to determine whether a hereditary factor increasing OC risk could be identified in the exons of the genome.

WES and bioinformatic analyses of family F1528 identified 276 rare (minor allele frequency [MAF]<1%) variants shared between the two sisters diagnosed with OC. Under the assumption that OC is a dominantly inherited trait, variants shared in a homozygous state were excluded as candidate risk variants. At the time of analyses a candidate gene approach was employed, investigating variants in DNA repair pathway

Family 1496



Family 1497



Family 1528

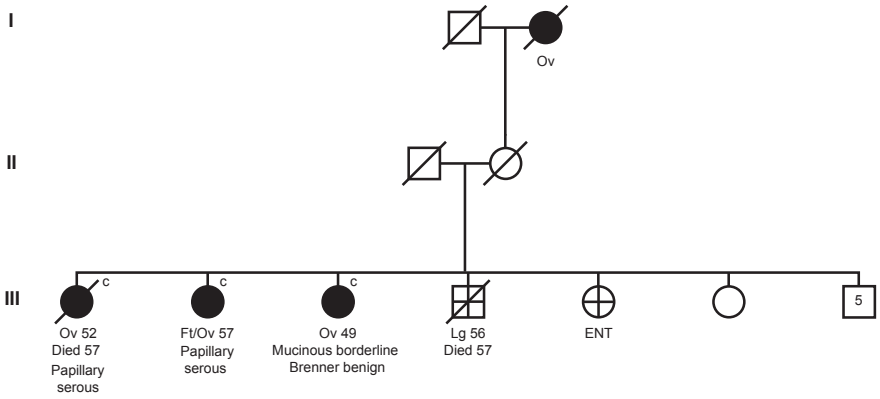


Figure 1.4. Pedigrees of French Canadian hereditary ovarian cancer families. Cancer type [ovarian (Ov), fallopian tube (Ft), prostate (Pro), bladder (Bla), melanoma (Mel), and lung (Lu)] and age of diagnosis or age of death are shown along with *BRCA1* and *BRCA2* pathogenic variant carrier status; c next to a symbol denotes a confirmed cancer case.

Adapted from Tonin et al.[123] and Fierheller et al.[124].

genes, especially those involved in the FA-HR pathway, as approximately 20% of sporadic OC cases were identified to have a germline potentially pathogenic variant in one of these genes[69] and *BRCA1* and *BRCA2* are involved in this pathway. The only candidate variant identified in FA-HR pathway genes based on this candidate gene approach was *FANCI* c.1813C>T; p.L605F in family F1528.

Preliminary in silico analyses of this variant suggested *FANCI* c.1813C>T is located at a highly conserved locus (GERP++) and the amino acid change of leucine to phenylalanine was predicted to affect protein function (SIFT and PolyPhen2). Loci that are highly conserved across species are considered to be important residues as conservation across divergent species implies functional constraint[125]. This is exemplified by the conservation of FANCI, FANCD2, and FANCL proteins across eukaryotes[126], though the core complex members are found only in vertebrates[127]. Additionally, variants that are predicted to affect protein function where a loss of protein function occurs could be encoded by tumour suppressor genes[128].

Initial data showed that *FANCI* c.1813C>T was more common in familial (1/7, 14.3%) compared to sporadic OC cases (7/273, 2.5%). If a variant is more common in familial cancer cases compared to sporadic cancer cases or cancer-free controls, this supports a role in conferring increased risk to cancer[73].

Population genetic databases indicated that the allele frequency of *FANCI* c.1813C>T was higher than expected (0.2-0.76%) compared to individual pathogenic variants in OC predisposing genes such as *BRCA1* and *BRCA2* (0.001%)[129,130]. The rare variant hypothesis suggests that rare (MAF<1% in the population) variants are more likely to be risk alleles in diseases[131]. Additionally, the precedence for variants that

are more common than expected in the general population and that are nonetheless involved in cancer disease risk is exemplified by *CHEK2* c.1100del; p.Thr367MetfsTer15 in BC. This *CHEK2* variant has an allele frequency of 0.2% (range 0-0.9%) in population database Genome Aggregation Database (gnomAD) non-cancer controls, an allele frequency that is similar to the *FANCI* c.1813C>T allele (0.7%, range 0-1.8%)[132]. A case-control analyses of this *CHEK2* variant did not initially identify an association with BC risk[133]. However, stratification of cases into those with or without a family history of cancer found that the prevalence of *CHEK2* c.1100del in individuals with a strong family history of BC was significantly higher compared to controls[133]. This example illustrates that variants more common than expected in the general population may be involved in disease risk.

Preliminary functional analysis of the protein was therefore performed to determine the effect that p.L605F may have on FANCI protein function, which revealed that *FANCI* c.1813C>T encodes a protein that does not function properly. Similar to the above in silico prediction, the consequence of loss of protein function for a tumour suppressor gene could promote tumourigenesis.

1.7 *FANCI*

Fanconi anemia complementation group I was established by somatic cell hybridization analysis[134]; *KIAA1794* was identified as the *FANCI* gene in 2007 by three different groups[135–137] in the pursuit of the gene associated with the FA complementation group I. *FANCI* was identified using a *FANCD2* protein sequence-based homology approach[135,137] and genome-wide linkage analyses[136]. *FANCI* is located in chromosome region 15q26.1. These initial reports showed that *FANCI*:

homozygous or compound heterozygous variants were identified in FA complementation group I patients; is ubiquitinated, likely by the FA core complex; protein formed a complex with FANCD2; protein participates in DNA repair pathways and localized to sites of DNA damage; and patient cells are deficient in ubiquitinated FANCD2. Though there are four known and 47 proposed transcripts for *FANCI* in *Homo sapiens,* the canonical transcript (NM_001113378.2 genome assembly GRCh37/hg19) contains 38 exons and encodes a protein that has 1,328 amino acids (Figure 1.5).

FANCI has five protein domains: leucine zipper[138], SUMO-like domain-interacting motif (SLIM)[139], Armadillo repeat (ARM)[137,140], EDGE (consists of Glutamic acid [E] - Aspartic acid [D] - Glycine [G] - Glutamic acid [E] amino acids)[141], and nuclear localization signal (NLS)[141] (Figure 1.5). FANCI is an integral member of the FA-HR pathway and acts as the molecular switch to turn on this pathway[142], such that FANCI is required for the activation of the FA core complex when an endogenous or exogenous agent creates a DNA interstrand crosslink[143,144] (Figure 1.6). The FA core complex is responsible for the ubiquitination of FANCI and its heterodimeric binding partner FANCD2, which form the ID2 complex and are the only known ubiquitination targets of the FA core complex[145]. The ubiquitination of this ID2 complex is necessary for downstream pathway function and the phosphorylation of FANCI, by ATR, is required for this process[142,146–149]. The monoubiquitination of the ID2 complex is integral to the FA pathway as most FA patients have lost the ability to carry out this biochemical process. Interestingly, MSH2, a known OC predisposing gene, is required for ubiquitination and chromatin loading of FANCI and FANCD2[150]. Stability and protection

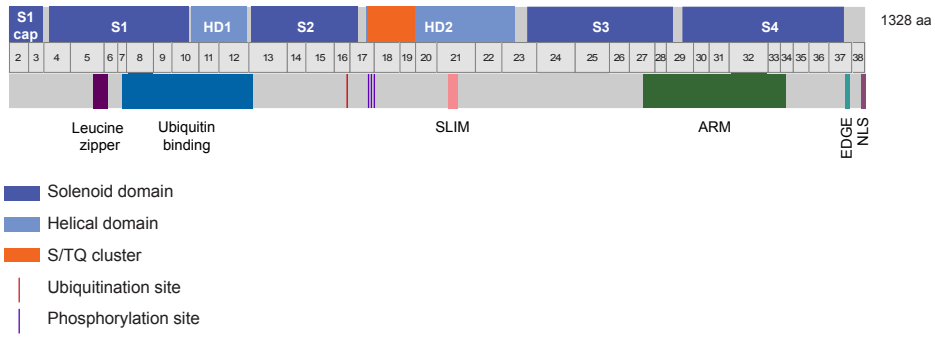

S1 cap
S1
HD1
S2
HD2
S3
S4
1328 aa
2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38
Leucine zipper
Ubiquitin binding
SLIM
ARM
EDGE
NLS
Solenoid domain
Helical domain
S/TQ cluster
Ubiquitination site
Phosphorylation site

Figure 1.5. Structure of FANCI protein.

FANCI domains were adapted from pfam (https://pfam.xfam.org). *FANCI* exon locations adapted from University of California Santa Cruz Genome Browser (https://genome.ucsc.edu). Solenoid domain: antiparallel pairs of α-helices that form α-α superhelix segments; Helical domain: α-helices; Ubiquitination site: site of monoubiquitination by the FA core complex to allow downstream FA pathway function, located at K523[152,153]; S/TQ cluster: location of conserved phosphorylation sites[154]. Phosphorylation sites (556, 559, and 565aa): sites of phosphorylation that stabilize the association of FANCI with DNA and FANCD2[148]. Leucine zipper (130-151aa): may be related to protein-protein interactions, DNA binding, or RNA binding, but the leucine zipper found at the N-terminus of FANCI has been shown not to bind to DNA[138]. Ubiquitin binding (175-377aa): this region binds to the ubiquitin on FANCD2[155]. SUMO-like domain-interacting motif (SLIM; 682-696aa): binds to the SUMO-like domain 2 (SLD2) of UAF1 promoting FANCD2 deubiqutination which is required for FA pathway function[139]. Armadillo repeat (ARM; 985-1207aa): forms a super helix of helices, which can also be found in FANCD2[153]. EDGE motif (1300-1303aa): this motif consists of Glutamic acid (E) - Aspartic acid (D) - Glycine (G) - Glutamic acid (E) and is required for DNA crosslink repair function[141,152,153]. Nuclear localization site (NLS; 1323-1328aa): required for localization to the nucleus where subsequent function in the FA pathway can occur[141].

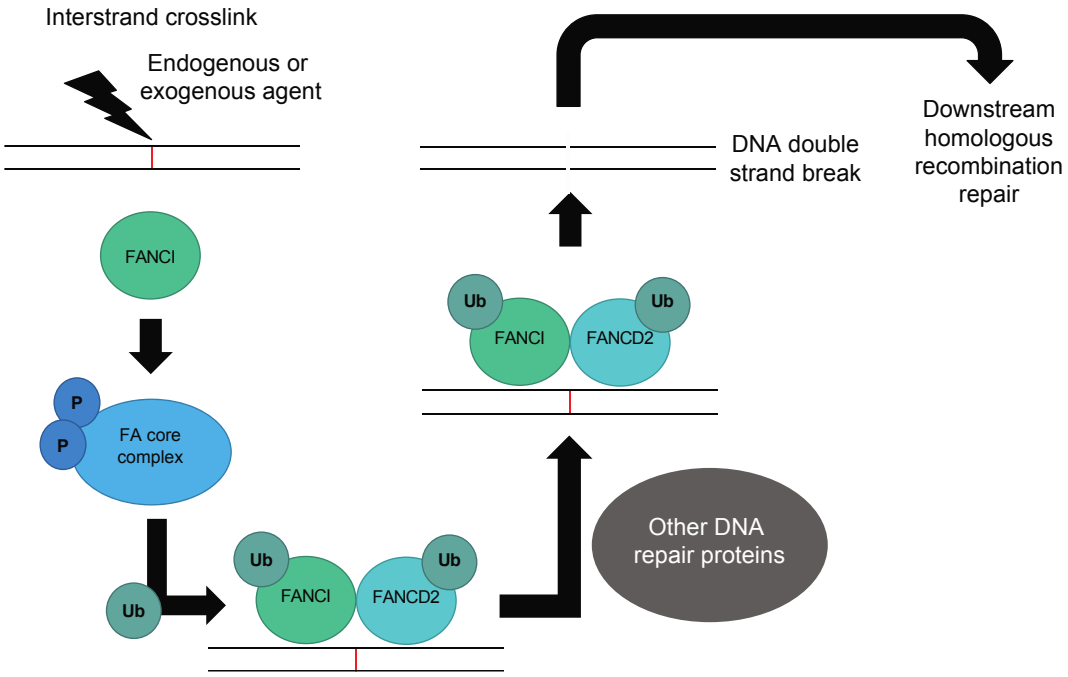


Figure 1.6. Function of FANCI protein in the Fanconi anemia pathway upstream of the homologous recombination DNA repair pathway. After an interstrand crosslink is created by an endogenous or exogenous DNA damaging agent FANCI activates the Fanconi anemia core complex. This complex then ubiquitinates both FANCI and its heterodimeric binding partner FANCD2. Other DNA repair proteins are recruited through a largely unknown process and a double stranded DNA break is created and repaired by downstream homologous recombination pathway proteins.

FA: Fanconi anemia; Ub: ubiquitin; P: phosphorylation

from degradation of the ID2 complex is also dependent on another OC predisposing gene, *BRIP1* (*FANCJ*)[151].

The ubiquitinated ID2 complex localized to the DNA interstrand crosslink site was initially thought to lead to the recruitment of other proteins involved in HR, translesion synthesis, and nucleotide excision repair[156]. It has since been shown that ubiquitination of the ID2 complex leads to a conformational change that clamps the complex onto double-stranded DNA[155,157]. This ubiquitinated ID2 complex clamped to DNA does not exhibit any direct protein-protein interactions, but leads to filamentous array formation on double-stranded DNA[157]. FANCI is also SUMOylated during this process[139] and dosage of the ID2 complex at sites of DNA damage is related to the SUMOylation that occurs on the proteins[158]. SUMOylated FANCI binds to the SUMO-like domain 2 (SLD2) of UAF1 promoting FANCD2 de-ubiquitination which is required for FA pathway function[139]. The ID2 complex interacts throughout the cell cycle and in the absence of ATR or the core complex, suggesting that this interaction is between the non-ubiquitinated forms of FANCI and FANCD2[159]. FANCI is able to bind DNA in both the ubiquitinated and non-ubiquitinated form[160], which occurs early in S-phase and remains bound for the duration of this phase of the cell cycle[161]. FANCI is also able to homodimerize leading to stabilization of the protein through a process mediated by UBL5[162], which may account for the increased levels of FANCI protein compared to that of FANCD2 observed in cells. The ID2 complex exhibits roles outside the FA pathway such as regulation of splicing factors[163], histone chaperone binding at Holliday junctions[164], binding to R-loops[165], and protection of stalled replication forks[166]. FANCI

also has functions independent of the ID2 complex such as dormant origin firing[167], regulation of Akt signaling[168], and ribosome biogenesis[169].

FA associated with pathogenic variants in *FANCI* is rare and comprises approximately 1% of FA cases[117] (Table 1.1). There are relatively few cases described in the literature (n=29 cases from 27 families)[134–136,170–177] and those with clinical phenotype information available present with young age of onset bone marrow failure (average age = 6.1 years, range 0.5-15) and VACTERL-H association (7/12, 58%) (Table 1.2). No cases of FA complementation group I have been reported to be diagnosed with any cancer. The spectrum of variants identified in this group is broad, where biallelic missense variants may also be disease causing, which is exemplified by four different cases (Table 1.3). Missense variants were also identified in the context of canonical (n=1) or non-canonical (n=3) splice site variants. A canonical splice site variant was identified in a case with a secondary synonymous variant predicted to affect skipping of exon 4 (n=1)[171]. Loss-of-function variants were identified but the second identified variant was either missense (n=7), non-canonical splice site (n=4), loss-of-function variant within the last 100 amino acids of the encoded protein (n=5), or both loss-of-function variants were within the last 100 amino acids of the encoded protein (n=3). Finally, one case was identified with a homozygous start loss variant. Since it has been documented that missense variants may exhibit residual protein function; splice variants may be leaky, leading to production of both wild type and altered mRNAs; and variants within the last 100 amino acids of a protein, especially those within the last exon, can exhibit hypomorphic effects with some protein function maintained[73,178–180] it is plausible that FA complementation group I cases exhibit some residual function of

FANCI. Complete gene knockout of *FANCI* may not be viable, accounting for the breadth of clinical phenotypes observed and rarity of cases. All variants are inferred to be in trans, and they have been confirmed in nine (Table 1.3).

This *FANCI* c.1813C>T; p.L605F variant is a plausible candidate to pursue because of:

1. The initial observation of an increased carrier frequency in OC families compared to sporadic OC cases;
2. Its known function as an integral member of the FA-HR pathway;
3. Preliminary data suggesting abrogated protein function;
4. Conservation of the protein and loci across organisms; and
5. In silico data predicting this variant would affect protein function.

Table 1.2 Clinical features of *FANCI* associated Fanconi anemia cases.

Patient ID	Sex	Origin	VACTERL-H association (at least 3 features)	Bone marrow failure onset (years)	Age (years)	Vital status
EUFA592	F	Turkey	Yes	2.5	6.5	Deceased
BD952[1]	F	India	NR	7.3	23	Deceased
1428	M	NR	NR	7.3	15	Deceased
EUFA816[2]	M	Hungary	NR	6	12	Deceased
EUFA480	M	NR	NR	4.8	24	Alive
EUFA961	M	Austria	Yes	8	21	Deceased
EUFA1399	M	Germany	Yes	1	30.5	Alive
F010095	NR	NR	No	6	21	Deceased
F010191	NR	NR	No	NR	31	Alive
IFAR663	NR	NR	Yes	6	10	Alive
NCI-82-I	F	NR	Yes	10.5	11.5	Deceased
NCI-253-I	F	NR	Yes	5	6	Alive
NCI-309-I	F	NR	Yes	NR	9	Alive
FA-1201	NR	Finland	No	No	6	Alive
Case 96	M	Japan	Yes	0.5	4	Deceased
Case 97	M	Japan	No	15	17	Alive
IN066[3]	M	Indian	NR	NR	8	Alive
IN138[3]	M	Indian	NR	NR	7	Alive
FA14	F	Europe	NR	NR	NR	NR
EUFA695	M	USA	NR	NR	12	Deceased
20074	M	NR	NR	NR	NR	NR
2480	F	NR	NR	NR	NR	NR
6	F	NR	NR	NR	15	Alive
Unc3	NR	NR	NR	NR	NR	NR
Sam11	NR	NR	NR	NR	NR	NR
20070	NR	NR	NR	NR	NR	NR
20071	NR	NR	NR	NR	NR	NR
20071	NR	NR	NR	NR	NR	NR
20072	NR	NR	NR	NR	NR	NR

[1]Sibling of 1428; [2]Sibling of EUFA480; [3]One of these cases is reported to exhibit VACTERL-H association though it is unknown which; F: female; M: male; NR: not reported

Table 1.3 Variants identified in *FANCI* associated Fanconi anemia cases.

Patient ID	Variant 1	Variant 2	Source
EUFA592	c.2T>C; p.?[1]	c.2T>C; p.?[2]	134–136,170
BD952	c.3854G>A; p.R1285Q[1]	c.3854G>A; p.R1285Q[2]	134,136,170
1428	c.3854G>A; p.R1285Q[1]	c.3854G>A; p.R1285Q[2]	136,170
EUFA816	c.3853C>T; p.R1285X[1]	c.3437_3455del; H1146LfsX12[2]	134,136,170
EUFA480	c.3853C>T; p.R1285X[1]	c.3437_3455del; H1146LfsX12[2]	136,170
EUFA961	c.3437_3455del; H1146LfsX12[1]	c.2572C>T; p.H858Y[2]	134,136,170
EUFA1399	c.3895C>T; p.R1299X	c.3895C>T; p.R1299X	136,170
F010095	c.2572C>T; p.H858Y	c.3437_3455del; H1146LfsX12	135,170
F010191	c.1840C>T; p.R614X	c.3895C>T; p.R1299X	135,170
IFAR663	c.2509G>T; p.E837X	c.3901dup; D1301GfsX3	135,170
NCI-82-I	c.3801_3804del; p.S1268RfsX5	c.3901dup; D1301GfsX3	170
NCI-253-I	c.1461T>A; p.Y487X	c.3041G>A; p.C1014Y	170
NCI-309-I	c.1039T>C; p.S347P	c.1202del; p.G401EfsX35	170
FA-1201	c.3041G>A; p.C1014Y	c.2957_2969del; V986AfsX39	173
Case 96	c.158-2A>G	c.288G>A (skip exon 4)	171
Case 97	c.3346dup; S1116FfsX16	c.3006+3A>G	171
IN066	c.1813C>T; p.L605F	c.1813C>T; p.L605F	172
IN138	c.295del; H99IfsX10	c.3907G>T; p.E1303X	172
FA14	c.1264G>A; p.G422R[1]	c.1583+142C>T[2]	170,174
EUFA695	c.3006+3A>G	c.1264G>C; p.G422R	136,170
20074	c.1461T>A; p.Y487X	c.3058+4A>G	181
2480	c.157+78G>A[1]	c.3493del; p.D1165TfsX34[2]	175
6	c.3853C>G; p.R1285X [1]	c.3853C>G; p.R1285X [2]	170,176
Unc3	c.3041G>A; p.C1014Y	c.3041G>A; p.C1014Y	177
Sam11	c.2509G>T; p.E837X	Deletion exon 38	177
20070	c.3604G>C; p.G1202R[1]	c.3508+1G>A	181
20071	c.756-25_756-19del[1]	c.3521C>T; p.T1174I[2]	181
20071	c.3622_3623del; L1208VfsX11[1]	c.866T>C; p.L289P	181
20072	c.2248T>C; p.C750G	c.2509G>T; p.E837X	181

[1]Confirmed maternal variant
[2]Confirmed paternal variant

1.8 Rationale, hypothesis, and objectives

The major OC predisposing genes, *BRCA1*, and *BRCA2*, were discovered almost 30 years ago and newer OC predisposing genes discovered since then have only accounted for a small proportion of the remaining OC families (approximately 9%)[31]. This lack of success in discovering new OC predisposing genes can be attributed to a combination of very rare variants with high penetrance, uncommon variants with low to moderate penetrance, and common variants with low penetrance. Many studies have focused on very rare variants (MAF<0.1%) or common variants (MAF>2%), though these have still accounted for only a small proportion (approximately 1% for each new gene identified) of the remaining families. Studies of rare variants have mostly restricted analyses to loss-of-function variants, excluding missense variants as their potential to affect risk is more difficult to interpret. This is exemplified by the fact that over 80% of variants of unknown significance in *BRCA1* and *BRCA2* are missense variants in contrast to the less than 5% of pathogenic variants being missense variants[182,183]. However, it is known that missense variants can increase risk to OC, such as with *BRCA2* c.9004G>A; p.E3002K[184] and *RAD51D* c.620C>T; p.S207L[185].

Using the strategy of investigating families with multiple cases of OC from a genetically unique population combined with a candidate gene approach, we identified *FANCI* c.1813C>T; p.L605F. Based on the preliminary analysis and function of FANCI upstream in the FA-HR DNA repair pathway, I hypothesized that *FANCI* c.1813C>T is likely a pathogenic variant that modifies risk to OC. To test this hypothesis, I proposed three main objectives.

The objectives were to:

1. Investigate the candidacy of *FANCI* as a new OC predisposing gene by:
    a. Comparing the carrier frequency of *FANCI* c.1813C>T in defined study groups comprised of FC familial and sporadic OC cases and cancer-free controls;
    b. Determining the effect FANCI p.L605F had on protein function;
    c. Surveying the frequency of *FANCI* c.1813C>T and other potentially pathogenic *FANCI* variants in non-FC populations; and
    d. Examining the protein expression of FANCI in normal fallopian tube epithelium and HGSC tissue.
2. Investigate the exomic landscape of *FANCI* c.1813C>T OC carriers:
    a. On the germline level to identify other variants that co-occur globally; and
    b. On the somatic level to determine if commonly mutated genes, CNAs, and somatic mutational signatures, are akin to those observed in HGSC tumours.
3. Characterize *FANCI* as a cancer predisposing gene by identifying carriers of germline *FANCI* c.1813C>T diagnosed with different cancer types and identifying somatic *FANCI* variants across different cancer types.

Chapter 2.0 addresses objective 1 where I identified *FANCI* c.1813C>T; p.L605F in OC cases of FC ancestry as a new candidate OC predisposing gene. Chapter 3.0 addresses objectives 2 and 3 where I characterized aspects of *FANCI* as a new candidate CPG.

2.0 Chapter 2: A functionally impaired missense variant identified in French Canadian families implicates *FANCI* as a candidate ovarian cancer-predisposing gene

Authors: Caitlin T Fierheller[1,2], Laure Guitton-Sert[3,4], Wejdan M Alenezi[1,2,5], Timothée Revil[1,6], Kathleen K Oros[7], Yuandi Gao[3,4], Karine Bedard[8,9], Suzanna L Arcand[2], Corinne Serruya[2], Supriya Behl[1], Liliane Meunier[10], Hubert Fleury[10], Eleanor Fewings[11], Deepak N Subramanian[12], Javad Nadaf[1,6], Jeffrey P Bruce[13], Rachel Bell[13], Diane Provencher[10,14], William D Foulkes[1,2,7,15], Zaki El Haffaf[16], Anne-Marie Mes-Masson[10,17], Jacek Majewski[1], Trevor J Pugh[13,18,19], Marc Tischkowitz[11], Paul A James[12,20,21], Ian G Campbell[12,20], Celia M T Greenwood[1,7,22,23], Jiannis Ragoussis[1,6], Jean-Yves Masson[3,4], Patricia N Tonin[1,2,15]

Affiliations: [1]Department of Human Genetics, McGill University, Montreal, Quebec, Canada, [2]Cancer Research Program, The Research Institute of the McGill University Health Centre, Montreal, Quebec, Canada, [3]Genome Stability Laboratory, CHU de Québec-Université Laval Research Center, Oncology Division, Quebec City, Quebec, Canada, [4]Department of Molecular Biology, Medical Biochemistry and Pathology, Laval University Cancer Research Center, Quebec City, Quebec, Canada, [5]Department of Medical Laboratory Technology, Taibah University, Medina, Saudi Arabia, [6]McGill Genome Centre, McGill University, Montreal, Quebec, Canada, [7]Lady Davis Institute for Medical Research, Jewish General Hospital, Montreal, Quebec, Canada, [8]Laboratoire de Diagnostic Moléculaire, Centre Hospitalier de l'Université de Montréal (CHUM),

Montreal, Quebec, Canada, [9]Département de pathologie et biologie cellulaire, Université de Montréal, Montreal, Quebec, Canada, [10]Centre de recherche du Centre hospitalier de l'Université de Montréal and Institut du cancer de Montréal, Montreal, Quebec, Canada, [11]Department of Medical Genetics, National Institute for Health Research Cambridge Biomedical Research Centre, University of Cambridge, Cambridge, UK, [12]Cancer Genetics Laboratory, Peter MacCallum Cancer Centre, Melbourne, Victoria, Australia, [13]Princess Margaret Cancer Centre, University Health Network, Toronto, Ontario, Canada, [14]Division of Gynecologic Oncology, Université de Montréal, Montreal, Quebec, Canada, [15]Department of Medicine, McGill University, Montreal, Quebec, Canada, [16]Centre de recherche du Centre Hospitalier de l'Université de Montréal, Montreal, Quebec, Canada, [17]Department of Medicine, Université de Montréal, Montreal, Quebec, Canada, [18]Department of Medical Biophysics, University of Toronto, Toronto, Ontario, Canada, [19]Ontario Institute for Cancer Research, Toronto, Ontario, Canada, [20]Sir Peter MacCallum Department of Oncology, University of Melbourne, Melbourne, Victoria, Australia, [21]The Parkville Familial Cancer Centre, Peter MacCallum Cancer Centre and The Royal Melbourne Hospital, Melbourne, Victoria, Australia, [22]Gerald Bronfman Department of Oncology, McGill University, Montreal, Quebec, Canada, [23]Department of Epidemiology, Biostatistics & Occupational Health, McGill University, Montreal, Quebec, Canada

Published in:

Genome Med. 2021 Dec 3;13(1):186. doi: 10.1186/s13073-021-00998-5

2.1 Abstract

Background: Familial ovarian cancer (OC) cases not harbouring pathogenic variants in either of the *BRCA1* and *BRCA2* OC-predisposing genes, which function in homologous recombination (HR) of DNA, could involve pathogenic variants in other DNA repair pathway genes.

Methods: Whole exome sequencing was used to identify rare variants in HR genes in a *BRCA1* and *BRCA2* pathogenic variant negative OC family of French Canadian (FC) ancestry, a population exhibiting genetic drift. OC cases and cancer-free individuals from FC and non-FC populations were investigated for carrier frequency of *FANCI* c.1813C>T; p.L605F, the top-ranking candidate. Gene and protein expression were investigated in cancer cell lines and tissue microarrays, respectively.

Results: In FC subjects, c.1813C>T was more common in familial (7.1%, 3/42) than sporadic (1.6%, 7/439) OC cases (P = 0.048). Carriers were detected in 2.5% (74/2950) of cancer-free females though female/male carriers were more likely to have a first-degree relative with OC (121/5249, 2.3%; Spearman correlation = 0.037; *P* = 0.011), suggesting a role in risk. Many of the cancer-free females had host factors known to reduce risk to OC which could influence cancer risk in this population. There was an increased carrier frequency of *FANCI* c.1813C>T in *BRCA1* and *BRCA2* pathogenic variant negative OC families, when including the discovery family, compared to cancer-free females (3/ 23, 13%; OR = 5.8; 95%CI = 1.7–19; *P* = 0.005). In non-FC subjects, 10 candidate *FANCI* variants were identified in 4.1% (21/516) of Australian OC cases negative for pathogenic variants in *BRCA1* and *BRCA2*, including 10 carriers of *FANCI* c.1813C>T. Candidate variants were significantly more common in familial OC than in

sporadic OC (*P* = 0.04). Localization of FANCD2, part of the FANCI-FANCD2 (ID2) binding complex in the Fanconi anaemia (FA) pathway, to sites of induced DNA damage was severely impeded in cells expressing the p.L605F isoform. This isoform was expressed at a reduced level, destabilized by DNA damaging agent treatment in both HeLa and OC cell lines, and exhibited sensitivity to cisplatin but not to a poly (ADP-ribose) polymerase inhibitor. By tissue microarray analyses, FANCI protein was consistently expressed in fallopian tube epithelial cells and only expressed at low-to-moderate levels in 88% (83/94) of OC samples.

Conclusions: This is the first study to describe candidate OC variants in *FANCI*, a member of the ID2 complex of the FA DNA repair pathway. Our data suggest that pathogenic *FANCI* variants may modify OC risk in cancer families.

Keywords: *FANCI*, Ovarian cancer, Cancer-predisposing gene, Whole exome sequencing, Tissue microarray, Protein expression, DNA repair, Fanconi anaemia pathway, Familial aggregation of cancer, Hereditary cancer

2.2 Background

Ovarian cancer (OC), with an overall 5-year survival rate of 40%, is the leading cause of death in women with gynecologic cancer [1]. The overall lifetime risk for OC in the North American population is 1.3% [1]. However, twin studies suggest that 22% of OC risk can be attributed to heritable factors [2] and having an affected first- degree relative confers a 3–7-fold increase in risk to this disease [3, 4]. Carriers that are heterozygous for pathogenic variants in *BRCA1* (*FANCS*) or *BRCA2* (*FANCD1*) have an estimated lifetime risk for OC of 17–44% (by age 80 years), depending on the gene mutated [5]. Pathogenic *BRCA1* and *BRCA2* variants have been reported in 65–85% of

cancer syndromes featuring high-grade serous ovarian carcinoma (HGSC) [6], the most common histopathological subtype of epithelial OC [7], and in 10–20% of HGSC cases regardless of age at diagnosis [8]. Identifying carriers of *BRCA1* and *BRCA2* pathogenic variants for cancer prevention (prophylactic surgery [9, 10]) and management of OC using new therapies (e.g. poly (ADP-ribose) polymerase inhibitors (PARPi) [11–16]) is being offered in medical genetic and gynecologic oncology settings.

New cancer-predisposing gene (CPG) candidates have been investigated with a focus on members of the Fanconi anaemia (FA) DNA repair pathway involving BRCA1 and BRCA2 function. The most promising new OC-predisposing genes are from reports of heterozygous carriers of candidate variants in *BRIP1* (*FANCJ*) [17, 18], *RAD51C* (*FANCO*) [19–22], and *RAD51D* [23]. In cancer families, carriers of pathogenic *RAD51C* and *RAD51D* variants have been estimated to have cumulative risks to age 80 of 11% (95% confidence interval (CI) = 6–21) and 13% (95%CI = 7–23), respectively, for OC [24]. Collectively, carriers of pathogenic variants in these genes do not account for a large proportion of familial OC and breast cancer (BC) cases that have not been attributed to the known CPGs. Therefore, it is possible that new CPGs conferring risk to OC have yet to be discovered.

The low incidence of OC, rarity of pathogenic variants in each proposed CPG, and genetic heterogeneity of the general population pose major challenges in finding new OC-predisposing genes. An attractive strategy for finding additional CPGs focuses on the investigation of demographically (ethnically or geographically) defined populations that have a history of founder effects. Due to a relatively few number of ancestors, rapid expansion and geographic isolation during 1608~1760 of the small

founding immigrant French population of Quebec from Europe (EUR), a loss of genetic variation has occurred resulting in subsequent waves of expansion of carriers of specific variants [25–29]. As French Canadians (FC) are more likely to harbour frequently occurring germline pathogenic variants, candidate variants for OC may be readily identified by sequencing familial cases and/or by comparing allele frequencies in cancer cases versus cancer-free controls in contrast to studies involving the general population due to allelic heterogeneity [25, 26]. Though 42 different pathogenic *BRCA1* or *BRCA2* variants have been identified in FC cancer families of Quebec, five recurrent pathogenic variants account for 84% of all mutation-positive BC and/or OC families [30]. This is in contrast to the over 2000 different pathogenic *BRCA1* and *BRCA2* variants reported for undefined populations [31]. Specific pathogenic variants in *PALB2* (*FANCN*—c.2323C>T; p.Q775X) [32] and *RAD51D* (c.620C>T; p.S207L) [33] have also been identified in FC BC and HGSC cases, respectively.

Using whole exome sequencing (WES), we identified carriers of the *FANCI* c.1813C>T; p.L605F missense variant in a *BRCA1* and *BRCA2* pathogenic variant negative FC OC family. We investigated this variant based on a candidate gene approach as FANCI is the FA Complementation Group I gene, which is an essential member of the FA-homologous recombination (HR) pathway that repairs interstrand crosslink (ICL) DNA damage and acts as the molecular switch to activate this pathway [34–38].

To evaluate the potential pathogenicity of *FANCI* c.1813C>T; p.L605F, we applied a strategy that took advantage of the observed genetic drift in the FC population by investigating its allele frequency in FC OC and cancer-free subjects. We performed

in cellulo (HeLa and OC cell lines) and in vitro experiments to investigate the functional effects of the encoded p.L605F isoform and its response to therapies used in the treatment of OC. We also investigated FANCI expression in HGSC and normal tissues. Lastly, we investigated Canadian non-FC (CDN) and Australian (AUS) cancer cases for rare candidate *FANCI* variants.

2.3 Methods

2.3.1 Study subjects

Information about all study subjects obtained from various biobanking resources can be found in Additional file 1: Table S1.

The FC cancer samples were obtained from Réseau de recherche sur le cancer (RRCancer) Tumour and Data biobank. The OC samples from this biobank derived its collection from patients attending a major gynecologic oncology hospital centre in the province of Quebec. This centre largely services FCs, where it is estimated that 85% of samples come from participants who self-identify as FC [39]. Samples within this collection with a familial history of OC and/or BC have been extensively studied, where the majority self-report grandparental FC ancestry of index cancer affected cases [30, 40, 41]. The allele frequency of *FANCI* c.1813C>T was determined by investigating selected index OC or BC cases, defined based on family history of OC and/or BC or sporadic disease where cases were not selected based on family history of cancer, where all were self-reported FC ancestry as previously described [30, 40, 41] (see Additional file 1: Table S1). These cases were mostly ascertained over a 20-year period from the early 1990s to 2004. OC families had at least two epithelial OC cases within first-, second-, or third-degree relatives and the average age of diagnosis was 50 years

(range 24–77). Hereditary breast and ovarian cancer (HBOC) families had at least two invasive BC cases diagnosed under the age of 65 and one epithelial OC case in first-, second-, or third-degree relatives and the average age of diagnosis was 43.7 years (range 18–65). Hereditary breast cancer (HBC) families had at least three invasive BC cases diagnosed under the age 65 in first-, second-, or third-degree relatives and the average age of diagnosis was 44.6 years (range 22– 65). All first-, second-, and third-degree relations needed to be within the same branch of the family. The *FANCI* locus was investigated in available WES data from a subset of 157 OC or BC cases of FC ancestry (see Additional file 1: Table S1). Sporadic BC cases were diagnosed with invasive BC before the age of 70 (average = 52.7, range 25–69) [42]. We cannot exclude the possibility that some cases occurred in more than one study group: based on RRCancer biobanking sample number, OC cases from at least 13 families were also found in pedigrees from BC cases that were genotyped from the familial HBOC study group.

Carrier frequencies of candidate variants were investigated in cancer-free FC study subjects using genotyping data obtained from CARTaGENE [43], a resource containing biological samples, genetic and health data for up to 43,000 adult residents in Quebec. The subjects investigated were recruited between 2009 and 2014, and had an overall average age of 54.7 years (range 39–71) [43] and included 2950 females (average age = 54.3 years; range 39–71) and 2299 males (average age = 55.3 years; range 39–70). Selection criteria for individuals with genotyping data are biased towards individuals with higher quantity of health data (see Additional file 1: Table S2). Individuals were defined as FC if they were born in the province of Quebec, their

parents and all four grandparents were born in Canada, and French was the first language learned.

Variants in the *FANCI* locus were investigated in available OC Association Consortium (OCAC) and BC Association Consortium (BCAC) data. These study groups and accompanying genotyping data have been described elsewhere [44–46]. Data from 25,509 epithelial OC cases (22,406 invasive cancer) and 40,491 controls of EUR ancestry [44] were available from OCAC, including those for histopathological subtypes for the entire cohort as have been previously reported [44]. Data from 46,785 BC cases and 42,892 controls of EUR ancestry [45, 46] were available from BCAC.

The *FANCI* locus was investigated in the AUS population from available germline sequencing data derived from WES analysis of HGSC cases as previously de- scribed [47]. Briefly, all AUS cases had ovarian, fallopian tube, or peritoneal cancer (*n* = 516) and did not carry pathogenic variants in *BRCA1* and/or *BRCA2* (see Additional file 1: Table S1). Genetic data from AUS controls (*n* = 4878) were available from the lifepool project as previously described [48].

The *FANCI* locus was investigated in germline sequencing data available from other non-FC CDN study groups comprised of female subjects with OC, BC, or pancreatic cancer (*n* = 63) who were recruited from health care research centres in the province of Quebec (Additional file 1: Table S1). All recruited individuals had a strong family history of BC. A BRCAPro score [49], which is based on studies of Ashkenazi Jewish and EUR ancestry individuals, was generated to predict the likelihood of families carrying pathogenic variants in *BRCA1* or *BRCA2*. Individuals with a BRCAPro score of

>10%, but with no pathogenic variants in these genes were selected. Of this set, 14 individuals were of Ashkenazi Jewish ancestry.

Kaplan-Meier curves for overall survival analysis was performed using available gene expression data from 35 cancer types (*n* = 12,373, including *n* = 425 OC cases), from The Cancer Genome Atlas (TCGA) Pan-Cancer data set of the TGCA Project and was obtained from (Additional file 1: Table S1) University of California Santa Cruz Xena Browser [50]. The *FANCI* locus was investigated in germline sequencing data available for 412 Pan-Cancer OC cases downloaded from TCGA. Characteristics for TCGA samples are available via the National Cancer Institute Genomic Data Commons and cBioPortal for Cancer Genomics.

To further protect the anonymity of study subjects, all samples were assigned a unique identifier and pedigrees were modified. This project has received approval from The McGill University Health Centre (MUHC) REB (MP-37-2019-4783 and 2017-2722). All participants provided informed consent and the research conformed to the principles of the Helsinki Declaration.

2.3.2 Identification of candidate *FANCI* c.1813C>T variant

The *FANCI* c.1813C>T variant was initially discovered in family F1528 and has since been updated to include new information, including histopathology of OC and a reported case of ear, nose, and throat cancer (Fig. 1). Peripheral blood lymphocyte (PBL) DNA (~ 500 ng) from two sisters from this family was captured with the Agilent SureSelect 50 Mb exome capture oligonucleotide library, and then sequenced with paired-end 100 bp reads on Illumina HiSeq 2000. After removing putative PCR-generated duplicate reads using Picard (V.1.48), sequencing reads were aligned to

human genome assembly hg19 using a Burroughs–Wheeler algorithm (BWA V.0.5.9). Sequence variants were called using Samtools (V.0.1.17) mpileup and varFilter meeting the following criteria: at least three variant reads, ≥ 20% variant reads for each called position, and Phred-like quality scores of ≥ 20 for SNPs and ≥ 50 for small insertions or deletions. Annovar [51] and custom scripts were used to annotate variants according to the type of variant, Single Nucleotide Polymorphism database designation (dbSNP), Sorting Intolerant from Tolerant (SIFT) score [52], and allele frequency data from the 1000 Genomes Project [53] and National Heart, Lung and Blood Institute (NHLBI) Exome Sequencing Project (ESP) v.2014 [54]. Then, the variant list was organized to select top candidate variants that were shared in common among the two sisters by de-prioritizing the following: (1) synonymous or intronic variants other than those affecting the consensus splice sites; (2) variants seen in more than 5 of 416 exomes from patients with rare, monogenic diseases unrelated to cancer that were independently sequenced and available at the McGill Genome Centre (MGC); and (3) variants with a frequency ≥ 1% in either the 1000 Genomes Project or NHLBI exome datasets. Using a candidate gene approach, we then further prioritized the list of candidates based on their role in FA-HR pathways. Using this strategy, *FANCI* c.1813C>T was the only candidate remaining on the list of prioritized variants (*n* = 276) shared in common between the two sisters in family F1528. The presence of the *FANCI* variant was verified using Integrative Genomics Viewer (IGV) [55]. The *FANCI* c.1813C>T variant was validated by targeted PCR analysis and bidirectional Sanger sequencing at the MGC using standard methods (see Additional file 1: Table S3).

Since the initial discovery of the *FANCI* variant in family F1528, newer WES capture kit technology and bioinformatic tools became available, and thus we repeated our analysis with DNA from the same sisters from this family. WES and bioinformatic analyses were again performed at the MGC using Roche NimbleGen SeqCap® EZ Exome Kit v3.0 (Roche Sequencing) followed by HiSeq 100bp paired-end sequencing (Illumina) applying the manufacturer's protocols. Sequencing reads were aligned to human genome assembly hg19 using BWA-MEM v0.7.17, then deduplicated using Picard v2.9.0 (Broad Institute). Bases were recalibrated using the GATK best practices. Variants were called using HaplotypeCaller available from GATK v3.5 (Broad Institute) and recalibrated according to GATK best practices. The filtered variants were then annotated and loaded into a GEMINI v0.19.1 database as per the recommended workflow. Data was filtered for non-synonymous rare variants (variant allele frequency [VAF] < 1%) deduced from a publicly available database Genome Aggregation Database (gnomAD) v2.1.1 [56] identified in genes with reported function in DNA repair pathways (*n* = 276 [57]). *FANCI* c.1813C>T was once again the only variant directly involved in the FA-HR DNA repair pathway identified in both sisters. The presence of the *FANCI* variant was again confirmed by IGV [55] and validated by PCR analysis and Sanger sequencing at the MGC using standard methods (see Additional file 1: Table S3).

2.3.3 Genetic analyses of candidate *FANCI* variants in FC cancer cases and cancer-free controls

In FC cancer cases, carriers of *FANCI* c.1813C>T were identified by targeted genotyping of PBL DNA samples or from surveying available WES data (subjected to

the same latest WES technology and data analysis pipeline as described above) from affected cases in our study groups (see Additional file 1: Table S1). PBL DNA from OC or BC cases were genotyped using a custom TaqMan® genotyping assay [58] based on established methods (see Additional file 1: Table S4). Where PBL DNA was no longer available from the study case, genomic DNA extracted from the tumour (if available) was provided by the RRCancer biobank for genotyping. PBL DNA from sporadic BC cases were genotyped using Sequenom® iPLEX® Gold Technology at the MGC [42]. Samples that were removed from the analysis were due to poor DNA quality (*n* = 30), duplication (*n* = 1), or were from cases exceeding age limit criteria (70 years or older when diagnosed with first invasive BC; *n* = 2). Results from a total of 558 cases were evaluated for *FANCI* c.1813C>T carrier status. The *FANCI* locus (NC_ 000015.9: g.89828441C>T) was reviewed in WES data, validated by IGV analysis, and *FANCI* c.1813C>T variant carriers verified by Sanger sequencing as described (see Additional file 1: Table S3).

To identify carriers of *FANCI* c.1813C>T in CARTaGENE FC cancer-free controls, data was extracted from available genotyping sets derived from germline DNA of subjects that were genotyped in three different batches using two different genotyping platforms (Illumina and Affymetrix; see Additional file 1: Table S2). Data was imputed when there was no representative probe for a locus on the genotyping array using the Sanger Imputation Service with Haplotype Reference Consortium (release 1.1) as the reference panel [59]. Pre-phasing and imputation was performed using Eagle2 [60] and the positional Burrows-Wheeler transform (PBWT) [61]. Samples were removed as part of quality control to improve imputation of the array (see Additional file 1: Table S2).

Two-sided Fisher's exact test was used to compare frequencies of *FANCI* c.1813C>T carriers in the cases and controls or between different study subjects, where a *p* value ≥ 0.05 was considered significant. Odds ratios (ORs) and 95% CIs were estimated for all study subjects for this allele.

2.3.4 Identification of candidate *FANCI* in various populations

Candidate *FANCI* variants were identified by investigating genotyping data available from OCAC, BCAC, and TCGA biobank resources or derived from the genetic analysis of AUS and CDN study groups (see Additional file 1: Table S1). Rare (VAF < 1%) *FANCI* variants were subjected to bioinformatic analyses using 13 in silico tools, to predict the effect of the nucleotide change(s), which includes four tools for conservation and three tools to predict splice site variants. These tools were selected for the best predictive performance [62]. Conservation tools included the following: Genomic Evolutionary Rate Profiling (GERP++) [63], Phylogenetic P values (PhyloP) 100 way in vertebrates [64], Phylogenetic Analysis with Space/Time models Conservation (PhastCons) 100 way in vertebrates [65], and Site-specific Phylogenetic analysis (SiPhy) 29 way in mammals [66], where variants were conserved if ≥ 2 in GERP++ and ≥ 0.4 in all other tools. In silico tools for missense variants included the following: Combined Annotation Dependent Depletion (CADD) [67] v1.6, Consensus Deleteriousness (Condel) [68], Eigen [69] v1.1, Meta-analytic Logistic Regression (MetaLR) [70], Meta-analytic Support Vector Machine (MetaSVM) [70], Variant Effect Scoring Test (VEST) [71] v4.0, and Rare Exome Variant Ensemble Learner (REVEL) [72], where variants were candidates if ≥ 15 in CADD and ≥ 0.4 in all other tools. Splice site variants were analysed with Maximum Entropy Modeling of Short Sequence Motifs

(MaxEntScan) [73] (splicing change if difference ≥ |2| and Database Splicing Consensus Single Nucleotide Variant v4.0 (dbscSNV) tools, AdaBoost (ADA) and Random Forest (RF) [74] (splicing change if score ≥ 0.4). Variants were considered candidates if they were predicted to be pathogenic/deleterious in ≥ five out of seven tools and ≥ two out of four conservation tools for missense variants or all three tools for splice site variants (± 5 nucleotides from the exon-intron junction). Nonsense and frameshift variants were considered candidates, but in-frame deletions were not. Variants were annotated using the Ensembl Variant Effect Predictor [75].

2.3.5 Genetic analysis of *FANCI* locus in OC and BC cases and controls from consortia databases

The *FANCI* locus was investigated in available OCAC and BCAC data. The $\log_2$OR, standard error (SE), χ2, and *p* value for 25,509 epithelial OC cases (22,406 invasive cancer) and 40,491 controls of EUR ancestry [44] were derived from OCAC resource. The $\log_2$OR, SE, χ2, and *p* value for 46,785 cases and 42,892 controls of EUR ancestry [45, 46] were derived from BCAC resource. Data was also available for carriers of *BRCA1* c.4327C>T and rs8037137 loci, which were used as comparators (see Additional file 1: Table S5). All rare (VAF < 1%) *FANCI* variants identified in the OCAC and BCAC resource were subjected to the same bioinformatic analyses using in silico tools as described.

2.3.6 Genetic analysis of *FANCI* locus in AUS HGSC cases and controls

The *FANCI* locus was investigated in germline sequencing data available from WES analysis of 516 AUS HGSC cases as previously described [47] (see Additional file 1: Table S1) and 4878 AUS controls from the lifepool study [48]. The identified rare

(VAF < 1%) variants found in *FANCI* were subjected to the same bioinformatic analyses using in silico tools as described.

2.3.7 Genetic analysis of *FANCI* locus in CDN BC cases

The *FANCI* locus was investigated in germline sequencing data available from other non-FC CDN study groups subjected to WES analysis of PBL DNA from subjects with OC, BC, or pancreatic cancer (*n* = 63) (Additional file 1: Table S1). FANCI variants were selected from PE125 WES data that was generated using the Nextera Rapid Capture Exome enrichment kit (Illumina) followed by HiSeq-4000 sequencing performed by the CRUK CI genomics core facility in the UK. Variant Call Format files were generated with a standard pipeline following GATK Best Practices recommendations for WES data. The identified rare (VAF < 1%) variants found in *FANCI* were subjected to the same bioinformatic analyses using in silico tools as described.

2.3.8 Genetic analysis of *FANCI* locus in TCGA Pan-Cancer cases

Processed *FANCI* mRNA expression and clinical data from TCGA Pan-Cancer data set were downloaded from University of California Santa Cruz Xena Browser [50]. Kaplan-Meier curves for overall survival were performed for all 35 cancer types from the Pan-Cancer TCGA [76]. Samples were dichotomized into high and low *FANCI* expression groups based on the median. For OC cases, data was parsed based on *BRCA1* and *BRCA2* pathogenic variant status (germline and somatic) according to TCGA reporting of variants. WES data from 412 OC cases of the Pan-Cancer TCGA set was downloaded and annotated using wANNOVAR [51]. The identified rare (VAF < 1%) variants found in *FANCI* were subjected to the same bioinformatic analyses using in silico tools as described.

2.3.9 Genetic analysis of variants in known OC-predisposing genes and DNA repair genes in FC *FANCI* c.1813C>T carriers

Rare (VAF < 1%) variants that were identified in known high-risk epithelial OC-predisposing genes in the analysis of WES data from *FANCI* c.1813C>T carriers was investigated using various bioinformatic tools. *BRCA1* and *BRCA2* variants were classified for their pathogenicity using BRCA Exchange [31] and ClinVar [77]. Rare (VAF < 1%) variants in DNA repair pathway genes (*n* = 276 [57]) were evaluated in *FANCI* c.1813C>T carriers. The only variant identified that was shared in all cases was *POLG* c.2492A>G (see Additional file 2) and it was pursued further as described below.

The allele frequency of *POLG* c.2492A>G was determined by investigating selected index OC or BC FC cases as above. Carriers of *POLG* c.2492A>G were identified by targeted genotyping of PBL DNA samples or from surveying available WES data from affected cases from our study groups as described. PBL DNA from OC or BC cases were genotyped using a custom TaqMan® genotyping assay [58] based on established methods (see Additional file 1: Table S4). *POLG* c.2492A>G was reviewed in available WES data as above. Genotyping data from CARTaGENE for cancer-free FC controls was investigated as above, including imputation (see Additional file 1: Table S2). *POLG* c.2492A>G was subjected to the same bioinformatic analyses using in silico tools as described.

2.3.10 Cell lines, cell culture, and reagents

HeLa cells and OVCAR-4 cells were grown in Dulbecco's modified Eagle's medium (CorningTM cellgroTM) and Dulbecco's modified Eagle's medium/Nutrient Mixture F-12 (GibcoTM) respectively, both supplemented with 10% foetal bovine serum

(GibcoTM), at 37°C, 5% CO_2, and 20% O_2. OVCAR-3 cells were grown in RPMI supplemented with 0.01 mg/ml bovine insulin and 20% foetal bovine serum (GibcoTM), at 37°C, 5% CO_2. HeLa cells knockout (KO) for *FANCI* were obtained using the ALT-R CRISPR-Cas9 system from Integrated DNA Technologies™. Cells were transfected with crRNA: tracrRNA:Cas9 RNP-complexes (crRNA sequence: AATCCCCCGATTCCACCAAC), according to the manufacturer's guidelines for RiboNucleoProtein transfection using RNAimax. After transfection, genomic DNA from the pool of transfected cells was extracted using QIAamp DNA Mini Kit (Qiagen, ref 51306). A 500-bp DNA region containing the sgRNA complementing sequence was amplified by PCR from 400 ng of genomic DNA with the Thermo Scientific™ Phusion™ High-Fidelity DNA Polymerase and verified by sequencing using the following primers: Forward: 5′-GTTACTGG ACTTCTCAAAAGCTGTAAG-3′ and Reverse: 5′-CTAGGTTGGGCACTTAAGTTTTCCT-3′. Sequencing results from non-transfected cells and genetically altered cells were compared using TIDE software to estimate the percentage of genetically altered cells. Clones were then generated and selected based on FANCI protein depletion using western blot analysis. Two clones, clones 1 and 2, were used in this study.

When specified, cells were treated with mitomycin C (MMC) from *Streptomyces caespitosus* (Millipore-Sigma, ref M0440) or formaldehyde (BAKER ANALYZED® ACS, J.T. Baker®, ref CAJT2106). For protein stability as- says, cycloheximide (CHX) (Millipore-Sigma, ref C4859) was used at a final concentration of 100 µg/ml.

2.3.11 siRNA transfection and complementation assays

Approximately 2.5 × 10^5 HeLa cells were transfected with 50 nM of siCTL (UUCGAACGUGUCACGUCAA) or siFANCI (UGGCUAAUCACCAAGCUUAA) with RNAimax (Invitrogen) according to the manufacturer's protocol. Then, after 24 h, cells were transfected again with the same siRNAs. After 6 h, cells were complemented with the indicated pcDNA3-Flag-FANCI constructs of Flag-FANCI or pcDNA3 empty vector (EV) using Lipofectamine 2000 according to the manufacturer's protocol, using the following quantities of plasmids: 1 µg of wild type (WT) and EV, 3 µg of p.L605F, and 1.5 µg of p.P55L. In the case of HeLa FANCI$^{-/-}$ cells, 3.5 × 10^5 cells were seeded and directly transfected with pcDNA3 or pEYFP-C1 constructs after 24 h. For immunofluorescence, peGFP and piRFP670-N1 plasmids, respectively, were co-transfected at a volume corresponding to 10% of the quantity of transfected pcDNA plasmid construct. Approximately 3 × 10^5 OVCAR-3 or OVCAR-4 cells were transfected with 50 nM of siCTL or siFANCI with RNAimax (Invitrogen) according to the manufacturer's protocol. After 24 h, cells were complemented with the indicated constructs of pcDNA3-Flag-FANCI constructs or pcDNA3 EV using Lipofectamine 3000 according to the manufacturer's protocol, using the following quantities of plasmids: 2 µg of WT FANCI or EV, 3 µg of p.L605F.

2.3.12 Protein extraction and immunoblotting

Cells were collected by trypsinization and rinsed once in cold PBS. Cell pellets were then incubated in lysis buffer (10 mM HEPES pH 7.4, 10 mM KCl, 1% Triton, 150 mM NaCl, 30 mM $Na_2P_2O_7.10H_2O$, 1 mM EDTA and 1 µg/ ml Leupeptin, 3.4 µg/ml Aprotinin, 1% PMSF, 5 mM NaF, 1 mM Na_3VO_4, Complete™ EDTA-free Protease Inhibitor Cocktail (Roche)) for 30 min on ice. Cell lysates were then sonicated for 5 min

(30 s on, 30 s off, high, Bioruptor) and centrifuged for 30 min, 13,000 rpm, 4°C. Supernatant was then processed for immunoblotting analysis using the indicated antibodies.

2.3.13 Antibodies for western blotting and immunofluorescence assays

The antibodies used were anti-FANCI (A7) (Santa Cruz Biotechnology, ref sc-271316, 1:100 for western blot), anti-FANCD2 (Novus, ref NB100-182D1, 1:5000 for western blot, 1:1000 for immunofluorescence), anti-Flag (Cell signaling Technologies, ref 8146, 1:1600 for immunofluorescence), and anti-vinculin (Sigma, ref V9131, 1:100,000 for western blot). Horseradish peroxidase- conjugated anti-rabbit IgG or anti-mouse (1:10,000; Jackson ImmunoResearch) were used as secondary antibodies for western blot. For immunofluorescence, Alexa Fluor secondary antibodies from Life Technologies (Goat anti-mouse Alexa fluor 568 A-11004, Goat anti-rabbit Alexa fluor 568 A-11011, Goat anti-rabbit Alexa fluor 488 A-11008) were used at a 1:1000 dilution.

2.3.14 Cisplatin and olaparib cell survival assays

Approximately 3 × 10^5 HeLa FANCI$^{-/-}$ cells were seeded into one well of a six-well plate. After 24 h, cells were complemented with the indicated Flag-FANCI con-struct using Lipofectamine 2000 (Invitrogen), and then after another 24h seeded in triplicate into a Corning 3603 black-sided clear bottom 96-well microplate at a density of 3500 cells per well. The remaining cells were stored at − 80°C until processing for protein extraction and immunoblotting as described above. Once attached to the plate, the cells were exposed to different concentrations of either 0–300 nM cisplatin (Tocris, #2251) or 0–2.5 μM olaparib. After 3 days of treatment, nuclei were stained with Hoechst 33342 (Invitrogen) at 10 μg/ ml in media for 45 min at 37°C. Images of entire

wells were captured at x 4 magnification using a Cytation^{TM} 5 Cell Imaging Multi-Mode Reader and Hoechst-stained nuclei were quantified with the Gen5 Data Analysis Software v3.03 (BioTek Instruments). Cell viability was expressed as percentage of cell survival in cisplatin or olaparib-treated cells relative to vehicle (DMSO)-treated cells. Results represent the mean ± standard error of the mean (SEM) of at least three independent biological replicates, each performed in technical triplicate.

2.3.15 Protein stability assays

To test the stability of Flag-FANCI variants, HeLa FANCI$^{+/+}$ cells were first transfected with siRNA targeting FANCI and then complemented with Flag-FANCI constructs as described above. For HeLa FANCI$^{-/-}$ clones, and OVCAR-3 or OVCAR-4, cells were directly transfected with Flag-FANCI constructs. Twenty-four hours after DNA transfection, cells were seeded in 6-well plates at 5 × 10^5 cells/well for HeLa and 3.5 × 10^5 cells/well for OVCAR-3 or OVCAR-4 and grown overnight. Cells were then treated with CHX (100µg/ml) and MMC (50 ng/ml) or formaldehyde (300 µM) or no genotoxic treatment for the indicated times. At each time point (t0, t1.5h, t3h, t4h, t5h, t6h, and t8h), cells were collected by trypsinization and snap-frozen after a wash in cold PBS. Samples were prepared for immunoblotting as described above. A first western blot was performed with all t0 timepoints to adjust quantity of samples to load for the whole kinetic in order to have comparable amounts of Flag-FANCI constructs at t0. Flag-FANCI WT and Flag-FANCI p.L605F were run on the same gel.

2.3.16 Immunofluorescence analyses

HeLa FANCI$^{-/-}$ cells were complemented with either FLAG-FANCI variants (1 µg of WT, 3 µg of p.L605F, and 1.5 µg of p.P55L) and 0.1 µg of transfection control peGFP

to identify transfected cells, or pEYFP-C1-FANCI (1 µg of WT, 3 µg of p.L605F) and 0.1 µg of transfection control piRFP670-N1 to identify transfected cells. One microgram of pcDNA3 or pEYFP-C1 was used as EV. After 18 h, cells were seeded on a glass coverslip for 8 h and then treated with 50 ng/ml MMC for 18 h and processed for immunofluorescence with anti-FANCD2 (Novus, ref NB100-182D1, 1:1000) antibody according to the protocol provided by Cell Signaling Technologies for Flag antibody (ref 8146). Briefly, cells were fixed in PBS-PFA 4% for 15min at room temperature and blocked and permeabilized in Blocking Buffer (1× PBS / 5% normal serum / 0.3% TritonTM X-100) for 30 min at room temperature. Incubation with anti-FANCD2 antibody, diluted in Antibody Dilution Buffer: (1X PBS / 1% BSA / 0.3% TritonTM X-100), was performed for 2 h, room temperature. After three washes of 5 min in PBS, Alexa Fluor secondary antibody from Life Technologies (Goat anti-rabbit Alexa fluor 568 A-11011) was used at 1:1000 in antibody dilution buffer and incubated for 1 h at room temperature. Finally, slides were incubated in DAPI for 15 min and washed two more times in PBS for 5 min, and ProLong® Gold Antifade Mountant (Invitro- gen^{TM}) was used as mounting medium. FANCD2 and YFP-FANCI foci were counted in transfected cells ac- cording to the transfection control used (peGFP- or iRFP-positive cells). HeLa FANCI$^{+/+}$ cells were transfected with siRNA and complemented with siRNA-resistant FANCI variants or EV as described above. After 18 h, cells were seeded on a glass coverslip for 8 h and then treated with 50 ng/ml MMC for 18 h and processed for immunofluorescence with anti-Flag (Cell signaling Technologies, ref 8146, 1:1600) and anti-FANCD2 (Novus, ref NB100-182D1, 1:1000) as described before except that incubation with primary antibody was performed at 4°C, overnight in a humid chamber.

Alexa Fluor secondary antibodies from Life Technologies (ref A-11008, A-11004) were used at 1:1000. In HeLa FANCI$^{-/-}$ cells, FANCD2 foci were counted in GFP-positive cells. In that case, only Flag-positive cells were taken into consideration for the quantification of FANCD2 foci. Each dot represents a nucleus and the red line corresponds to the mean of FANCD2 or FANCI foci per nucleus and error bars the SEM. Statistical analysis was performed using GraphPad Prism 8 (Kruskal-Wallis test).

2.3.17 Anti-Flag pulldown assays

After siRNA transfection and complementation with Flag-FANCI WT or p.L605F, HeLa FANCI$^{+/+}$ cells were lysed in lysis buffer (50 mM Tris–HCl, pH 7.5, 150 mM NaCl, 0.5% NP-40) containing protease and phosphatase inhibitors (1 mM PMSF, 0.019TI U/ml aprotinin, 1 mg/ ml leupeptin, 5 mM NaF, and 1 mM Na_3VO_4) incubated for 30 min on ice, and lysed by sonication. Insoluble material was removed by high-speed centrifugation (13,000 rpm at 4°C) and each immunoprecipitation was carried out using soluble protein extract in 1 ml of lysis buffer. Fifty millilitres of anti-Flag M2 affinity gel (Sigma) and 70 U of DNase I were added and incubated at 4°C for 2.5 h. Beads were washed three times with washing buffer (50 mM Tris–HCl, pH 7.5, 250 mM NaCl, 0.5% NP-40), and proteins were eluted with 60 µl of Laemmli buffer. Proteins were visualized by western blotting using the appropriate antibodies. Flag-FANCI p.L605F variant immunoprecipitation was overloaded in order to have the same amount of protein immunoprecipitated as in the Flag-FANCI WT lane and be able to compare co-immunoprecipitated FANCD2. Experiment has been performed twice.

2.3.18 FANCI protein expression by immunohistochemistry (IHC) analysis of HGSC tumours and normal tissues

Slides containing 4 micron slices of tissue microarrays (TMAs) containing 0.6 mm formalin-fixed paraffin-embedded (FFPE) tissue cores (spaced 0.2 mm apart) of HGSC (*n* = 101) [78] and normal fallopian tube (*n* = 15) tissues, and *FANCI* c.1813C>T carrier tumour tissues (*n* = 8) were stained using the BenchMark XT automated stainer (Ventana Medical System Inc., Roche). Antigen retrieval was carried out with Cell Conditioning 1 solution for 1 h. The FANCI polyclonal antibody (Sigma HPA039972 dilution 1/200) was automatically dispensed and the TMAs were incubated at 37°C for 1 h. The Ultra View DAB detection kit was used, and the slide was counterstained with haematoxylin. The TMAs were scanned with a 20 × 0.75 NA objective by VS-110 Olympus.

Staining patterns were evaluated by two independent observers. Intensity of staining was scored for all cores using a 4-point system; zero referring to no detectable staining to three referring to the highest staining intensity. As each sample was present in the TMA in duplicate, each case received four scores (two for the first core and two for the second core). The mode score was used for analysis where possible; otherwise, the average score was used. The interobserver correlation for IHC analysis of the TMA of HGSC samples was 89%. Staining patterns and analyses from the TMA containing HGSC samples and normal fallopian tube samples were evaluated without prior knowledge of carrier status for *FANCI* c.1813C>T. All HGSC and normal fallopian tube samples were genotyped for *FANCI* c.1813C>T variant as described, and one previously known carrier was identified (PT0004). Samples that could not be scored

were removed from further analysis (*n* = 7 HGSC samples, *n* = 2 fallopian tube epithelium [FTE] samples). A second TMA that contained 10 samples from eight *FANCI* c.1813C>T carriers (in duplicate) were also scored separately: the results from one sample from this TMA was removed from analysis due to poor tissue quality.

Spearman correlation was used to measure the strength of the correlation of staining intensity and survival data with clinical data as continuous variables. Survival curve was calculated according to Kaplan-Meier method coupled with a log rank test. Univariable Cox hazard models were used to estimate the hazard ratio as categorical data. All statistical analysis was done using Statistical Package for the Social Sciences software version 24 (SPSS, Inc) and results deemed statistically significance at $p \leq 0.05$.

2.4 Results

2.4.1 Discovery of *FANCI* c.1813C>T as a candidate

We previously reported a rare *BRCA1* and *BRCA2* pathogenic variant negative OC family (F1528) in a study of the histopathology of OC and *BRCA1* and *BRCA2* pathogenic variant carrier status of FC cancer families [79]. To clarify, *BRCA1* and *BRCA2* pathogenic variants were not identified in either sibling using two different WES platforms, which is consistent with independent clinical genetic testing results. To investigate if other candidate variants could be contributing to cancer risk in this family, WES and bioinformatic analyses were performed on PBL DNA available from two affected siblings both of whom had HGSC [79]. We selected rare (VAF < 1%) variants (*n* = 276) as candidates that were inherited in the heterozygous state and shared in common with the affected sisters. The only DNA repair pathway gene identified with a

variant was *FANCI* (c.1813C>T; p.L605F). This was an intriguing candidate to investigate given that family F1528 is predicted to harbour a pathogenic variant in *BRCA1* or *BRCA2* (Manchester score [80, 81]: *BRCA1* = 29, *BRCA2* = 20). As *FANCI* plays a role in FA-HR pathway it may be associated with phenotypically similar cancer families that have implicated *BRCA1* and *BRCA2* [79, 82] (Fig. 1). Preliminary in silico tools predicted this variant, located within the S/TQ phosphorylation cluster [34] of the encoded protein, to be highly conserved and probably damaging. However, at the time of discovery, the overall allele frequency of *FANCI* c.1813C>T from available databases was 0.76% in the NHLBI ESP v.2014 [54] and 0.2% in the 1000 Genomes Project [53]. These allele frequencies were notably higher than expected for individual pathogenic variants found in known OC-predisposing genes, such as *BRCA1* and *BRCA2* (0.001%). Therefore, we performed a molecular investigation before pursuing extensive genetic analyses of our study groups.

2.4.2 In cellulo and in vitro analysis revealed FANCI p.L605F isoform behaves differently than WT protein

FANCI belongs to the FA-HR DNA repair pathway that has been mainly described to be involved in ICL repair induced by DNA cross-linking agents, such as MMC. Briefly, when DNA replication forks are blocked by the presence of an ICL, FANCM recognizes the lesion, recruits the FA core complex which will ubiquitinate the heterodimer FANCI-FANCD2. Essential to downstream FA pathway function is this interdependent ubiquitination of both FANCI and FANCD2 [36–38], leading to DNA repair through DNA lesion excision of the DNA crosslink, DNA translesion synthesis, and HR. The functionality of this pathway can be assessed in cellulo by monitoring the

ubiquitination of FANCI and FANCD2 after MMC treatment. To investigate the functional impact of FANCI p.L605F isoform, both HeLa CRISPR FANCI KO (Fig. 2a–g, Additional file 3: Fig. S1, Additional file 4) or HeLa FANCI siRNA knockdown (KD) cells (Additional file 3: Fig. S1, Additional file 4) were complemented with the FANCI p.L605F isoform and treated with MMC. Western blot analysis first showed decreased levels of FANCI p.L605F isoform, unlike the FANCI p.P55L isoform encoded by variant c.164C>T which has been reported to exhibit WT function [37] (Fig. 2a, Additional file 3: Fig. S1, Additional file 4). Increasing the quantity of transfected *FANCI* c.1813C>T DNA by threefold did not overtly increase the level of protein expression comparable to that seen in the WT FANCI or p.P55L isoform (Additional file 3: Fig. S1, Additional file 4). We then looked at the impact of MMC treatment on FANCI and FANCD2 ubiquitination depending on FANCI status (Fig. 2a and Additional file 3: Fig. S1, Additional file 4). In WT or siCTL cells, both proteins are modified, as shown by the presence of the upper band (H). As expected, in the absence of FANCI, FANCD2 ubiquitination is lost. Complementation with WT FANCI or FANCI p.P55L isoform partially rescued the phenotype, though rescue was less evident in cells complemented with FANCI p.L605F. To confirm this, we then looked at FANCD2 ubiquitination after immunoprecipitation of FANCI WT or FANCI p.L605F isoform in presence of MMC. Given that the level of FANCI p.L605F isoform is lower than the WT in the input (Fig. 2b, left panel, Additional file 4), we overloaded the immunoprecipitated fraction of the variant in order to have the same signal in both lanes to be able to compare the two conditions. Though FANCI p.L605F isoform co-immunoprecipitates with FANCD2, ubiquitination levels of FANCD2 were severely diminished as compared to those in FANCI WT expressing cells (Fig. 2b,

right panel, Additional file 4) confirming our results (Fig. 2a, Additional file 4). These observations suggest that while physical interactions between FANCI p.L605F isoform and FANCD2 proteins are maintained the altered FANCI isoform may affect ubiquitination of FANCD2. As ubiquitination of FANCD2 is required to form MMC-induced foci, we then looked at FANCD2 foci formation in both KO and KD cells. Consistent with this role, the expression of FANCI p.L605F led to a significant reduction in the number of FANCD2 foci in transfected cells, while both WT and FANCI p.P55L isoforms were able to rescue the loss of FANCD2 foci observed in absence of FANCI (Fig. 2c, Additional file 3: Fig. S1, Additional file 4). Moreover, a concomitant reduction of GFP-FANCI p.L605F was also observed (Additional file 3: Fig. S2, Additional file 4).

As the expression of FANCI p.L605F appeared to be lower than the WT or p.P55L isoforms, even when increasing the quantity of plasmid (Fig. 2a, Additional file 3: Fig. S1, Additional file 4), we suspected that this protein isoform was unstable. Upon MMC or formaldehyde treatment, both known to induce DNA damage repaired by the FA-HR pathway, cells expressing FANCI WT protein, or either of the p.L605F and p.P55L isoforms, were treated with CHX to inhibit protein synthesis. FANCI protein levels decreased over time in response to both DNA damaging agents (Fig. 2d–f, Additional file 3: Fig. S1, Additional file 4). We recapitulated our findings in OVCAR-3 and OVCAR-4 cell lines to determine if FANCI is also unstable in an OC cell line background (Additional file 3: Fig. S2, Additional file 4). The effect was more prominent in FANCI p.L605F expressing cells as compared to WT FANCI or p.P55L expressing cells. These observations suggest that treatment with genotoxic agents exacerbates FANCI p.L605F protein instability, as it has been previously described for BRCA2

protein [83]. This is in agreement with the observation that FANCI p.L605F failed to complement survival of the HeLa FANCI$^{-/-}$ cells that were challenged with the platinum compound cisplatin (Fig. 2g, Additional file 3: Fig. S1, Additional file 4), a drug known to induce DNA crosslinks. In contrast, albeit in accordance with the literature [84], FANCI$^{-/-}$ cells were not sensitive to olaparib, a PARPi (Fig. 2g, Additional file 3: Fig. S1, Additional file 4).

2.4.3 *FANCI* c.1813C>T carriers are enriched in familial OC cases of FC ancestry

With these promising results in hand, we assessed *FANCI* c.1813C>T carrier frequency in available PBL DNA from index OC or BC cases of FC ancestry to determine if this variant plays a role in conferring risk in phenotypically defined cancer families [17–20, 22, 23, 30, 32, 33, 40, 41, 79, 85–90]. These OC or BC cases were selected based on their family history of OC or BC, or regardless of cancer family history (sporadic cases), where *BRCA1* and *BRCA2* pathogenic variant carrier status was known [30, 32, 33, 39–41, 85, 86, 91–95]. Index OC cases from OC families (3/42, 7%) had a higher carrier frequency of *FANCI* c.1813C>T than sporadic OC cases (7/439, 1.6%, $P = 0.048$, Fisher's exact) and sporadic BC cases (8/558, 1.4%, $P = 0.035$, Fisher's exact). Index OC cases from OC families (3/42, 7%) and index BC cases from HBOC families (3/82, 3.7%) had a higher carrier frequency than BC cases from HBC families (3/ 158, 1.9%), though these differences were not statistically significant ($P = 0.11$ and $P = 41$, respectively, Fisher's exact) (Table 1, Additional file 3: Fig. S3). When including the discovery OC family, there was an increased carrier frequency of c.1813C>T in *BRCA1* and *BRCA2* pathogenic variant negative OC families

versus sporadic OC cases (*P* = 0.01, Fisher's exact) and cancer-free females (3/23, 13%; OR = 5.8; 95%CI = 1.7-19; *P* = 0.005).

2.4.4 Cancer-free FC *FANCI* c.1813C>T carriers are significantly correlated with having a first-degree relative with OC

Recently, new data has become available from the CARTaGENE biobank enabling the evaluation of allele frequencies in study subjects from a cancer-free female FC population, and thus providing a more comparable reference group to our FC cancer subjects [43]. Using data from three different genotyping platforms, we estimated a 1.3% VAF in cancer-free FC females (Additional file 1: Table S6). This is not significantly different from the 1% estimated VAF in non-Finnish EURs, a population most likely to share common ancestry with FCs (France) [25, 26], as reported in the gnomAD [96] (Additional file 1: Table S6). In this database, the estimated VAF was 0.67% for the total of all study populations and varied across populations: highest in Estonians (2.1%) to none in East Asians. Rare homozygous carriers (17/134,154, 0.013%) were also identified in gnomAD. This finding did not dissuade us from pursuing this candidate variant as the in cellulo findings suggest that it may behave as a hypomorph.

The estimated carrier frequency at 2.5% in cancer-free FC females was lower than that observed in index cancer cases from OC (7.1%) and HBOC (3.7%) families, but higher than observed in sporadic OC cases (1.6%) and index BC cases from HBC families (1.9%), though these differences were not statistically significant (Table 1). Additional information was available from the CARTaGENE subjects to investigate *FANCI* c.1813C>T carrier frequency in the context of cancer family history (first-degree

only), reproductive history, oral contraceptive pill use, oophorectomy, and fallopian tube ligation; all of which are host factors that are known to significantly impact lifetime risk of OC [97, 98]. We observed that cancer-free carriers (female/male) were significantly correlated with having a first-degree relative with OC (Spearman correlation = 0.037; *P* = 0.011) compared to non-carriers, when analysing data from subjects genotyped with arrays that included probes for the variant allele (*n* = 4645) (Additional file 1: Table S7). The correlation is still significant, though slightly weaker, when adding data from cancer-free subjects where genotypes were imputed (*n* = 604; Spearman correlation = 0.027; *P* = 0.047) (Additional file 1: Table S2 and S7). No other cancer type was significantly correlated with carrier status. The majority of cancer-free FC females were parous (78%, 2315/2950) and had experienced oral contraceptive pill use, oophorectomy, and/or tubal ligation (91.8%, 2710/2950) (Additional file 1: Table S8). Only 8.1% (6/74) of c.1813C>T carriers reported no risk-reducing host factors.

2.4.5 Other candidate FANCI variants are rare in OC cases of FC ancestry

To determine if there are other *FANCI* variants (VAF < 1%) in FCs, we investigated available WES data from 80 familial and/or young age of onset OC cases, regardless of *BRCA1* or *BRCA2* pathogenic variant carrier status (Additional file 1: Table S1). We identified seven rare variants among 32 index OC familial cases, where one carrier was heterozygous for *FANCI* c.1573A>G; p.M525V (Additional file 1: Table S9). Although this missense variant is predicted to be highly conserved (all four conservation tools used) and damaging by in silico tools, in cellulo analyses suggested that it does not encode an aberrantly functioning protein (data not shown). Thus, *FANCI*

c.1813C>T is the only plausible candidate variant identified in *FANCI* in FC OC cases (Fig. 3a).

2.4.6 Co-occurrence of other candidate variants in OC-predisposing genes in *FANCI* c.1813C>T carriers

We analysed WES data from *FANCI* c.1813C>T OC (*n* = 12) carriers for the co-occurrence of pathogenic variants in known high-risk OC-predisposing genes [99]. No additional carriers of *BRCA1*, *BRCA2*, *BRIP1*, *RAD51C*, and *RAD51D* pathogenic variants were found in our familial cases. None of the sporadic OC cases (*n* = 7) carried pathogenic variants in *BRCA1*, *BRCA2*, *BRIP1*, *RAD51C*, and *RAD51D*. Moreover, the *FANCI* c.1813C>T variant did not co-occur in carriers of recurrent *BRCA1* [39], *BRCA2* [39], and *RAD51D* [33] in the FC population.

2.4.7 OC and BC cases of non-FC ancestry also carry candidate *FANCI* variants

We identified 99 unique *FANCI* variants (VAF < 1%) in 516 AUS HGSC *BRCA1* and *BRCA2* pathogenic variant negative cases [47] and 4878 AUS cancer-free controls from available WES data (Additional file 1: Table S1). Based on in silico tools, there were 10 candidate missense variants in 516 HGSC cases (4.1%), where 10 (1.9%) cases carried *FANCI* c.1813C>T and 11 (2.1%) cases carried other variants (Table 2, Fig. 3b, Additional file 1: Table S9). We identified 42 different candidate missense variants in 190/4878 (3.9%) AUS controls, where 95 (1.9%) carried *FANCI* c.1813C>T (Fig. 3d). The number of carriers of candidate variants in *FANCI* was not significantly different between AUS cases and controls (*P* = 0.48), including for *FANCI* c.1813C>T alone (*P* = 0.81). There was no significant difference in allele frequencies of *FANCI* variants in AUS cases compared to AUS controls (Additional file 1: Table S10), though

for five of eight rarest candidate *FANCI* variants (VAF < 0.1%) odds ratios were > 12 when compared to gnomAD cancer-free controls (Table 3). In contrast, *FANCI* c.1813C>T was the only variant identified in 1/63 (1.6%) familial CDN *BRCA1* and *BRCA2* pathogenic variant negative BC cases (Fig. 3c, Additional file 1: Table S9) and the carrier was known to be of Greek Canadian origin.

With respect to family history of cancer in AUS cases, five of the c.1813C>T carriers had a family history of OC within third-degree relatives (5/98, 5.1%), which was significantly higher than the carrier frequency of this variant in combined isolated cases of HGSC cases (5/418, 1.2%; $P = 0.025$) (Table 2). In contrast, there was no significant difference in the carrier frequency of *FANCI* c.1813C>T in combined cases with a reported family history of OC and/or BC (6/262, 2.3%) than those without (4/254, 1.6%; $P = 0.75$). *FANCI* c.1813C>T co-occurred with another candidate missense variant, *FANCI* c.2366C>T; p.A789V, in a HGSC case diagnosed at 75 years with a family history of OC. Three carriers of other candidate variants in *FANCI* (c.1573A>G; p.M525V, c.1264G>A; p.G422R, and c.1412C>G; p.P471R), with an average age of diagnosis of 52 years, also had a family history of OC (Table 2), where the carrier of p.M525V had no family history of BC. The number of carriers of candidate *FANCI* variants with a family history of OC (8/98, 8.2%) was significantly higher than isolated cases of HGSC ($P = 0.04$), but there was no significant difference when accounting for family history of OC and/or BC ($P = 0.66$).

We investigated rare variants (VAF < 1%) from imputed SNP array data that was available from two case-control studies: OCAC [44] and BCAC [45, 46]. In all, nine missense and one splice site variant were identified in OCAC and BCAC databases. *FANCI*

c.1813C>T and c.824T>C were the only candidate missense variants, but the splicing variant c.3007-1G>A may be a candidate (Additional file 1: Table S11). The data imputed from the OCAC database [44] revealed that the OR for *FANCI* c.1813C>T and c.824T>C was highest in HGSC and endometrioid subtypes compared to all histopathological subtypes combined, though there was no significant difference in allele frequency in OC cases compared to controls (Additional file 1: Table S5). To compare to a known OC pathogenic variant, OCAC data was investigated for the most common pathogenic *BRCA1* variant reported in FCs, c.4327C>T; p.R1443X, and also found repeatedly in populations of EUR ancestry [39]. As similarly observed with *FANCI* c.1813C>T, the OR was highest in HGSC cases, though there was no significant difference in allele frequency when all OC subtype cases were compared to controls (Additional file 1: Table S5). There was no significant difference in allele frequency between BC cases and controls for *FANCI* c.1813C>T in a similar analysis of BCAC case-control data [45, 46] (*BRCA1* c.4327C>T; p.R1443X and *FANCI* c.824T>C data was not available in the BCAC database) (Additional file 1: Table S5).

2.4.8 Clinical features of OC from *FANCI* c.1813C>T carriers are similar to those of HGSC cases

We reviewed available clinical characteristics of OC in *FANCI* c.1813C>T carriers. Given the paucity of *FANCI* variants, we focused on 13 FC OC carriers of *FANCI* c.1813C>T from familial and sporadic FC OC study groups. The seven carriers found in the context of sporadic OC cases (Additional file 1: Table S12) were reported as HGSC subtype. The remaining six carriers were identified in OC cases with a known family history of cancer (Additional file 1: Table S12), where five had serous subtype OC

and one had a mucinous subtype OC. There appeared to be no striking differences in the ages of the diagnosis for OC in carriers where age ranged from 40 to 81 years (average = 59.2 years) as compared with non-carriers in the sporadic OC group (average = 61years, range 36–81years) (Additional file 1: Table S12). Similarly, AUS HGSC *FANCI* carriers were diagnosed with OC between the ages of 31–82 years (average = 60 years) (Additional file 1: Table S12). Although sample size was limited, age range of carriers in FC cases was consistent and aligned with average age of diagnosis of OC in the North American population [1].

Available survival data showed that all seven *FANCI* c.1813C>T carriers from the sporadic OC group were deceased by the time of our investigation. They had an average survival of 61.1 months (range 9–163). Due to sample size, we were unable to perform survival analysis using Kaplan-Meier estimation as 57% (4/7) OC cases did not survive past five years (Additional file 1: Table S12). The two carriers with survival past 61 months (2/7; 28%) is comparable to survival of non-carrier sporadic HGSC cases (100/334; 30%).

2.4.9 FANCI protein is expressed at low-to-moderate levels in HGSC tumour samples

We performed IHC analysis of an available TMA containing cores from FFPE HGSC tumour tissues and FTE cells, a proposed tissue of origin for the HGSC subtype [100– 106], staining for FANCI protein. Though a dualistic origin has been proposed for HGSC [107–109], we were only able to study FTE tissue. IHC analysis revealed strong nuclear and low-to-moderate cytoplasmic staining in FTE cells, in contrast to stromal cell components where staining was low or undetectable (Fig. 4a, Additional file 3: Fig. S4). In contrast, IHC analysis of tumour cells in HGSC tissue cores exhibited variable

staining (Fig. 4b, Additional file 3: Fig. S4), where the majority (83/94, 88.3%) exhibited low-to-moderate nuclear and cytoplasmic staining in epithelial components, compared to the stromal cell components where staining intensity was low or undetectable. Using Kaplan-Meier survival analysis, we found no correlation of staining intensity in epithelial tumour cell components of the HGSC tissue cores with overall or disease-free survival (Additional file 3: Fig. S4). Age at diagnosis, disease stage, residual disease, chemotherapy type, and survival (disease-free and 5-year) were not correlated with the intensity of protein staining.

A separate IHC analysis of tumour tissues available from eight *FANCI* c.1813C>T carriers revealed a range of staining intensity (Additional file 3: Fig. S4), consistent with the expectation that the variant encoded protein could be expressed in tumours (Fig. 2a). We were not able to similarly investigate by correlative or Kaplan-Meier analyses *FANCI* variant c.1813C>T carriers due to the small number of cases.

2.4.10 *FANCI* mRNA expression is associated with survival in TCGA OC cases

Using Kaplan-Meier analysis of TCGA Pan-Cancer cases, we found that adrenocortical cancer, kidney chromophobe, lower-grade glioma, lower-grade glioma and glioblastoma, lung adenocarcinoma, melanoma, mesothelioma, pancreatic cancer, and sarcoma along with OC were showed significant association with survival for *FANCI* mRNA expression (Additional file 1: Table S13). OC cases with high *FANCI* mRNA expression had significantly better overall survival compared to cases with low *FANCI* mRNA expression (Fig. 5a). We found that *BRCA1* and *BRCA2* pathogenic variant carriers did not show this survival benefit (Fig. 5b) and non-carriers had a stronger survival signal (Fig. 5c).

Ten rare *FANCI* variants were identified in 18/412 (4.4%) TCGA OC cases from germline WES data, where four variants are candidates based on in silico tools, including c.1813C>T (Additional file 1: Table S14). Six carriers of c.1813C>T were identified (1.5%), which is comparable to the frequency of carriers identified in the FC sporadic OC study group (1.6%). Of the 10 carriers of the 4 candidate *FANCI* variants, 3 cases had co-occurring pathogenic variants in *BRCA1* (*n* = 1) or *BRCA2* (*n* = 2). No *FANCI* carriers had co-occurring candidate variants in *BRIP1*, *RAD51C*, or *RAD51D*. Age of diagnosis was similar to FC OC cases ranging from 38 to 81 (average = 58.9; *n* = 9 cases).

2.5 Discussion

FANCI c.1813C>T was the only candidate *FANCI* variant identified in our study of FC OC cases. Our strategy for the discovery of new CPGs in OC was predicated upon the genetic drift observed in FCs of Quebec and thus the expectation that candidate risk alleles frequently occur and can readily be identified due to common ancestors in this population [25, 26]. Our findings are reminiscent of the identification of specific variants in familial FC cancer populations of Quebec, such as *RAD51D* c.620C>T; p.S207L in familial and sporadic OC cases [33], *PALB2* c.2323C>T; p.Q775X in BC cases and HBC families [32, 110], and *MSH6* c.10C>T; p.Q4X in colorectal cancer (Lynch Syndrome) families [111]. Given the unique genetic architecture of the FC population of Quebec, it is likely that carriers of *FANCI* c.1813C>T have common ancestors as has been shown with carriers of frequently occurring pathogenic variants in *BRCA1* [40, 112], *BRCA2* [40, 92, 113], and *MSH6* [111] in cancer families. As expected, given the genetic heterogeneity observed with the

above examples of CPGs in non-FC populations, we identified 10 candidate *FANCI* variants in AUS HGSC cases and 4 in TCGA OC cases, which included our *FANCI* variant. Although a recent genome-wide discovery study of AUS HGSC cases did not report *FANCI* among the list of potential new CPGs for OC [47], missense variants were not investigated [47].

FANCI c.1813C>T might exert its deleterious effect as a hypomorphic variant, as suggested by the instability of the encoded isoform in our cell line models, which include OC cell lines. Though tumour DNA was not available for all of our variant carriers, Sanger sequencing of DNA from FFPE tumour cells suggest loss of the WT allele and retention of the variant allele had occurred in two FC HGSC *FANCI* c.1813C>T carriers, as shown in Additional file 3: Fig. S5. Interestingly, tumour samples from a bilateral OC case predominantly exhibited the *FANCI* variant allele suggesting that loss of the WT allele could have been an early event in tumour progression in this case. Also, HGSC samples from both cases had acquired somatic pathogenic variants in *TP53*, a known major driver of tumourigenicity in the majority of HGSCs [114, 115]. Our IHC analyses showed differential FANCI protein expression, with a high proportion of HGSC tumour cells exhibiting low-to-moderate levels of protein expression. This is in contrast to consistent FANCI protein expression observed in FTE cells. These findings suggest loss of *FANCI* may play a role in OC akin to that suggested by other CPGs in the HR pathway, such as *BRCA1* and *BRCA2* [116]. In light of the dualistic origin of epithelial OC [107–109], future studies involving ovarian surface epithelial cells could also define the role of FANCI in OC. Results from analyses of TCGA data also suggest the role of *FANCI* in OC where OC cases with higher *FANCI* mRNA expression had a better

overall survival outcome. In keeping with this hypothesis is that loss of the chromosome 15q arm, which contains the *FANCI* locus (15q26.1), has been reported in 55% of 978 HGSC samples by TCGA project [115]. Though the curves of the TCGA Kaplan-Meier plots are separated at the 5-year mark, future analyses of a large sample group, focusing on 5-year survival, could potentially have more clinical relevance as the majority of HGSC patients (>75%) are deceased in this time period.

The highest frequency of carriers was in *BRCA1* and *BRCA2* pathogenic variant negative OC index cases from OC families (13%), when also including the multi-case discovery family in this group, which is significantly higher compared to sporadic OC cases ($P = 0.01$, Fisher's exact). Variant carriers in *BRCA1* and *BRCA2* pathogenic variant negative OC families were also more frequent when compared to cancer-free FC females by including the OC discovery family in our analysis ($P = 0.02$, Fisher's exact).

Based on available genetic data from non-Finnish EURs, the allele frequency of *FANCI* c.1813C>T at 1% is higher than expected as compared to many pathogenic variants in established CPGs. Similarly, the carrier frequency of c.1813C>T in AUS cancer-free controls at 1.9% was more common than anticipated. The carrier frequency of c.1813C>T in the general population is reminiscent of the pathogenic *CHEK2* c.1100delC, a moderate-risk BC-predisposing variant, which also has a similarly high carrier frequency of 1.4% in population controls as compared with other pathogenic variants in known CPGs for BC and OC [117]. This *CHEK2* variant was also found more frequently in BC cases from HBC families than sporadic BC cases, relative to healthy controls [117]. Although our estimates of overall risk to OC using OCAC data was

inconclusive, carriers of *FANCI* c.1813C>T in FC *BRCA1* and *BRCA2* pathogenic variant negative OC families have an increased risk based on the OR of 5.8 (95%CI = 1.7–20; $P = 0.005$). Though the confidence interval is wide, due to the small sample size, our findings are supported by the observation that cancer-free *FANCI* c.1813C>T carriers (female/male) were more likely to have a first-degree relative with OC in the FC population.

Given the allele frequencies observed among OC cases and controls, it is clear that penetrance is low for *FANCI* variant carriers. Although we cannot obtain a precise estimate given the numbers of carriers available, penetrance for *FANCI* will evidently be much lower than penetrance for pathogenic variants in *BRCA1* and *BRCA2*. Assuming that *FANCI* is a risk variant for OC, it is possible that other variants modify this risk. Although we did not identify other strong candidates in our WES analyses, it may be possible in the future to estimate a polygenic risk for OC based on a set of common variants, and then to explore the *FANCI*-associated risk of OC after controlling for the polygenic background, as has been done for BC and other diseases [118]. Similarly, the effect of risk modifiers in the CARTaGENE cancer-free controls in the context of *FANCI* variant carriers is unknown. We are mindful of the fact that FC cancer cases were recruited during a different time period than FC cancer-free controls, and it is possible that risk modifiers could be different across these groups, though this information is not available for FC cancer cases.

It is interesting that *rs8037137*, which is located 1.68 mega-base pairs downstream of *FANCI* c.1813C>T, was among the polymorphic genetic markers found significantly associated with risk to either invasive epithelial or HGSC subtype OC in a

large genome-wide association analysis of OCAC data [44]. Consistent with these findings is our observation that the OR for *FANCI* c.1813C>T and c.824T>C in the OCAC study groups are highest in endometrioid and HGSC subtype OC cases. A similar analysis of other candidate *FANCI* variants identified in our study was not possible as corresponding genetic data was not available in the OCAC database. The possibility that *FANCI* c.1813C>T is a moderate-risk allele with variable penetrance is consistent with our observations, though we are mindful of the limitations of our study due to sample size. Based on the allele frequency, we would require an estimated sample size of approximately 100 OC families and 7000 female cancer-free controls or 13,000 HGSC cases and 115,000 female cancer-free controls to achieve 80% power, numbers that are currently unattainable in FCs.

During the course of this investigation, *FANCI* loss-of- function and missense variants in a targeted analysis of selected DNA repair genes in OCAC cases (*n* = 6385) and controls (*n* = 6115) were reported, where only *PALB2* showed significant differences [119]. Based on sample size, the study was not sufficiently powered to identify moderate-risk alleles. Interestingly, 49 candidate *FANCI* variants, including loss-of-function variants (frameshift, nonsense, and splicing), and missense variants were reported (see Additional file 1: Table S15 and Additional file 3: Fig. S6). Although we were able to analyse *FANCI* c.1813C>T in OCAC, this variant was not listed among the candidates, as only variants with VAF < 1% were investigated in this study.

Although *FANCI* c.1813C>T variant carriers were found in FC familial BC cases, there were proportionally more carriers in BC cases from HBOC families than in HBC families. We also identified a variant carrier in a BC family of Greek Canadian origin, a

family from the same catchment area as our FC cancer families. These findings are in part reminiscent of the variable penetrance for BC and OC for known high-risk CPGs, where carriers are more likely to harbour pathogenic variants in *BRCA1* or *BRCA2* (or *PALB2*) based on family history of BC and OC [40]. There have been independent reports of BC cases carrying other *FANCI* variants with VAF 10^{-3} to 10^{-6} in cancer-free individuals. At least 19 different variants have been described in familial and/or sporadic BC cases: four nonsense, three frameshift, two splicing, and 10 missense (see Additional file 1: Table S16 and Additional file 3: Fig. S7) [120–127]. These *FANCI* variants were reported in Finnish [125] (4/1524, 0.3%), Chinese [124] (1/99, 1%), and two Spanish [122, 126] (1/154, 0.6% and 1/94, 1.1%) studies. The role of *FANCI* in other cancer types remains to be determined, though there have been reports of *FANCI* variant carriers in a variety of cancer types such as prostate cancer [125, 128, 129], sarcoma [130], malignant pleural mesothelioma [131], acute myeloid leukaemia [132], head and neck carcinoma [133], and colorectal cancer [134] (see Additional file 1: Table S17 and Additional file 3: Fig. S7).

FANCI regulates the recruitment of the FA core complex to sites of interstrand crosslinks, and thus plays an important function upstream in the FA-HR DNA repair pathway [135]. FANCI encodes one of only two proteins that comprise the ID2 complex, the other being FANCD2. In cellulo modeling using cell lines, pathogenic variants or gene knockouts of *BRCA1*, *PALB2*, or *RAD51D* have exhibited sensitivity to cisplatin and PARPi's, providing some insight into their role in DNA repair [33, 136–138]. We observed sensitivity to cisplatin but not to the PARPi olaparib in cell lines expressing the FANCI p.L605F isoform. Although the mechanism is unknown, these findings are

consistent with a report showing lack of sensitivity to a PARPi (KU0058948) in a fibroblast cell line transduced with HPV E6/E7 from a *FANCI* FA patient, as well as in cell lines generated from *FANCA*, *FANCL*, *FANCD2*, and *FANCJ* (*BRIP1*) patients [84]. The indirect role of FANCI in HR DNA repair and recent evidence suggesting that FANCI also has functions independent from the FA DNA repair pathway [139–143] may be consistent with our in cellulo studies. Further investigation of FANCI p.L605F in additional cell lines, including normal cell lines which are more karyotypically normal such as those that are representative of the various origins of epithelial OCs, could lend support to the influence of this variant on protein function in this cancer context.

Biallelic inactivation of *FANCI* has been associated with FA, a rare autosomal recessive disease that is characterized by congenital defects and developmental disabilities [36–38]. FA is a heterogenous genetic disease with 22 known causal genes, where *FANCI* implicated cases comprise approximately 1% of all FA diagnoses [144]. No clear genotype/phenotype association has been identified for *FANCI*-associated FA, though 7/16 (44%) patients show at least three features of the VACTERL-H association [145], which is a disease characterized by a non-random association of birth defects (typically at least three) affecting multiple parts of the body. *FANCI* c.1813C>T; p.L605F has been reported in ClinVar as benign or likely benign (n = 6 submissions) in the context of *FANCA* associated FA (n = 1), *FANCI*-associated FA (n = 2), or unspecified conditions (n = 3) with only in silico (no in cellulo or in vitro) evidence provided to classify this variant and no information on zygosity in carriers nor cancer context. Mild or no FA phenotypes have been observed for other homozygous hypomorphic variants in FA genes (*BRCA1* (*FANCS*), *BRCA2* (*FANCD1*), *FANCA*, and *PALB2* (*FANCN*)) [146–

149]. Hypomorphic variants in *RB1*, the causative gene of retinoblastoma have been found to confer significantly lower penetrance (< 25%) as compared to more common loss-of-function variants which are highly penetrant (> 95%) for the disease [150]. As *FANCI*-associated FA cases are rare, the incidence of cancer in biallelic carriers has not been reported. Heterozygous carriers of *FANCI* c.1264G>A; p.G422R, a pathogenic variant that has been reported in two *FANCI*-associated FA cases [38, 151], were identified in AUS cases and controls in our study. Although there was no information about cancer incidence, a Fanci KO mouse model was recently reported describing phenotypes consistent with developmental defects, though they also reported a low Mendelian ratio [152].

2.6 Conclusions

This is the first study to describe candidate variants in *FANCI* in the context of familial OC and in a member of the ID2 complex of the FA DNA repair pathway. Our strategy of investigating a limited number of familial and sporadic cancer cases from a population showing genetic drift found an increased frequency of carriers in OC cases. In cellulo and in vitro analysis of a missense variant found to recur in cancer cases implicates *FANCI* as a new candidate OC-predisposing gene. This study emphasizes the importance of pursuing missense variants during the gene discovery phase, especially when plausible candidates are revealed by analyses of defined cancer families. Indeed, a large number of pathogenic variants in known CPGs, such as *BRCA1* and *BRCA2*, are missense variants where they have been vetted using in cellulo and/or in vitro functional studies [31]. Although some of the identified *FANCI* variants are predicted to affect gene function as shown by in cellulo analyses of FANCI

p.L605F isoform, further studies are warranted to evaluate their role in OC risk. Our study suggests the possibility that *FANCI* variants might confer moderate risk to OC akin to *CHEK2* variants to BC risk and question the classification of *FANCI* c.1813C>T as benign or likely benign but support that it is likely pathogenic [77]. We were not able to estimate penetrance due to sample size and inability to perform extensive familial studies associating carrier status with affected cases as has recently been performed with *PALB2* risk [153]. Establishing risk is important in the context of familial aggregations of OC and host behaviours known to affect risk, such as has been shown with oral contraceptive pill use in carriers of pathogenic *BRCA1* and *BRCA2* variants. Risk-reducing surgery may not be necessary for *FANCI* variant carriers having significantly reduced risk due to oral contraceptive pill usage [154]. An investigation of carriers of candidate *FANCI* variants is also warranted given the intriguing observation of sensitivity to cisplatin but not to olaparib in the in cellulo studies of *FANCI* c.1813C>T, as this might impact the efficacy of PARPi's in the treatment of HGSC in these cases.

2.7 References

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2019. Ca Cancer J Clin. 2019; 69(1):7–34. https://doi.org/10.3322/caac.21551.
2. Lichtenstein P, Holm NV, Verkasalo PK, Iliadou A, Kaprio J, Koskenvuo M, et al. Environmental and heritable factors in the causation of cancer: analyses of cohorts of twins from Sweden, Denmark, and Finland. N Engl J Med. 2000;343(2):78–85.
3. Lynch JF, Butts M, Snyder C, Bewtra C, Lynch J, Godwin AK, et al. Hereditary ovarian cancer: molecular genetics, pathology, management, and heterogeneity. Mol Oncol. 2009;3(2):97–137.
4. Stratton JF, Pharoah P, Smith SK, Easton D, Ponder B. A systematic review and meta-analysis of family history and risk of ovarian cancer. Br J Obstet Gynaecol. 1998;105(5):493–9.
5. Kuchenbaecker KB, Hopper JL, Barnes DR, Phillips K, Mooij TM, Jervis S, et al. Risks of breast, ovarian, and contralateral breast cancer for BRCA1 and BRCA2 mutation carriers. JAMA. 2017;317(23):2402–16.
6. Peres LC, Cushing-haugen KL, Ko M, Harris HR, Berchuck A, Rossing MA, et al. Invasive epithelial ovarian cancer survival by histotype and disease stage. J Natl Cancer Inst. 2019;111(1):60–68, 1, DOI: https://doi.org/10.1093/ jnci/djy071.
7. Vang R, Shih LM, Kurman RJ. Ovarian low-grade and high grade serous carcinoma: pathogenesis, clinicopathologic and molecular biologic features, and diagnostic problems. Adv Anat Pathol. 2009;16(5):267–82.

8. Mucci LA, Hjelmborg JB, Harris JR, Czene K, Havelick DJ, Scheike T, et al. Familial risk and heritability of cancer among twins in nordic countries. JAMA. 2016;315(1):68–76. https://doi.org/10.1001/jama.2015.17703.
9. Finch APM, Lubinski J, Møller P, Singer CF, Karlan B, Senter L, et al. Impact of oophorectomy on cancer incidence and mortality in women with a BRCA1 or BRCA2 mutation. J Clin Oncol. 2014;32(15):1547–54.
10. Finch A, Beiner M, Lubinkski J, Lynch HT, Moller P, Rosen B, et al. Salpingo-oophorectomy and the risk of ovarian, fallopian tube, and peritoneal cancers in women with a BRCA1 or BRCA2 mutation. JAMA. 2006;296(2): 185–92.
11. Fong PC, Yap TA, Boss DS, Carden CP, Mergui-roelvink M, Gourley C, et al. Poly (ADP)-ribose polymerase inhibition: frequent durable responses in BRCA carrier ovarian cancer correlating with platinum-free interval. J Clin Oncol. 2010;28(15):2512–9.
12. Audeh MW, Carmichael J, Penson RT, Friedlander M, Powell B, Bell-mcguinn KM, et al. Oral poly (ADP-ribose) polymerase inhibitor olaparib in patients with BRCA1 or BRCA2 mutations and recurrent ovarian cancer: a proof-of- concept trial. Lancet Oncol. 2010;376:245–51.
13. Gelmon KA, Tischkowitz M, Mackay H, Swenerton K, Robidoux A, Tonkin K, et al. Olaparib in patients with recurrent high-grade serous or poorly differentiated ovarian carcinoma or triple-negative breast cancer: a phase 2 , multicentre , open-label , non-randomised study. Lancet Oncol. 2011;12: 852–61.

14. Kaufman B, Shapira-Frommer R, Schmutzler RK, Audeh MW, Friedlander M, Balmana J, et al. Olaparib monotherapy in patients with advanced cancer and a germline BRCA1/2 mutation. J Clin Oncol. 2015;33(3):244–50.
15. Pujade-lauraine E, Ledermann JA, Selle F, Gebski V, Penson RT, Oza AM, et al. Olaparib tablets as maintenance therapy in patients with platinum-sensitive, relapsed ovarian cancer and a BRCA1/2 randomised, placebo-controlled, phase 3 trial. Lancet Oncol. 2017;18(9):1274–84. https://doi.org/10.1016/S1470-2045(17)30469-2.
16. Ledermann J, Harter P, Gourley C, Friedlander M, Vergote I, Rustin G, et al. Olaparib maintenance therapy in patients with platinum-sensitive relapsed serous ovarian cancer: a preplanned retrospective analysis of outcomes by BRCA status in a randomised phase 2 trial. Lancet Oncol. 2014;15:852–61.
17. Song H, Ramus SJ, Kjaer SK, Hogdall E, Dicioccio RA, Whittemore AS, et al. Tagging single nucleotide polymorphisms in the BRIP1 gene and susceptibility to breast and ovarian cancer. PLoS One. 2007;2(3):e268.
18. Rafnar T, Gudbjartsson DF, Sulem P, Jonasdottir A, Sigurdsson A, Jonasdottir A, et al. Mutations in BRIP1 confer high risk of ovarian cancer. Nat Genet. 2011;43(11):1104–7.
19. Meindl A, Hellebrand H, Wiek C, Erven V, Wappenschmidt B, Niederacher D, et al. Germline mutations in breast and ovarian cancer pedigrees establish RAD51C as a human cancer susceptibility gene. Nat Genet. 2010;42(5):410– 4.

20. Zheng Y, Zhang J, Hope K, Niu Q, Huo D, Olopade OI. Screening RAD51C nucleotide alterations in patients with a family history of breast and ovarian cancer. Breast Cancer Res Treat. 2010;124:857–61.
21. Bagley CM, Young RC, Canellos GP, DeVita VT. Treatment of ovarian carcinoma: possibilities for progress. N Engl J Med. 1972 Oct 26;287(17):856–62. https://doi.org/10.1056/NEJM197210262871705.
22. Somyajit K, Subramanya S, Nagaraju G. RAD51C: A novel cancer susceptibility gene is linked to Fanconi anemia and breast cancer. Carcinogenesis. 2010;31(12):2031–8.
23. Loveday C, Turnbull C, Ramsay E, Hughes D, Ruark E, Frankum JR, et al. Germline mutations in RAD51D confer susceptibility to ovarian cancer. Nat Genet. 2011;43(9):879–82. https://doi.org/10.1038/ng.893.
24. Yang X, Song H, Leslie G, Engel C, Hahnen E, Auber B, et al. Ovarian and breast cancer risks associated with pathogenic variants in RAD51C and RAD51D. J Natl Cancer Inst. 2020;112(12):1242–50.
25. Scriver CR. Human genetics: Lessons from Quebec populations. Annu Rev Genomics Hum Genet. 2001;2(1):69–101. https://doi.org/10.1146/annurev.genom.2.1.69.
26. Laberge A-M, Michaud J, Richter A, Lemyre E, Lambert M, Brais B. Population history and its impact on medical genetics in Quebec. Clin Genet. 2005;68:287–301.

27. Roy-Gagnon M-H, Moreau C, Bherer C, St-onge P, Labuda D, Sinnett D, et al. Genomic and genealogical investigation of the French Canadian founder population structure. Hum Genet. 2011;129:521–31.
28. Bherer C, Labuda D, Roy-Gagnon M-H, Houde L, Tremblay M, Vezina H. Admixed ancestry and stratification of Quebec regional populations. Am J Phys Anthropol. 2011;144(3):432–41.
29. Fierheller CT, Alenezi WM, Tonin PN. The genetic analyses of French Canadians of Quebec facilitate the characterization of new cancer predisposing genes implicated in hereditary breast and/or ovarian cancer syndrome families. Cancers (Basel). 2021;13(14):3406.
30. Oros KK, Ghadirian P, Greenwood CMT, Perret C, Shen Z, Paredes Y, et al. Significant proportion of breast and/or ovarian cancer families of French Canadian descent harbor 1 of 5 BRCA1 and BRCA2 mutations. Int J Cancer. 2004;112:411–9.
31. Cline MS, Liao RG, Parsons MT, Paten B, Alquaddoomi F, Antoniou A, et al. BRCA Challenge: BRCA Exchange as a global resource for variants in BRCA1 and BRCA2. PLoS Genet. 2018;14(12):e1007752.
32. Tischkowitz M, Sabbaghian N, Hamel N, Pouchet C, Foulkes WD, Mes- Masson A-M, et al. Contribution of the PALB2 c.2323C>T [p. Q775X] Founder mutation in well-defined breast and/or ovarian cancer families and unselected ovarian cancer cases of French Canadian descent. BMC Med Genet. 2013;14:5.
33. Rivera B, Di Iorio M, Frankum J, Nadaf J, Fahiminiya S, Arcand SL, et al. Functionally null RAD51D missense mutation associates strongly with ovarian

carcinoma. Cancer Res. 2017;77(16):4517–29. https://doi.org/10.1158/ 0008-5472.CAN-17-0190.

34. Ishiai M, Kitao H, Smogorzewska A, Tomida J, Kinomura A, Uchida E, et al. FANCI phosphorylation functions as a molecular switch to turn on the Fanconi anemia pathway. Nat Struct Mol Biol. 2008;15(11):1138–46.

35. Nakanishi K, Yang Y, Pierce AJ, Taniguchi T, Digweed M, Andrea ADD, et al. Human Fanconi anemia monoubiquitination pathway promotes homologous DNA repair. PNAS. 2005;102(4):1110–5. https://doi.org/10.1073/ pnas.0407796102.

36. Sims AE, Spiteri E, Sims RJ, Arita AG, Lach FP, Landers T, et al. FANCI is a second monoubiquitinated member of the Fanconi anemia pathway. Nat Struct Mol Biol. 2007;14(6):564–7.

37. Smogorzewska A, Matsuoka S, Vinciguerra P, McDonald ER III, Hurov KE, Luo J, et al. Identification of the Fanconi anemia (FANC) I protein, a monoubiquitinated FANCD2 paralog required for crosslink repair. Cell. 2007; 129(2):289–301.

38. Dorsman JC, Levitus M, Rockx D, Rooimans MA, Oostra AB, Haitjema A, et al. Identification of the Fanconi anemia complementation group I gene. FANCI. Cell Oncol. 2007;29:211–8.

39. Belanger MH, Dolman L, Arcand SL, Shen Z, Chong G, Mes-Masson A-M, et al. A targeted analysis identifies a high frequency of BRCA1 and BRCA2 mutation carriers in women with ovarian cancer from a founder population. J Ovarian Res. 2015;8:1.

40. Tonin PN, Mes-Masson AM, Futreal PA, Morgan K, Mahon M, Foulkes WD, et al. Founder BRCA1 and BRCA2 mutations in French Canadian breast and ovarian cancer families. Am J Hum Genet. 1998;63:1341–51.
41. Tonin PN, Mes-Masson A-M, Narod SA, Ghadirian P, Provencher D. Founder BRCA1 and BRCA2 mutations in French Canadian ovarian cancer cases unselected for family history. Clin Genet. 1999 May;55(5):318–24.
42. Behl S, Hamel N, de Ladurantaye M, Lepage S, Lapointe R, Mes-Masson A-M, et al. Founder BRCA1/BRCA2/PALB2 pathogenic variants in French-Canadian breast cancer cases and controls. Sci Rep. 2020;10:6491.
43. Awadalla P, Boileau C, Payette Y, Idaghdour Y, Goulet J, Knoppers B, et al. Cohort profile of the CARTaGENE study: Quebec's population-based biobank for public health and personalized genomics. Int J Epidemiol. 2013; 42:1285–99.
44. Phelan CM, Kuchenbaecker KB, Tyrer JP, Kar SP, Lawrenson K, Winham SJ. Identification of 12 new susceptibility loci for different histotypes of epithelial ovarian cancer. Nat Genet. 2017;49(5):680–91.
45. Michailidou K, Hall P, Gonzalez-neira A, Ghoussaini M, Milne RL, Schmidt MK, et al. Large-scale genotyping identified 41 new loci associated with breast cancer risk. Nat Genet. 2013;45(4):353–61.
46. Michailidou K, Beesley J, Lindstrom S, Canisius S, Dennis J, Lush M, et al. Genome-wide association analysis of more than 120,000 individuals identifies 15 new susceptibility loci for breast cancer. Nat Genet. 2015;47(4):373–80.
47. Subramanian DN, Zethoven M, Mcinerny S, Morgan JA, Rowley SM, Lee JEA, et al. Exome sequencing of familial high-grade serous ovarian carcinoma reveals

heterogeneity for rare candidate susceptibility genes. Nat Commun. 2020;11:1640.

48. Rowley SM, Mascarenhas L, Devereux L, Li N, Amarasinghe KC, Zethoven M, et al. Population-based genetic testing of asymptomatic women for breast and ovarian cancer susceptibility. Genet Med. 2018;21(4):913–22. https://doi.org/10.1038/s41436-018-0277-0.

49. Berry BDA, Iversen ES, Gudbjartsson DF, Hiller EH, Garber JE, Peshkin BN, et al. BRCAPRO validation, sensitivity of genetic testing of BRCA1/BRCA2, and prevalence of other breast cancer susceptibility genes. J Clin Oncol. 2002;20(11):2701–12.

50. Goldman MJ, Craft B, Hastie M, Repečka K, McDade F, Kamath A, et al. Visualizing and interpreting cancer genomics data via the Xena platform. Nature biotechnology. 2020;38:675–8.

51. Yang H, Wang K. Genomic variant annotation and prioritization with ANNOVAR and wANNOVAR. Nat Protoc. 2015;10(10):1556–66. https://doi.org/10.1038/nprot.2015.105.

52. Sim N, Kumar P, Hu J, Henikoff S, Schneider G, Ng PC. SIFT web server: predicting effects of amino acid substitutions on proteins. Nucleic Acids Res. 2012;40:452–7.

53. 1000 Genomes Project Consortium, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, et al. A global reference for human genetic variation. Nature. 2015;526(7571):68–74.

54. Exome Variant Server, NHLBI GO Exome Sequencing Project (ESP) [Internet]. Seattle, WA. [cited 2014 Feb 21]. Available from: http://evs.gs.washington.edu/EVS/
55. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative Genomics Viewer. Nat Biotechnol. 2011;29(1):24–6. https://doi.org/10.1038/nbt.1754.
56. Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, et al. Analysis of protein-coding genetic variation in 60,706 humans. Nature. 2016; 536:285–91.
57. Knijnenburg TA, Wang L, Zimmermann MT, Chambwe N, Gao GF, Cherniak AD, et al. Genomic and molecular landscape of DNA damage repair deficiency across The Cancer Genome Atlas resource. Cell Rep. 2018;23:239–54.
58. Fedick A, Su J, Jalas C, Northrop L, Devkota B, Ekstein J, et al. High- throughput carrier screening using TaqMan allelic discrimination. PLoS One. 2013;8(3):e59722.
59. Mccarthy S, Das S, Kretzschmar W, Delaneau O, Wood AR, Teumer A, et al. A reference panel of 64,976 haplotypes for genotype imputation. Nat Genet. 2016;48(10):1279–83. https://doi.org/10.1038/ng.3643.
60. Loh P, Danecek P, Palamara PF, Fuchsberger C, Reshef YA, Finucane HK, et al. Reference-based phasing using the Haplotype Reference Consortium panel. Nat Genet. 2016;48(11):1443–8.
61. Durbin R. Efficient haplotype matching and storage using the positional Burrows-Wheeler transform (PBWT). Bioinformatics. 2014;30(9):1266–72.

62. Ghosh R, Oak N, Plon SE. Evaluation of in silico algorithms for use with ACMG/AMP clinical variant interpretation guidelines. Genome Biol. 2017 28; 18(255), 1, DOI: https://doi.org/10.1186/s13059-017-1353-5.
63. Davydov EV, Goode DL, Sirota M, Cooper GM, Sidow A. Identifying a high fraction of the human genome to be under selective constraint using GERP ++. PLoS Comput Biol. 2010;6(12):e1001025.
64. Pollard KS, Hubisz MJ, Rosenbloom KR, Siepel A. Detection of nonneutral substitution rates on mammalian phylogenies. Genome Res. 2010;20(1):110–21
65. Siepel A, Bejerano G, Pedersen JS, Hinrichs AS, Hou M, Rosenbloom K, et al. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. Genome Res. 2005;15(8):1034–50. https://doi.org/10.1101/gr.3715005.
66. Garber M, Guttman M, Clamp M, Zody MC, Friedman N, Xie X. Identifying novel constrained elements by exploiting biased substitution patterns. Bioinformatics. 2009;25(12):i54–62.
67. Rentzsch P, Witten D, Cooper GM, Shendure J, Kircher M. CADD: predicting the deleteriousness of variants throughout the human genome. Nucleic Acids Res. 2019;47(D1):D886–94. https://doi.org/10.1093/nar/gky1016.
68. González-Pérez A, López-Bigas N. Improving the assessment of the outcome of nonsynonymous SNVs with a consensus deleteriousness score. Condel. Am J Hum Genet. 2011;88(4):440–9.

69. Ionita-Laza I, McCallum K, Xu B, Buxbaum JD. A spectral approach integrating functional genomic annotations for coding and noncoding variants. Nat Genet. 2016;48(2):214–20.
70. Dong C, Wei P, Jian X, Gibbs R, Boerwinkle E, Wang K, et al. Comparison and integration of deleteriousness prediction methods for nonsynonymous SNVs in whole exome sequencing studies. Hum Mol Genet. 2015;24(8):2125–37.
71. Douville C, Masica DL, Stenson PD, Cooper DN, Gygax DM, Kim R, et al. Assessing the pathogenicity of insertion and deletion variants with the variant effect scoring tool (VEST-Indel). Hum Mutat. 2016;37(1):28–35.
72. Ioannidis NM, Rothstein JH, Pejaver V, Middha S, McDonnell SK, Baheti S, et al. REVEL: an ensemble method for predicting the pathogenicity of rare missense variants. Am J Hum Genet. 2016;99(4):877–85. https://doi.org/10.1 016/j.ajhg.2016.08.016.
73. Shamsani J, Kazakoff SH, Armean IM, McLaren W, Parsons MT, Thompson BA, et al. A plugin for the Ensembl Variant Effect Predictor that uses MaxEntScan to predict variant spliceogenicity. Bioinformatics. 2019;35(13): 2315–7.
74. Jian X, Boerwinkle E, Liu X. In silico prediction of splice-altering single nucleotide variants in the human genome. Nucleic Acids Res. 2014;42(22): 13534–44.
75. Mclaren W, Gil L, Hunt SE, Riat HS, Ritchie GRS, Thormann A, et al. The Ensembl Variant Effect Predictor. Genome Biol. 2016;17(122) https://doi. org/10.1186/s13059-016-0974-4.
76. Huang K-L, Mashl RJ, Wu Y, Ritter DI, Wang J, Oh C, et al. Pathogenic germline variants in 10,389 adult cancers. Cell. 2018;173(2):355–370.e14.

77. Landrum MJ, Lee JM, Benson M, Brown GR, Chao C, Chitipiralla S, et al. ClinVar: Improving access to variant interpretations and supporting evidence. Nucleic Acids Res. 2018;46:1062–7.
78. Medrano M, Communal L, Brown KR, Iwanicki M, Normand J, Paterson J, et al. Interrogation of functional cell-surface markers identifies CD151 dependency in high-grade serous ovarian cancer. Cell Rep. 2017;18:2343–58.
79. Tonin PN, Maugard CM, Perret C, Mes-Masson AM, Provencher DM. A review of histopathological subtypes of ovarian cancer in BRCA-related French Canadian cancer families. Fam Cancer. 2007;6:491–7.
80. Evans DGR, Eccles DM, Rahman N, Young K, Bulman M, Amir E, et al. A new scoring system for the chances of identifying a BRCA1/2 mutation outperforms existing models including BRCAPRO. J Med Genet. 2004;41:474–80.
81. Evans DG, Harkness EF, Plaskocinska I, Wallace AJ, Clancy T, Woodward ER, et al. Pathology update to the Manchester Scoring System based on testing in over 4000 families. J Med Genet. 2017;54:674–81.
82. Oros KK, Ghadirian P, Maugard CM, Perret C, Paredes Y, Mes-Masson A-M, et al. Application of BRCA1 and BRCA2 mutation carrier prediction models in breast and/or ovarian cancer families of French Canadian descent. Clin Genet. 2006;70(4):320–9. https://doi.org/10.1111/j.1399-0004.2006.00673.x.
83. Tan SLW, Chadha S, Liu Y, Gabasova E, Perera D, Ahmed K, et al. A class of environmental and endogenous toxins induces BRCA2 haploinsufficiency and genome instability. Cell [Internet]. 2017;169(6):1105–1118.e15 Available from: https://pubmed.ncbi.nlm.nih.gov/28575672.

84. Kim Y, Spitz GS, Veturi U, Lach FP, Auerbach AD, Smogorzewska A. Regulation of multiple DNA repair pathways by the Fanconi anemia protein SLX4. Blood. 2013;121(1):54–64. https://doi.org/10.1182/blood-2012- 07-441212.
85. Tonin PN, Perret C, Lambert JA, Paradis AJ, Kantemiroff T, Benoît MH, et al. Founder BRCA1 and BRCA2 mutations in early-onset French Canadian breast cancer cases unselected for family history. Int J Cancer. 2001;95(3): 189–93. https://doi.org/10.1002/1097-0215(20010520)95:3<189::AID-IJC1032 >3.0.CO;2-N.
86. Arcand SL, Akbari MR, Provencher D, Foulkes WD, Narod SA, Tonin PN. Germline TP53 mutational spectrum in French Canadians with breast cancer. BMC Med Genet. 2015;16(1):24. https://doi.org/10.1186/s12881-015- 0169-y.
87. Novak DJ, Chen LQ, Ghadirian P, Hamel N, Zhang P, Rossiny V, et al. Identification of a novel CHEK2 variant and assessment of its contribution to the risk of breast cancer in French Canadian women. BMC Cancer. 2008;8(1): 239. https://doi.org/10.1186/1471-2407-8-239.
88. Tischkowitz M, Xia B, Sabbaghian N, Reis-filho JS, Hamel N, Li G, et al. Analysis of PALB2/FANCN-associated breast cancer families. PNAS. 2007; 104(16):6788–93. https://doi.org/10.1073/pnas.0701724104.
89. Cote S, Arcand SL, Royer R, Nolet S, Mes-Masson A, Ghadirian P, et al. The BRCA2 c.9004G>A (E2002K) [corrected] variant is likely pathogenic and recurs in breast and/or ovarian cancer families of French Canadian descent. Breast Cancer Res Treat. 2012 Jan;131(1):333–40. https://doi.org/10.1007/s1 0549-011-1796-4.

90. Osher DJ, De Leeneer K, Michils G, Hamel N, Tomiak E, Poppe B, et al. Mutation analysis of RAD51D in non-BRCA1/2 ovarian and breast cancer families. Br J Cancer. 2012;106(8):1460–3. https://doi.org/10.1038/bjc.2012.87.
91. Godard B, Foulkes WD, Provencher D, Brunet JS, Tonin PN, Mes-Masson AM, et al. Risk factors for familial and sporadic ovarian cancer among French Canadians: a case-control study. Am J Obstet Gynecol. 1998;179(2):403–10. https://doi.org/10.1016/S0002-9378(98)70372-2.
92. Oros KK, Leblanc G, Arcand SL, Shen Z, Perret C, Foulkes WD, et al. Haplotype analysis suggest common founder in carriers of recurrent BRCA2 mutation, 3398delAAAAG, in French Canadian hereditary breast and/ovarian cancer families. BMC Med Genet. 2006;7:23.
93. Arcand SL, Maugard CM, Ghadirian P, Robidoux A, Perret C, Zhang P, et al. Germline TP53 mutations in BRCA1 and BRCA2 mutation-negative French Canadian breast cancer families. Breast Cancer Res Treat. 2008;108(3):399–408.
94. Cavallone L, Arcand SL, Maugard C, Ghadirian P, Provencher D, Tonin PN. Haplotype analysis of TP53 polymorphisms, Arg72Pro and Ins16, in BRCA1 and BRCA2 mutation carriers of French Canadian descent. BMC Cancer. 2008;8:96.
95. Cavallone L, Arcand SL, Maugard CM, Nolet S, Gaboury LA, Mes-Masson AM, et al. Comprehensive BRCA1 and BRCA2 mutation analyses and review of French Canadian families with at least three cases of breast cancer. Fam Cancer. 2010;9:507–17.

96. Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. Nature. 2020;581(7809):434–43. https://doi.org/10.1038/s41586-020-2308-7.

97. Matulonis UA, Sood AK, Fallowfield L, Howitt BE, Sehouli J, Karlan BY. Ovarian cancer. Nat Rev Dis Prim. 2016;2:16061.

98. Reid BM, Permuth JB, Sellers TA. Epidemiology of ovarian cancer: a summary review. Cancer Biol Med. 2017;14(1):9–32.

99. PDQ Cancer Genetics Editorial Board. Genetics of breast and gynecologic cancers (PDQ®): health professional version. PDQ Cancer Information Summaries. Bethesda, MD: National Cancer Institute (US);

100. Kindelberger DW, Lee Y, Miron A, Hirsch MS, Feltmate C, Medeiros F, et al. Intraepithelial carcinoma of the fimbria and pelvic serous carcinoma: evidence for a causal relationship. Am J Surg Pathol. 2007;31(2):161–9.

101. Callahan MJ, Crum CP, Medeiros F, Kindelberger DW, Elvin JA, Garber JE, et al. Primary fallopian tube malignancies in BRCA-positive women undergoing surgery for ovarian cancer risk reduction. J Clin Oncol. 2007; 25(25):3985–90.

102. Carlson JW, Miron A, Jarboe EA, Parast MM, Hirsch MS, Lee Y, et al. Serous tubal intraepithelial carcinoma: its potential role in primary peritoneal serous carcinoma and serous cancer prevention. J Clin Oncol. 2008;26(25):4160–5.

103. Finch A, Shaw P, Rosen B, Murphy J, Narod SA, Colgan TJ. Clinical and pathologic findings of prophylactic salpingo-oophorectomies in 159 BRCA1 and BRCA2 carriers. Gynecol Oncol. 2006;100:58–64.

104. Labidi-galy SI, Papp E, Hallberg D, Niknafs N, Adleff V, Noe M, et al. High grade serous ovarian carcinomas originate in the fallopian tube. Nat Commun. 2017;8:1093.

105. Leeper K, Garcia R, Swisher E, Goff B, Greer B, Paley P. Pathologic findings in prophylactic oophorectomy specimens in high-risk women. Gynecol Oncol. 2002;87:52–6.

106. Powell CB, Kenley E, Chen L, Crawford B, Mclennan J, Zaloudek C, et al. Risk- reducing salpingo-oophorectomy in BRCA mutation carriers: role of serial sectioning in the detection of occult malignancy. J Clin Oncol. 2005;23(1): 127–32.

107. Hao D, Li J, Jia S, Meng Y, Zhang C, Wang L, et al. Integrated analysis reveals tubal- and ovarian-originated serous ovarian cancer and predicts differential therapeutic responses. Clin Cancer Res. 2017;23(23):7400–11.

108. Lo Riso P, Villa CE, Gasparoni G, Vingiani A, Luongo R, Manfredi A, et al. A cell-of-origin epigenetic tracer reveals clinically distinct subtypes of high- grade serous ovarian cancer. Genome Med. 2020;12(1):94.

109. Zhang S, Dolgalev I, Zhang T, Ran H, Levine DA, Neel BG. Both fallopian tube and ovarian surface epithelium are cells-of-origin for high-grade serous ovarian carcinoma. Nat Commun. 2019;10(1):5367. https://doi.org/10.1038/s41467-019-13116-2.

110. Foulkes WD, Ghadirian P, Akbari MR, Hamel N, Giroux S, Sabbaghian N, et al. Identification of a novel truncating PALB2 mutation and analysis of its contribution to early-onset breast cancer in French-Canadian women. Breast Cancer Res. 2007;9:R83.

111. Castellsagué E, Liu J, Volenik A, Giroux S, Gagné R, Maranda B, et al. Characterization of a novel founder MSH6 mutation causing Lynch syndrome in the French Canadian population. Clin Genet. 2015;87(6):536– 42. https://doi.org/10.1111/cge.12526.

112. Vezina H, Durocher F, Dumont M, Houde L, Szabo C, Tranchant M, et al. Molecular and genealogical characterization of the R1443X BRCA1 mutation in high-risk French-Canadian breast/ovarian cancer families. Hum Genet. 2005;117(2-3):119–32. https://doi.org/10.1007/s00439-005-1297-9.

113. Manning AP, Abelovich D, Ghadirian P, Lambert JA, Frappier D, Provencher D, et al. Haplotype analysis of BRCA2 8765delAG mutation carriers in French Canadian Yemenite Jewish hereditary breast cancer families. Hum Hered. 2001;52(2):116–20. https://doi.org/10.1159/000053364.

114. Wojnarowicz PM, Oros KK, Quinn MCJ, Arcand SL, Gambaro K, Madore J, et al. The genomic landscape of TP53 and p53 annotated high grade ovarian serous carcinomas from a defined founder population associated with patient outcome. PLoS One. 2012;7(9):e45484.

115. Bell D, Berchuck A, Birrer M, Chien J, Cramer DW, Dao F, et al. Integrated genomic analyses of ovarian carcinoma. Nature. 2011;474(7353):609–15.

116. Maxwell KN, Wubbenhorst B, Wenz BM, De Sloover D, Pluta J, Emery L, et al. BRCA locus-specific loss of heterozygosity in germline BRCA1 and BRCA2 carriers. Nat Commun. 2017;8:319.

117. Vahteristo P, Bartkova J, Eerola H, Syrja K, Ojala S, Kilpivaara O, et al. A CHEK2 genetic variant contributing to a substantial fraction of familial breast cancer. Am J Hum Genet. 2002;71:432–8.

118. Mavaddat N, Michailidou K, Dennis J, Lush M, Fachal L, Lee A, et al. Polygenic risk scores for prediction of breast cancer and breast cancer subtypes. Am J Hum Genet. 2019;104(1):21–34.

119. Song H, Dicks EM, Tyrer J, Intermaggio M, Chenevix-Trench G, Bowtell DD, et al. Population-based targeted sequencing of 54 candidate genes identifies PALB2 as a susceptibility gene for high-grade serous ovarian cancer. J Med Genet. 2021;58(5):305–13.

120. Lin P-H, Kuo W-H, Huang A-C, Lu Y-S, Lin C-H, Kuo S-H, et al. Multiple gene sequencing for risk assessment in patients with early-onset or familial breast cancer. Oncotarget. 2016;7(7):8310–20.

121. Girard E, Eon-Marchais S, Olaso R, Renault A, Dondon M, Barjhoux L, et al. Familial breast cancer and DNA repair genes: insights into known and novel susceptibility genes from the GENESIS study, and implications for multigene panel testing. Int J Cancer. 2018;144(8):1962–74.

122. Garcia MJ, Fernandez V, Osorio A, Barroso A, Fernandez F, Urioste M, et al. Mutational analysis of FANCL, FANCM and the recently identified FANCI suggests that among the 13 known Fanconi anemia genes, only FANCD1/

BRCA2 plays a major role in high-risk breast cancer predisposition. Carcinogenesis. 2009;30(11):1898–902. https://doi.org/10.1093/carcin/ bgp218.

123. Ellingson MS, Hart SN, Kalari KR, Suman V, Schahl KA, Dockter TJ, et al. Exome sequencing reveals frequent deleterious germline variants in cancer susceptibility genes in women with invasive breast cancer undergoing neoadjuvant chemotherapy. Breast Cancer Res Treat. 2015;153:435–43.

124. Yang X, Wu J, Lu J, Liu G, Di G, Chen C. Identification of a comprehensive spectrum of genetic factors for hereditary breast cancer in a Chinese population by next- generation sequencing. PLoS One. 2015;10(4):e0125571.

125. Mantere T, Haanpää M, Hanenberg H, Schleutker J, Kallioniemi A, Kahkonen M, et al. Finnish Fanconi anemia mutations and hereditary predisposition to breast and prostate cancer. Clin Genet. 2015;88:68–73.

126. Bonache S, Esteban I, Moles-Fernández A, Tenés A, Duran-Lozano L, Montalban G, et al. Multigene panel testing beyond BRCA1/2 in breast/ ovarian cancer Spanish families and clinical actionability of findings. J Cancer Res Clin Oncol. 2018;144(12):2495–513.

127. Tedaldi G, Tebaldi M, Zampiga V, Danesi R, Arcangeli V, Ravegnani M, et al. Multiple-gene panel analysis in a case series of 255 women with hereditary breast and ovarian cancer. Oncotarget. 2017;8(29):47064–75.

128. Paulo P, Maia S, Pinto C, Pinto P, Monteiro A, Peixoto A, et al. Targeted next generation sequencing identifies functionally deleterious germline mutations

in novel genes in early-onset/familial prostate cancer. PLoS Genet. 2018;14(4):e1007355.

129. Velho PI, Qazi F, Hassan S, Carducci MA, Denmeade SR, Markowski MC, et al. Efficacy of radium-223 in bone-metastatic castration-resistant prostate cancer with and without homologous repair gene defects. Eur Urol. 2018; 76(2):170–6. https://doi.org/10.1016/j.eururo.2018.09.040.

130. Chan SH, Lim WK, Diana N, Ishak B, Li S, Goh WL, et al. Germline mutations in cancer predisposition genes are frequent in sporadic sarcomas. Sci Rep. 2017;7(1):10660.

131. Betti M, Casalone E, Ferrante D, Aspesi A, Morleo G, Biasi A, et al. Germline mutations in DNA repair genes predispose asbestos-exposed patients to malignant pleural mesothelioma. Cancer Lett. 2017;405:38–45. https://doi. org/10.1016/j.canlet.2017.06.028.

132. Maung KZY, Leo PJ, Bassal M, Casolari DA, Gray JX, Bray SC, et al. Rare variants in Fanconi anemia genes are enriched in acute myeloid leukemia. Blood Cancer J. 2018;8:50.

133. Chandrasekharappa SC, Chinn SB, Donovan FX, Chowdhury NI, Kamat A, Adeyemo AA, et al. Assessing the spectrum of germline variation in Fanconi anemia genes among patients with head and neck carcinoma before age 50. Cancer. 2017;123(20):3943–54.

134. Zhunussova G, Afonin G, Abdikerim S, Jumanov A, Perfilyeva A, Kaidarova D, et al. Mutation spectrum of cancer-associated genes in patients with early onset of colorectal cancer. Front Oncol. 2019;9:673.

135. Castella M, Jacquemont C, Thompson EL, Yeo JE, Cheung RS, Huang JW, et al. FANCI regulates recruitment of the FA core complex at sites of DNA damage independently of FANCD2. PLoS Genet. 2015;11(10):e1005563.

136. Farmer H, McCabe N, Lord CJ, Tutt ANJ, Johnson DA, Richardson TB, et al. Targeting the DNA repair defect in BRCA mutant cells as a therapeutic strategy. Nature. 2005;434:917–21.

137. Wiltshire T, Ducy M, Foo TK, Hu C, Lee KY, Nagaraj AB, et al. Functional characterization of 84 PALB2 variants of uncertain significance. Genet Med. 2019;22(3):622–32. https://doi.org/10.1038/s41436-019-0682-z.

138. Rodrigue A, Margaillan G, Gomes TT, Coulombe Y, Montalban G, Carvalho S, et al. A global functional analysis of missense mutations reveals two major hotspots in the PALB2 tumor suppressor. Nucleic Acids Res. 2019;47(20): 10662–77.

139. Chatterjee N, Lin Y, Wilson JH. Fanconi anemia pathway regulates convergent trancription-induced cell death at trinucleotide repeats in human cells. Postdoc J. 2016;4(5):46–54. https://doi.org/10.14304/surya. jpr.v4n5.1.

140. Chen Y, Jones MJK, Yin Y, Crist SB, Colnaghi L, Sims RJ III, et al. ATR-mediated phosphorylation of FANCI regulates dormant origin firing in response to replication stress. Mol Cell. 2016;58(2):323–38.

141. Sato K, Shimomuki M, Katsuki Y, Takahashi D, Kobayashi W, Ishiai M, et al. FANCI-FANCD2 stabilizes the RAD51-DNA complex by binding RAD51 and protects the 5’-DNA end. Nucleic Acids Res. 2016;44(22): 10758–71.

142. Zhang X, Lu X, Akhter S, Georgescu M, Legerski RJ. FANCI is a negative regulator of Akt activation. Cell Cycle. 2016;15(8):1134–43. https://doi.org/10.1080/15384101.2016.1158375.

143. Sondalle SB, Longerich S, Ogawa LM, Sung P, Baserga SJ. Fanconi anemia protein FANCI functions in ribosome biogenesis. PNAS. 2019;116(7):2561–70.

144. Mehta PA, Tolar J. Fanconi Anemia. In: Pagon R, Adam M, Ardinger H, editors. GeneReviews(®) [Internet] [Internet]. Seattle (WA): University of Washington, Seattle; 2002. Available from: http://www.ncbi.nlm.nih.gov/pubmed/9590114

145. Savage SA, Ballew BJ, Giri N, Dceg NCI, Genomics C, Chandrasekharappa SC, et al. Novel FANCI mutations in Fanconi anemia with VACTERL Association. Am J Med Genet Part A. 2015;170A(2):386–91. https://doi.org/10.1002/a jmg.a.37461.

146. Bottega R, Nicchia E, Cappelli E, Ravera S, De Rocco D, Faleschini M, et al. Hypomorphic FANCA mutations correlate with mild mitochondrial and clinical phenotype in Fanconi anemia. Haematologica. 2018;103(3):417–26. https://doi.org/10.3324/haematol.2017.176131.

147. Keupp K, Hampp S, Hübbel A, Maringa M, Kostezka S, Rhiem K, et al. Biallelic germline BRCA1 mutations in a patient with early onset breast cancer, mild Fanconi anemia-like phenotype, and no chromosome fragility. Mol Genet Genomic Med. 2019;7(9):e863.

148. Byrd PJ, Stewart GS, Smith A, Eaton C, Taylor AJ, Guy C, et al. A hypomorphic PALB2 allele gives rise to an unusual form of FA-N associated with lymphoid tumour development. PLoS Genet. 2016;12(3):e1005945.

149. Caburet S, Heddar A, Dardillac E, Creux H, Lambert M, Messiaen S, et al. Homozygous hypomorphic BRCA2 variant in primary ovarian insufficiency without cancer or Fanconi anaemia trait. J Med Genet. 2020. https://doi. org/10.1136/jmedgenet-2019-106672.

150. Lohmann DR, Gallie BL. Retinoblastoma. In: Adam MP, Ardinger HH, Pagon RA, Wallace SE, Bean LJH, Stephens K, et al., editors. Seattle (WA); 1993.

151. Chandrasekharappa SC, Lach FP, Kimble DC, Kamat A, Teer JK, Donovan FX, et al. Massively parallel sequencing , aCGH , and RNA-Seq technologies provide a comprehensive molecular diagnosis of Fanconi anemia. Blood. 2013;121(22):e138–48.

152. Dubois EL, Guitton-Sert L, Béliveau M, Parmar K, Chagraoui J, Vignard J, et al. A Fanci knockout mouse model reveals common and distinct functions for FANCI and FANCD2. Nucleic Acids Res. 2019;47(14):7532–47. https://doi. org/10.1093/nar/gkz514.

153. Yang X, Leslie G, Doroszuk A, Schneider S, Allen J, Decker B, et al. Cancer risks associated with germline PALB2 pathogenic variants: an international study of 524 families. J Clin Oncol. 2020;38(7):674–85.

154. Moorman PG, Havrilesky LJ, Gierisch JM, Coeytaux RR, Lowery WJ, Urrutia RP, et al. Oral contraceptives and risk of ovarian cancer and breast

cancer among high-risk women: a systematic review and meta-analysis. J Clin Oncol. 2013;31(33):4188–98.

2.8 Main figures and tables

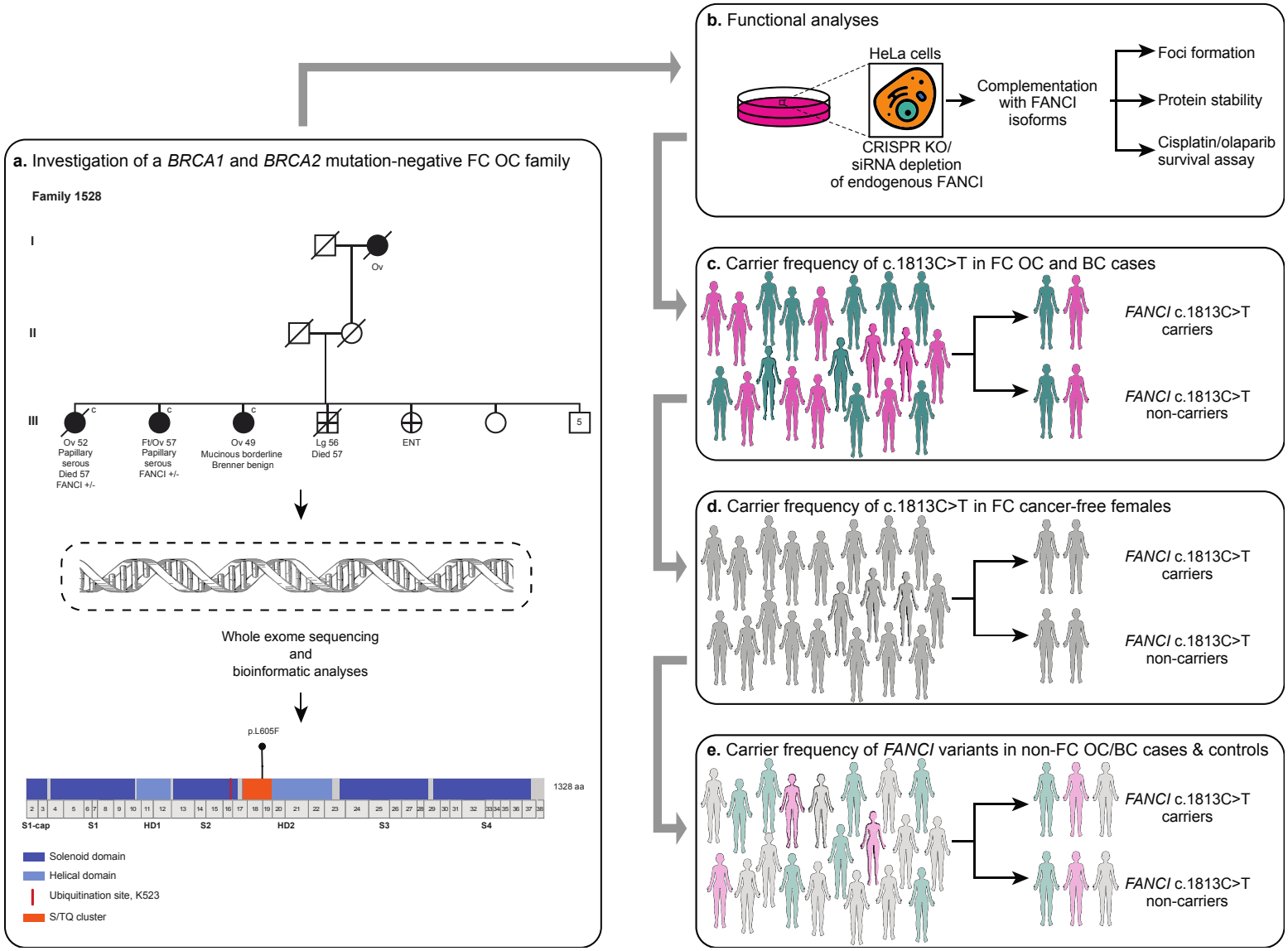


Figure 2.1. Study design for discovery and investigation of *FANCI* variants. **a** Pedigree F1528, a rare FC family with four cases of OC, in which *FANCI* c.1813C>T; p.L605F was discovered. WES was performed on the sisters, Ov 52 and FtOv 57 in generation III, who are *BRCA1* and *BRCA2* pathogenic variant negative. Cancer type (Ov: ovarian, Ft: fallopian tube, Lg: lung, and ENT: ear, nose, throat) and age of diagnosis are shown; c next to a symbol denotes a confirmed cancer case. The location of p.L605F is shown (bottom). Solenoid domain: antiparallel pairs of α-helices that form α-α superhelix segments; Helical domain: α-helices; Ubiquitination site, K523: site of monoubiquitination by the FA core complex to allow downstream FA pathway function [36, 37]; S/TQ cluster: location of conserved phosphorylation sites [34]. **b** Functional analyses of FANCI isoforms using HeLa cells. **c–e** Estimation of *FANCI* c.1813C>T; p.L605F carrier frequency in cases and controls. FANCI domains were adapted from pfam (https://pfam.xfam.org). *FANCI* exon locations adapted from University of California Santa Cruz Genome Browser (https://genome.ucsc.edu)

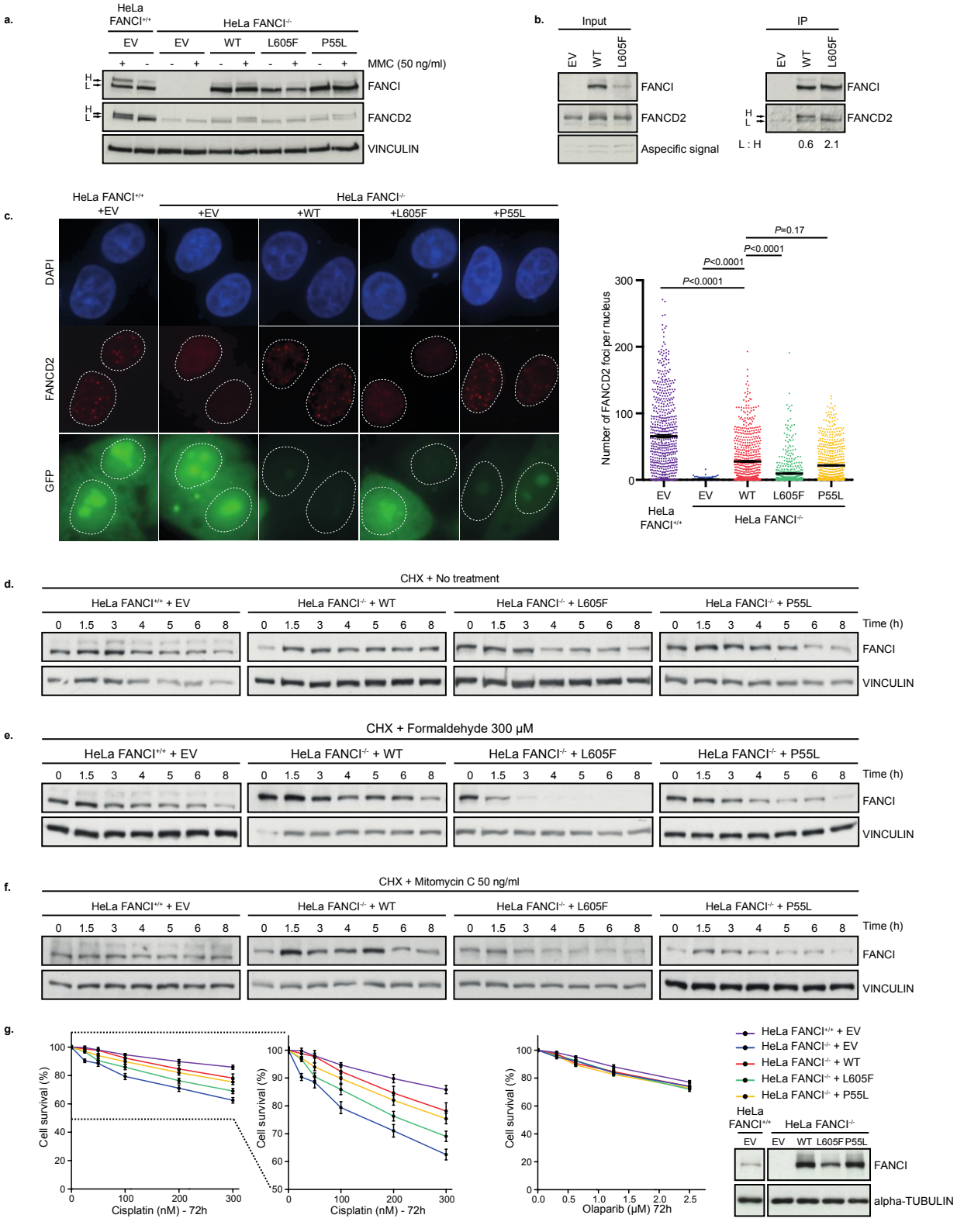

a.
HeLa FANCI+/+
HeLa FANCI-/-
EV
EV
WT
L605F
P55L
MMC (50 ng/ml)
FANCI
FANCD2
VINCULIN
b.
Input
IP
EV
WT
L605F
FANCI
FANCD2
Aspecific signal
L : H
0.6
2.1
c.
HeLa FANCI+/+ +EV
HeLa FANCI-/-
+EV
+WT
+L605F
+P55L
DAPI
FANCD2
GFP
Number of FANCD2 foci per nucleus
P=0.17
P<0.0001
P<0.0001
P<0.0001
d.
CHX + No treatment
HeLa FANCI+/+ + EV
HeLa FANCI-/- + WT
HeLa FANCI-/- + L605F
HeLa FANCI-/- + P55L
Time (h)
FANCI
VINCULIN
e.
CHX + Formaldehyde 300 µM
f.
CHX + Mitomycin C 50 ng/ml
g.
Cell survival (%)
Cisplatin (nM) - 72h
Olaparib (µM) 72h
HeLa FANCI+/+ + EV
HeLa FANCI-/- + EV
HeLa FANCI-/- + WT
HeLa FANCI-/- + L605F
HeLa FANCI-/- + P55L
alpha-TUBULIN

Figure 2.2. The isoform with the p.L605F variant impairs FANCI stability and function. **a** Western blots of HeLa cells with the *FANCI* gene (FANCI$^{+/+}$) or with the *FANCI* gene knocked out (FANCI$^{-/-}$). HeLa FANCI$^{-/-}$ cells from clone 1 were complemented with constructs of Flag-FANCI wild type (WT), p.L605F or p.P55L, or an empty vector (EV) and treated with 50 ng/ml MMC for 18 h. The upper band, H, shows the ubiquitination of FANCI and FANCD2 after treatment. The lower band, L, corresponds to non-ubiquitinated FANCI or FANCD2. VINCULIN was used as a loading control. Experiment was repeated three times. **b** HeLa FANCI$^{+/+}$ cells were transfected with siRNA targeting FANCI and then complemented with Flag-FANCI siRNA-resistant constructs or an EV. Cells were treated with 50 ng/ml MMC for 18 h followed by FLAG immunoprecipitation. The left panel shows FANCI constructs expression and the right panel the immunoprecipitated fractions. The p.L605F immunoprecipitation fraction sample was super-loaded to have the same signal after FANCI WT complementation. The ratio between the upper band (H) and lower band (L) for the immunoprecipitated FANCD2 is shown. **c** Immunofluorescence of HeLa FANCI$^{-/-}$ cells from clone 1 that were complemented with constructs of Flag-FANCI and 0.1 µg of empty GFP vector was used as a transfection control. The adjacent scatter plot shows the number of FANCD2 foci in GFP-positive cells after treatment with MMC (50 ng/ ml, 18 h). Mean with SEM is represented. The Kruskal-Wallis test was used to compare groups and the P value is shown for each test. Experiment has been performed in triplicate. **d–f** Western blot analysis of HeLa FANCI$^{-/-}$ cells from clone 1 that were complemented with constructs of Flag-FANCI and treated with cycloheximide (CHX) and either mock-treated (**d**) or treated with damaging agents formaldehyde (**e**) or MMC (**f**) for different lengths of time

at the indicated concentrations. At each time point, whole cell extracts were analysed by western blot to assess protein levels. Experiment has been done in triplicate. **g** Survival curves of HeLa FANCI$^{-/-}$ cells from clone 1 that were transfected with the different constructs of Flag-FANCI. Cell viability was monitored following cisplatin or olaparib treatments for 72 h and was assessed by counting remaining nuclei. Curves represent mean with SEM of three biological replicates. Western blots were used to monitor expression and shown here as an example. Alpha-tubulin was used as a loading control. Full blots are shown in Additional file 4

Table 2.1 Comparison of *FANCI* c.1813C>T carrier frequencies in cancer cases with French Canadian cancer-free women. All odds ratios are calculated comparing to cancer-free females.

Study group[1]	*BRCA1 and BRCA2* mutation status[1]	Case tested	Number of subjects	Number of c.1813C>T carriers (%)	OR	95% CI	*P*
OC families[2]	All	OC	42	3 (7.1)	3	0.9 – 9.9	0.073
	Negative		22	2[3] (9.1)	3.9	0.89 –17	0.071
	BRCA1 positive		14	1 (7.1)	3	0.39 –23	0.29
	BRCA2 positive		6	0	NA	NA	NA
Sporadic OC cases	All	OC	439	7 (1.6)	0.63	0.29 – 1.4	0.25
	Negative		400	7 (1.8)	0.69	0.32 – 1.5	0.36
	BRCA1 positive		18	0	NA	NA	NA
	BRCA2 positive		21	0	NA	NA	NA
HGSC cases	All	OC	341	7 (2.1)	0.81	0.37 – 1.8	0.61
	Negative		310	7 (2.3)	0.9	0.41 – 2	0.79
	BRCA1 positive		15	0	NA	NA	NA
	BRCA2 positive		16	0	NA	NA	NA
HBOC[2]	All	BC	82	3 (3.7)	1.5	0.46 – 4.8	0.52
	Negative		34	2 (5.9)	2.4	0.57 –10	0.23
	BRCA1 positive		29	0	NA	NA	NA
	BRCA2 positive		21	1 (4.8)	1.9	0.26 –15	0.52
HBC	All	BC	158	3 (1.9)	0.75	0.23 – 2.4	0.63
	Negative		93	2 (2.2)	0.85	0.21 – 3.5	0.83
	BRCA1 positive		20	1 (5)	2.1	0.27 – 15	0.49
	BRCA2 positive		45	0	NA	NA	NA
Sporadic BC cases	All	BC	558	8 (1.4)	0.57	0.27 – 1.2	0.13
	Negative		538	8 (1.5)	0.59	0.28 – 1.2	0.16
	BRCA1 positive		4	0	NA	NA	NA
	BRCA2 positive		17	0	NA	NA	NA
Cancer-free females	NA	NA	2,950	74 (2.5)	1		

[1]See **Additional file 1: Table S1** for details of study groups
[2]There is overlap of some families but individuals were counted only once
[3]Inclusion of the discovery family (F1528) leads to 3 *FANCI* c.1813C>T carrier families out of 23 *BRCA1* and *BRCA2* pathogenic variant negative (13%; OR=5.8; 95%CI=1.7-20.; *P*=0.005)
NA – not available

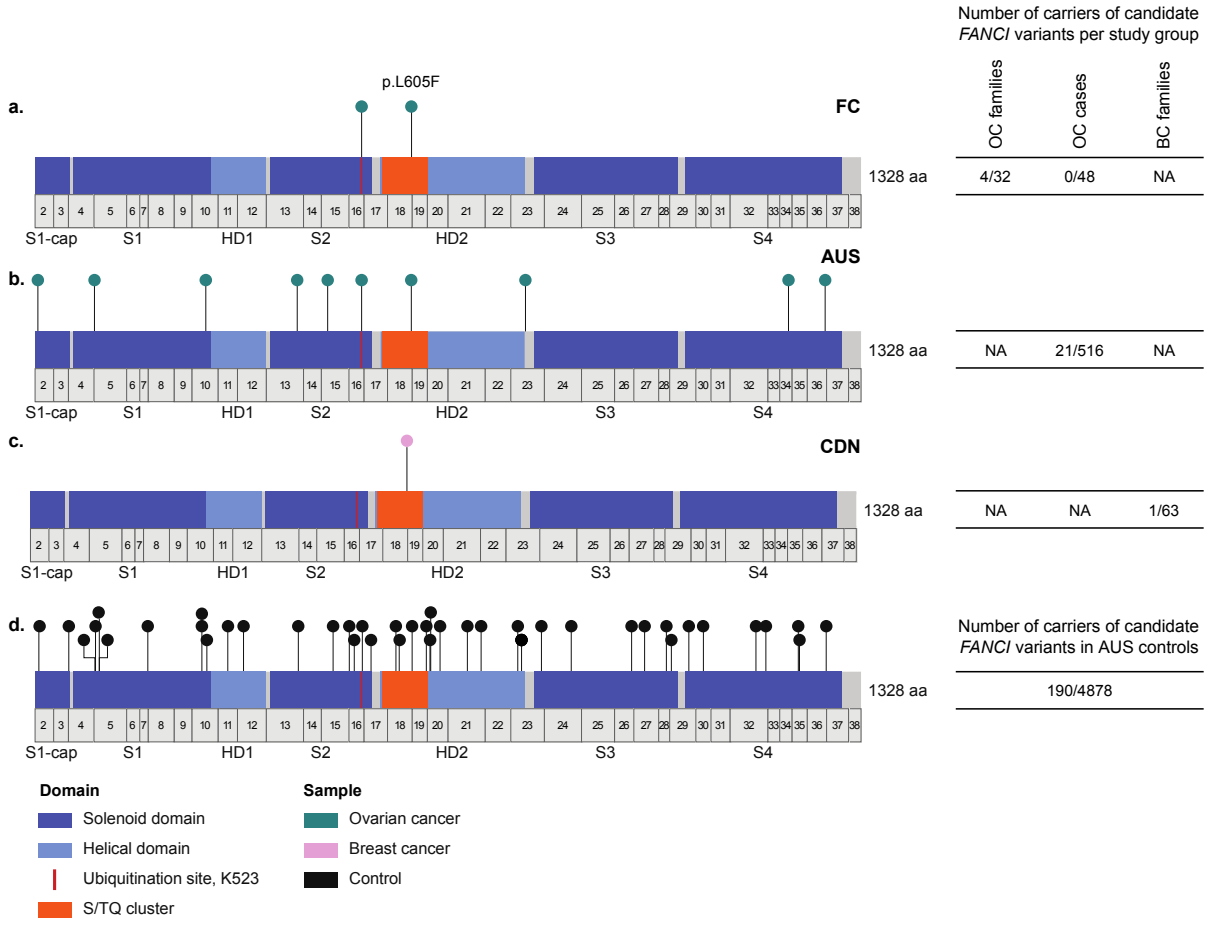


Figure 2.3. Schemata of the FANCI gene showing the location of candidate rare variants (< 1%) found in OC and/or BC in **a** French Canadian cases, **b** Australian cases, **c** Canadian non-French Canadian cases, and in **d** Australian controls. Refer to Supplementary Table 1 for study group descriptions. FANCI domains were adapted from pfam (https://pfam.xfam.org). *FANCI* exon locations adapted from University of California Santa Cruz Genome Browser (https://genome.ucsc.edu)

Table 2.2 Frequencies of carriers of candidate *FANCI* variants identified in Australian HGSC cases and controls.

Study group[1]	Number of subjects (%)	Number of variant carriers[3] (%)										
		c.13A>G p.I5V	c.286G>A p.E96K	c.824T>C p.I275T	c.1264G>A p.G422R	c.1412C>G p.P471R	c.1573A>G p.M525V	c.1813C>T p.L605F[4]	c.2366C>T p.A789V[4]	c.3635T>C p.F1212S	c.3812C>T p.S1271F	Total number of carriers[4] (%)
HGSC	516 (100)	1 (0.2)	1 (0.2)	1 (0.2)	2 (0.4)	1 (0.2)	3 (0.6)	10 (1.9)	1 (0.2)	1 (0.2)	1 (0.2)	21 (4.1)
Controls	4878 (100)	0	5 (0.1)	5 (0.1)	7 (0.1)	0	43 (0.88)	95 (1.9)	0	0	1 (0.02)	156 (3.2)
Family history of HGSC cases[2]												
≥2 OC cases (no BC)	7 (1)	0	0	0	0	0	1 (14)	0	0	0	0	1 (14)
1 OC case (no BC)	49 (10)	0	0	0	0	0	0	0	0	0	0	0
≥2 OC case and BC cases	42 (8)	0	0	0	1 (2.4)	1 (2.4)	0	5 (12)	1 (2.4)	0	0	7 (17)
Total with OC family history	**98 (19)**	**0**	**0**	**0**	**1 (1)**	**1 (1)**	**1 (1)**	**5 (5.1)**	**1 (1)**	**0**	**0**	**8 (8.2)**
≥2 BC cases (no OC)	45 (9)	0	0	0	0	0	1 (2.2)	0	0	0	0	1 (2.2)
1 BC case (no OC)	125 (24)	0	1 (0.8)	1 (0.8)	0	0	0	1 (0.8)	0	0	0	3 (2.4)
No OC or BC	248 (48)	1 (0.4)	0	0	1 (0.4)	0	1 (0.4)	4 (1.6)	0	1 (0.4)	1 (0.4)	9 (3.6)
Total isolated HGSC	**418 (81)**	**1 (0.2)**	**1 (0.2)**	**1 (0.2)**	**1 (0.2)**	**0**	**2 (0.5)**	**5 (1.2)**		**1 (0.2)**	**1 (0.2)**	**13 (3.1)**

[1]See **Additional file 1: Table S1** for description of study groups; 95% of the participants were of Western European descent

[2]First-, second-, or third-degree relatives reported for OC; first- and second-degree relatives only reported for BC

[3]See Supplementary Table 6 for more information on *FANCI* variants found in these study groups

[4]One HGSC case carried two *FANCI* variants: c.1813C>T; p.L605F and c.2366C>T; p.A789V (see **Additional file 1: Table S9**)

Table 2.3 Summary statistics for candidate *FANCI* variants in the AUS population as compared to cancer-free samples from gnomAD.

			Non-Finnish Europeans			All populations		
Coding DNA reference sequence[1]	**Amino acid change**	**dbSNP designation**	**OR**	**95% CI**	***P***	**OR**	**95% CI**	***P***
c.13A>G	p.I5V	rs200186938	NA	NA	NA	26.05	3.3 – 204	$1.9x10^{-3}$
c.286G>A	p.E96K	rs149243307	3.7	0.5 – 27.1	0.2	0.57	0.08– 4.1	0.58
c.824T>C	p.I275T	rs142906652	1.77	0.3 – 12.7	0.57	0.92	0.1 – 6.6	0.93
c.1264G>A	p.G422R	rs146040966	8.51	2 – 35.9	$3.5x10^{-3}$	17.39	4.2 – 73	$1x10^{-4}$
c.1412C>G	p.P471R	rs139072231	8.19	1.1 – 62.4	0.04	17.36	2.3 – 132	$5.8 x10^{-3}$
c.1573A>G	p.M525V	rs144908351	0.75	0.2 – 2.3	0.62	1.31	0.4 – 4.1	0.64
c.1813C>T	p.L605F	rs117125761	0.98	0.5 – 1.8	0.94	1.47	0.8 – 2.8	0.23
c.2366C>T	p.A789V	rs925359228	NA	NA	NA	NA	NA	NA
c.3635T>C	p.F1212S	rs775483853	57.4	5.2 – 634	$1x10^{-3}$	130.26	11.8 – 1439	$1x10^{-4}$
c.3812C>T	p.S1271F	rs202066338	19.9	2.3 – 171	$6.3x10^{-3}$	12.77	1.7 – 95.9	0.013

[1] Human GRCh37/hg19

NA – data not available for the controls

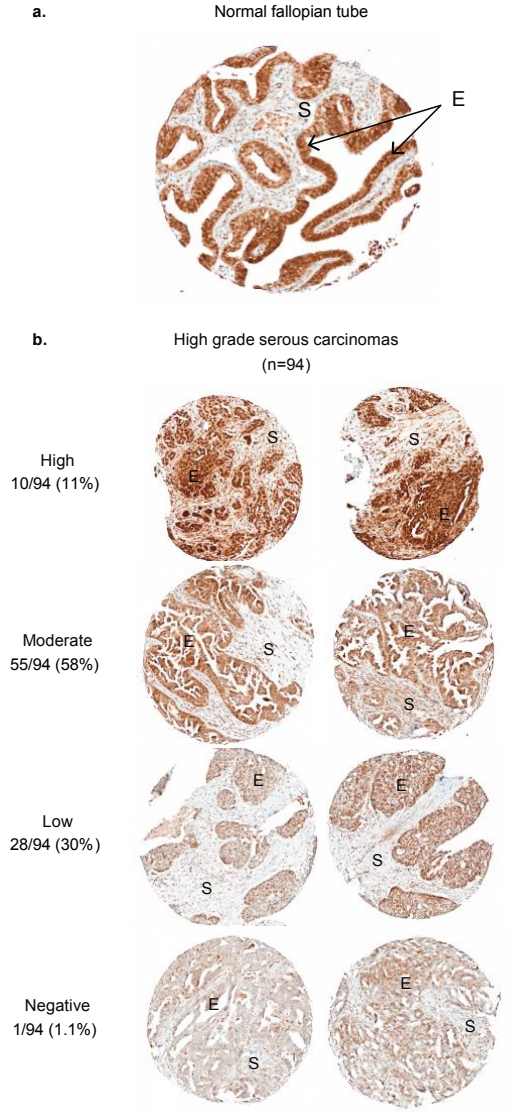


Figure 2.4. FANCI protein expression in HGSC by immunohistochemical analysis (IHC) of tissue microarrays. **a** An example of IHC analysis of FANCI protein of a paraffin-embedded normal fallopian tube tissue core. **b** Examples of different patterns of intensity of IHC analysis of FANCI protein of HGSC tissue cores in which the epithelial component is scored. E: epithelial component; S: stromal component

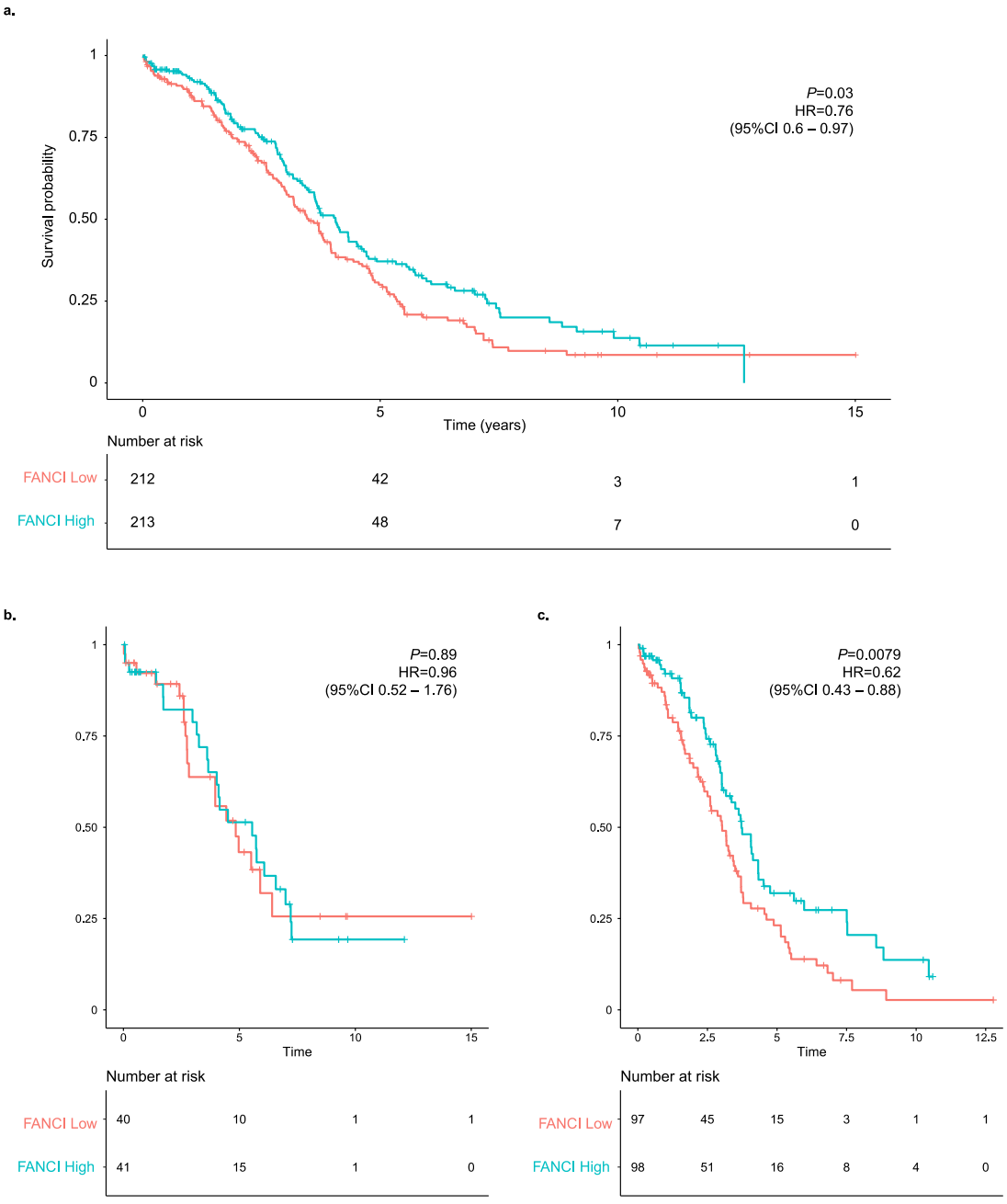

a.
Survival probability
1
0.75
0.50
0.25
0
P=0.03
HR=0.76
(95%CI 0.6 – 0.97)
0
5
10
15
Time (years)
Number at risk
FANCI Low
212
42
3
1
FANCI High
213
48
7
0
b.
P=0.89
HR=0.96
(95%CI 0.52 – 1.76)
Time
Number at risk
FANCI Low
40
10
1
1
FANCI High
41
15
1
0
c.
P=0.0079
HR=0.62
(95%CI 0.43 – 0.88)
2.5
7.5
12.5
Time
Number at risk
FANCI Low
97
45
15
3
1
1
FANCI High
98
51
16
8
4
0

Figure 2.5. Kaplan-Meier overall survival curves of OC cases from TCGA Pan-Cancer for *FANCI* mRNA expression. All cases (**a**), *BRCA1* or *BRCA2* pathogenic variant carriers (**b**), and *BRCA1* and *BRCA2* pathogenic variant carrier negative cases (**c**) are shown. Samples were dichotomized into high and low *FANCI* expression groups based on the median**.**

2.9 Supplementary note

To further characterize the germline genomic landscape of *FANCI* c.1813C>T carriers, we performed WES analysis of all seven carriers identified in the sporadic FC OC cases and compared it with WES data from five carriers identified in OC families (Table 1). We investigated DNA repair pathways genes (n=276 (1)) rationalizing that aberrant DNA repair is a hallmark of cancer and other variants in these genes may contribute to risk if shared among *FANCI* c.1813C>T carriers. Bioinformatic analyses of WES data identified a rare DNA polymerase γ (*POLG*) c.2492A>G; p.Y831C variant in the heterozygous state in all *FANCI* c.1813C>T carriers. *POLG* encodes the catalytic subunit for polymerase γ, the only known mitochondrial DNA polymerase, and has not been reported in association with cancer (2). *POLG* c.2492A>G has not been reported in association with a disease. Interestingly, *POLG* c.2492A>G is predicted to be damaging in 10/13 *in silico* tools (data not shown). Unlike *FANCI*, *POLG* is transcribed on the negative strand immediately downstream of *FANCI* where its 3' UTR is encoded in part by a genetic region shared in common with *FANCI*. As the estimated distance between *FANCI* c.1813C>T and *POLG* c.2492A>G is 36.6 kilobase pairs, it is possible that these alleles are in linkage disequilibrium in the FC population due to common ancestry (3,4). To investigate this possibility, all FC cancer cases were genotyped for *POLG* c.2492A>G carrier status. Only one case was found not to carry co-occurring *FANCI-POLG* variants: a *FANCI* variant carrying mucinous OC case. We also investigated available genotyping data from CARTaGENE for cancer-free FCs and found that *FANCI* c.1813C>T tended to co-occur with *POLG* c.2492A>G (Log2 Odds Ratio=15.2) suggesting that these rare alleles are likely in linkage disequilibrium in FCs.

2.8.1 References

1. Knijnenburg TA, Wang L, Zimmermann MT, Chambwe N, Gao GF, Cherniak AD, et al. Genomic and molecular landscape of DNA damage repair deficiency across The Cancer Genome Atlas resource. Cell Rep. 2018;23:239–54.
2. Walker RL, Anziano P, Meltzer PS. A PAC containing the human mitochondrial DNA polymerase gamma gene (POLG) maps to chromosome 15q25. Genomics. 1997;40(2):376–8.
3. Laberge A-M, Michaud J, Richter A, Lemyre E, Lambert M, Brais B. Population history and its impact on medical genetics in Quebec. Clin Genet. 2005;68:287–301.
4. Scriver CR. Human genetics: Lessons from Quebec populations. Annu Rev Genomics Hum Genet. 2001;2:69–101.

2.10 Supplementary figures

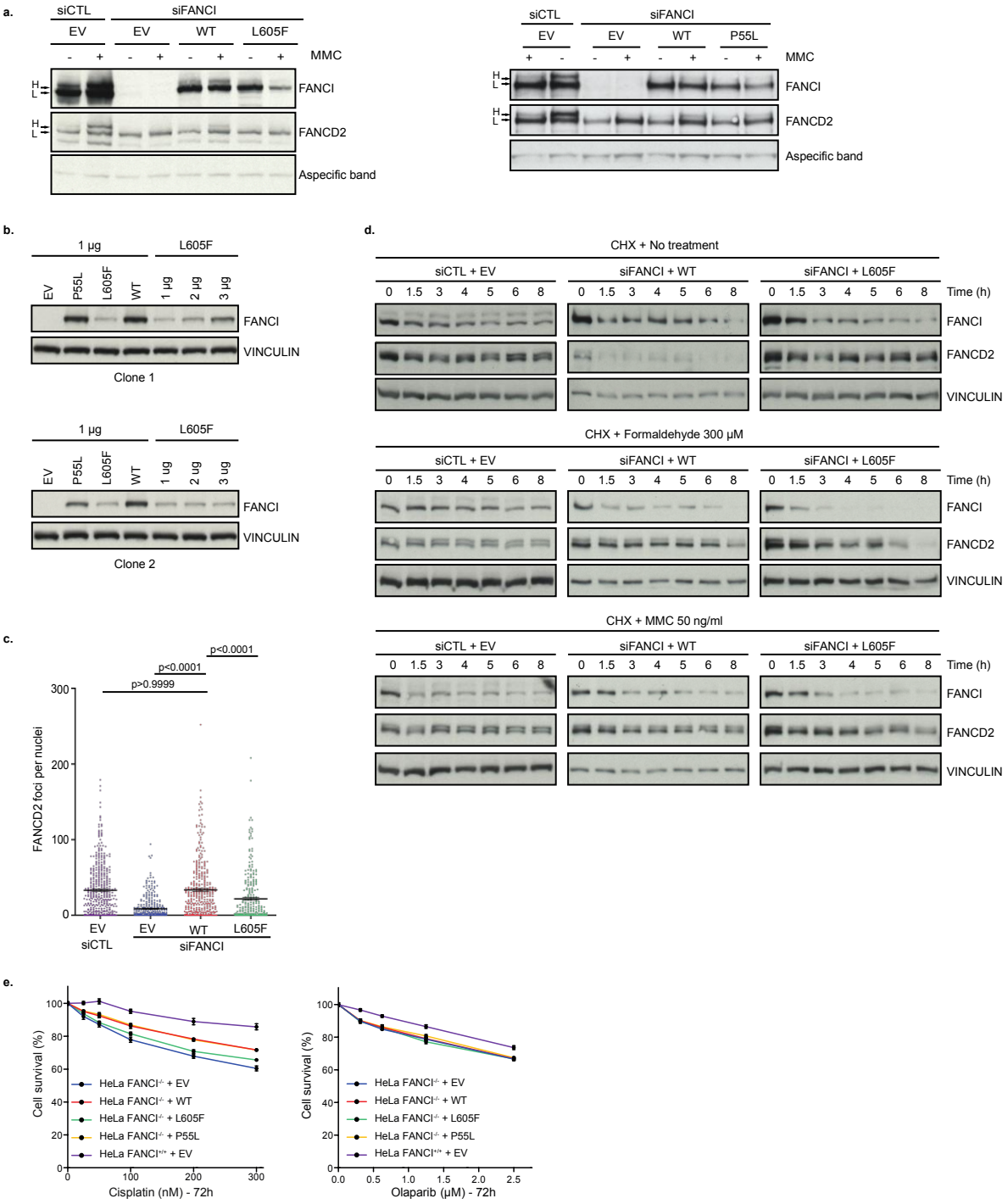


Fig. 2.S1. The isoform with the p.L605F variant impairs FANCI stability and function. a) Western blots of HeLa FANCI$^{+/+}$ cells transfected with siRNA control (siCTL) or targeting FANCI (siFANCI) and then complemented with Flag-FANCI constructs or empty vector (EV) and treated with 50 ng/ml MMC for 18 hours. The upper band, H, shows the ubiquitination of FANCI and FANCD2 after treatment. The lower band, L, corresponds to non-ubiquitinated FANCI or FANCD2. VINCULIN was used as a control. b) Western blot of HeLa FANCI$^{-/-}$ cells from clones 1 and 2 were complemented with increasing amounts of FANCI p.L605F plasmid. c) Immunofluorescence of HeLa FANCI$^{+/+}$ cells transfected with siRNA targeting FANCI and complemented with EV or Flag-FANCI siRNA resistant constructs. The number of FANCD2 foci in Flag positive cells after treatment with MMC (50 ng/ml, 18 hours) is shown. The upper and lower edge of the solid bars represents the SEM. The Kruskal-Wallis test was used to compare groups and the *P*-value is shown for each test. d) Western blots of HeLa FANCI$^{+/+}$ cells transfected with siRNA targeting FANCI and complemented with Flag-FANCI siRNA resistant constructs and treated with CHX and either mock-treated or treated with damaging agents formaldehyde or MMC for different lengths of time. At each time point, whole cell extracts were analyzed by Western blot to assess protein levels. Experiment has been performed twice. e) Survival curves of HeLa FANCI$^{-/-}$ cells from clone 2 that were complemented with constructs of Flag-FANCI variants or empty vector (EV) and plated in triplicate in a 96 well plate. Cell viability was monitored following cisplatin or olaparib treatments for 72 hours and was assessed by counting remaining nuclei. Experiments were performed in three biological replicates.

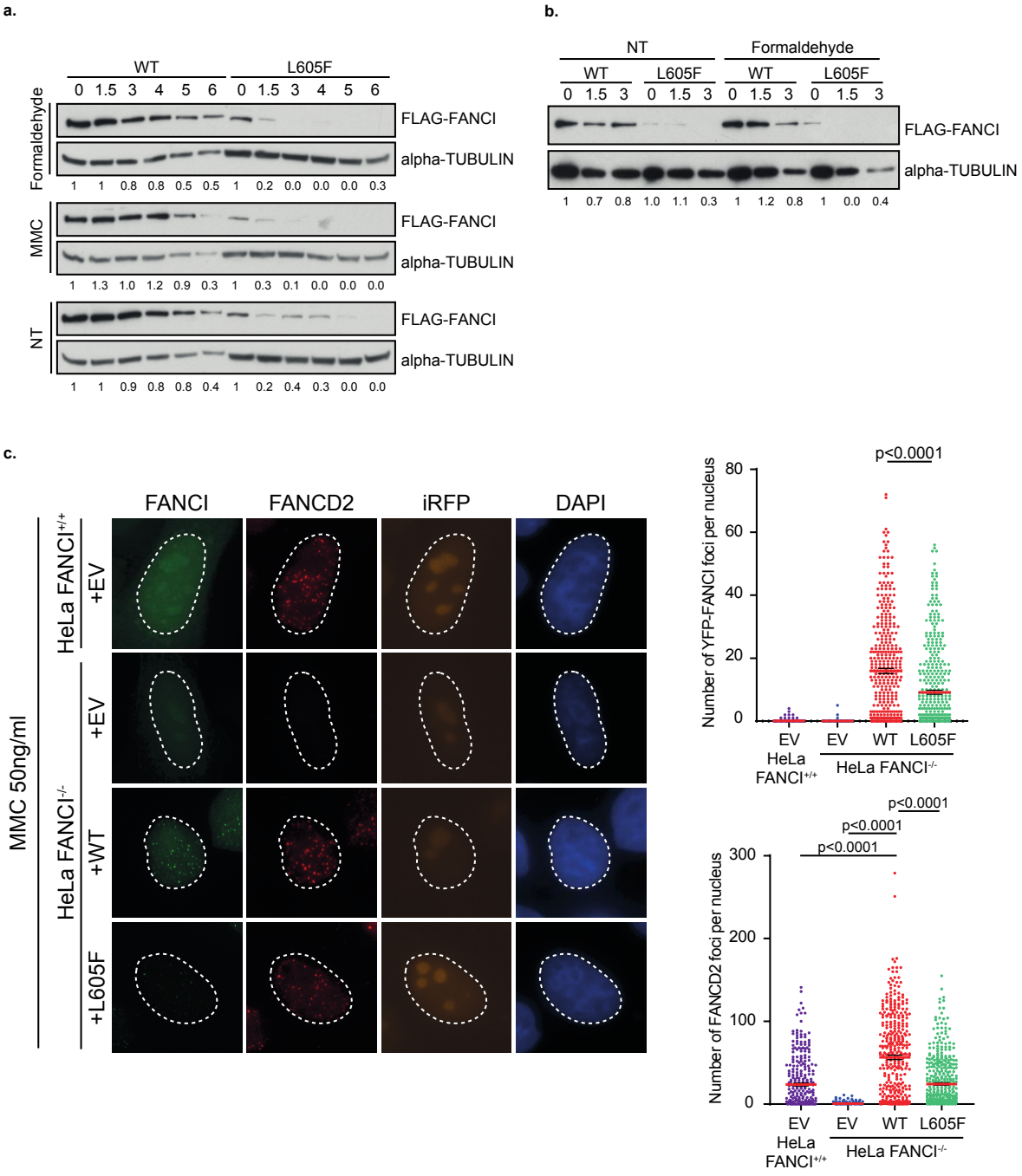
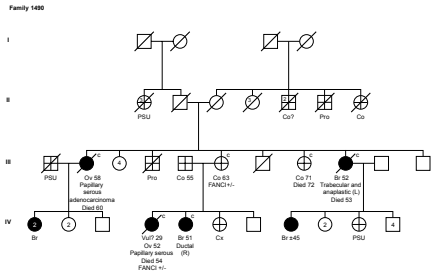
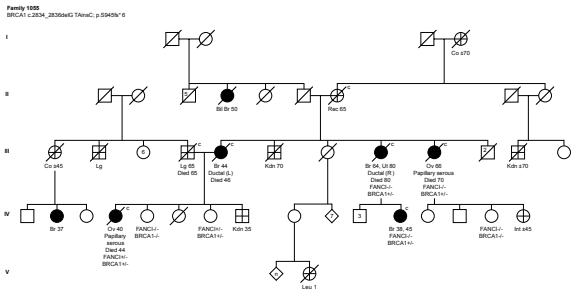
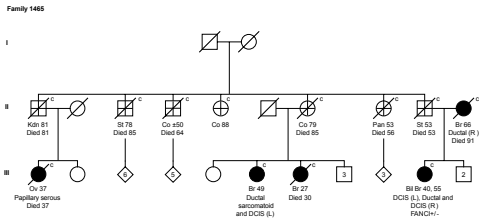
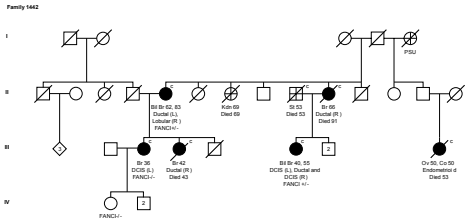

a.
WT
L605F
0 1.5 3 4 5 6 0 1.5 3 4 5 6
Formaldehyde
FLAG-FANCI
alpha-TUBULIN
1 1 0.8 0.8 0.5 0.5 1 0.2 0.0 0.0 0.0 0.3
MMC
FLAG-FANCI
alpha-TUBULIN
1 1.3 1.0 1.2 0.9 0.3 1 0.3 0.1 0.0 0.0 0.0
NT
FLAG-FANCI
alpha-TUBULIN
1 1 0.9 0.8 0.8 0.4 1 0.2 0.4 0.3 0.0 0.0
b.
NT
Formaldehyde
WT
L605F
WT
L605F
0 1.5 3 0 1.5 3 0 1.5 3 0 1.5 3
FLAG-FANCI
alpha-TUBULIN
1 0.7 0.8 1.0 1.1 0.3 1 1.2 0.8 1 0.0 0.4
c.
FANCI
FANCD2
iRFP
DAPI
MMC 50ng/ml
HeLa FANCI+/+
+EV
HeLa FANCI-/-
+EV
+WT
+L605F
Number of YFP-FANCI foci per nucleus
p<0.0001
80
60
40
20
0
EV
HeLa FANCI+/+
EV
WT
L605F
HeLa FANCI-/-
Number of FANCD2 foci per nucleus
p<0.0001
p<0.0001
p<0.0001
300
200
100
0
EV
HeLa FANCI+/+
EV
WT
L605F
HeLa FANCI-/-

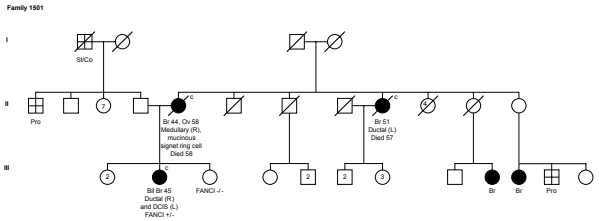
Fig. 2.S2. The isoform with the p.L605F variant impairs FANCI stability in OC cell lines and FANCI localization to DNA damage in HeLa cells. a) Western blots of OVCAR-4 cells expressing Flag-FANCI wild type (WT) or p.L605F. Cells were treated with CHX and either mock-treated (NT) or treated with formaldehyde (300 μM for the indicated times) or MMC (50 ng/ml for the indicated times). Alpha-tubulin was used as a loading control. b) Western blots of OVCAR-3 cells expressing Flag-FANCI wild type (WT) or p.L605F, treated with cycloheximide and either mock-treated or treated with formaldehyde (300 μM for the indicated times). Alpha- tubulin was used as a loading control. c) Immunofluorescence analysis of wild-type HeLa or HeLa FANCI-/- cells complemented with either YFP alone (EV), YFP-FANCI, or YFP-FANCI p.L605F constructs. iRFP was used as a transfection marker. The adjacent scatter plot shows the number of YFP-FANCI foci per nucleus or FANCD2 foci in iRFP-positive cells after treatment with MMC (50 ng/ml, 18 hours). Mean with SEM is represented. The Kruskal-Wallis test was used to compare groups and the P-value is shown for each test.

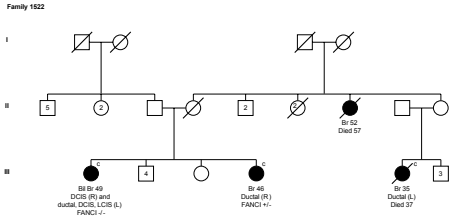

Family 1490
I
II
III
IV
PSU
Col7
Pro
Co
PSU
Ov 58
Papillary
serous
adenocarcinoma
Died 60
Pro
Co 55
Co 63
FANCI +/-
Co 71
Died 72
Br 52
Trabecular and
anaplastic (L)
Died 53
Br
Vul? 29
Ov 52
Papillary serous
Died 54
FANCI +/-
Br 51
Ductal
(R)
Cx
Br ±45
PSU

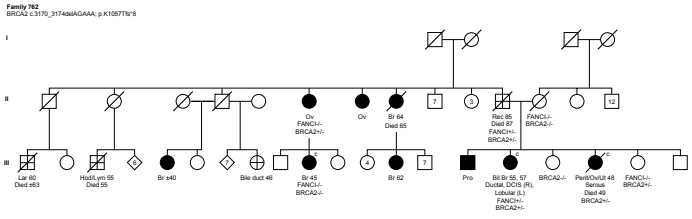

Family 1055
BRCA1 c.2834_2836delG TAinsC; p.S945fs*6
I
II
III
IV
V
Co ±70
Bil Br 50
Rec 65
Co ±45
Lg
Lg 65
Died 65
Br 44
Ductal (L)
Died 46
Kdn 70
Br 64, Ut 80
Ductal (R)
Died 80
FANCI+
BRCA1+/-
Ov 66
Papillary serous
Died 70
FANCI+/-
BRCA1+/-
Kdn ±70
Br 37
Ov 40,
Papillary
serous
Died 44
FANCI+/-
BRCA1+/-
FANCI-/-
BRCA1-/-
FANCI+/-
BRCA1+/-
Kdn 55
Br 38, 45
FANCI+/-
BRCA1+/-
FANCI-/-
BRCA1+/-
Int ±45
Leu 1

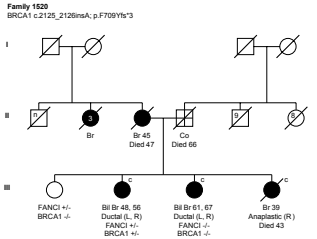

Family 1465
I
II
III
Kdn 81
Died 81
St 78
Died 85
Co ±50
Died 64
Co 88
Co 79
Died 85
Pan 53
Died 56
St 53
Died 53
Br 66
Ductal (R)
Died 91
Ov 37
Papillary serous
Died 37
Br 49
Ductal
sarcomatoid
and DCIS (L)
Br 27
Died 30
Bil Br 40, 55
DCIS (L), Ductal and
DCIS (R)
FANCI+/-


Family 1462
I
II
III
IV
PSU
Bil Br 62, 63
Ductal (L),
Lobular (R)
FANCI+/-
Kdn 69
Died 69
St 53
Died 53
Br 66
Ductal (R)
Died 91
Br 36
DCIS (L)
FANCI-/-
Br 42
Ductal (R)
Died 43
Bil Br 40, 55
DCIS (L), Ductal and
DCIS (R)
FANCI +/-
Ov 50, Co 50
Endometrioid
Died 53
FANCI-/-


Family 1501
I
II
III
StCo
Pro
Br 44, Ov 58
Medullary (R),
mucinous
signet ring cell
Died 58
Br 51
Ductal (L)
Died 57
Bil Br 45
Ductal (R)
and DCIS (L)
FANCI +/-
FANCI -/-
Br
Br
Pro


Family 1522
I
II
III
Br 52
Died 57
Bil Br 49
DCIS (R) and
ductal, DCIS, LCIS (L)
FANCI -/-
Br 46
Ductal (R)
FANCI +/-
Br 35
Ductal (L)
Died 37


Family 762
BRCA2 c.3170_3174delAGAAA; p.K1057Tfs*8
I
II
III
Lar 60
Died ±63
Hodg/Lym 55
Died 55
Br ±40
Bile duct 46
Ov
FANCI-/-
BRCA2+/-
Br 45
FANCI-/-
BRCA2-/-
Ov
Br 64
Died 65
Br 62
Rec 85
Died 87
FANCI+/-
BRCA2+/-
FANCI-/-
BRCA2-/-
Pro
Bil Br 55, 57
Ductal, DCIS (R),
Lobular (L)
FANCI+/-
BRCA2+/-
BRCA2-/-
Peritl/Ov/Ut 48
Serous
Died 49
BRCA2+/-
FANCI-/-
BRCA2+/-


Family 1520
BRCA1 c.2125_2126insA; p.F709Yfs*3
I
II
III
Br
Br 45
Died 47
Co
Died 66
FANCI +/-
BRCA1 -/-
Bil Br 48, 56
Ductal (L, R)
FANCI +/-
BRCA1 +/-
Bil Br 61, 67
Ductal (L, R)
FANCI -/-
BRCA1 -/-
Br 39
Anaplastic (R)
Died 43

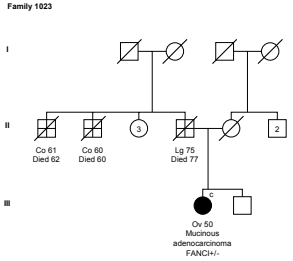

Family 1023
I
II
III
Co 61
Died 62
Co 60
Died 60
Lg 75
Died 77
Ov 50
Mucinous
adenocarcinoma
FANCI+/-

Fig. 2.S3. Pedigrees of OC and BC families with *FANCI* c.1813C>T; p.L605F. Cancer type (Ov: ovarian, Leu: leukemia, Br: breast, Cx: cervical, Co: colon, Rec: rectal, St: stomach, Kdn: kidney, Pro: prostate, Lg: lung, Perit: peritoneum, Ut: uterine, Lar: larynx, Hod: Hodgkin's lymphoma, Lym: lymphoma, and PSU: primary site unknown), risk reducing surgery (TAH/BSO: total abdominal hysterectomy/bilateral salpingo-oophorectomy), and age at diagnosis is indicated. *BRCA1* and *BRCA2* carrier status is shown for pathogenic variant positive families, F762, F1055, F1520; all other families are *BRCA1* and *BRCA2* pathogenic variant negative. c next to a symbol denotes a confirmed cancer case. Pedigrees may have been truncated to protect anonymity.

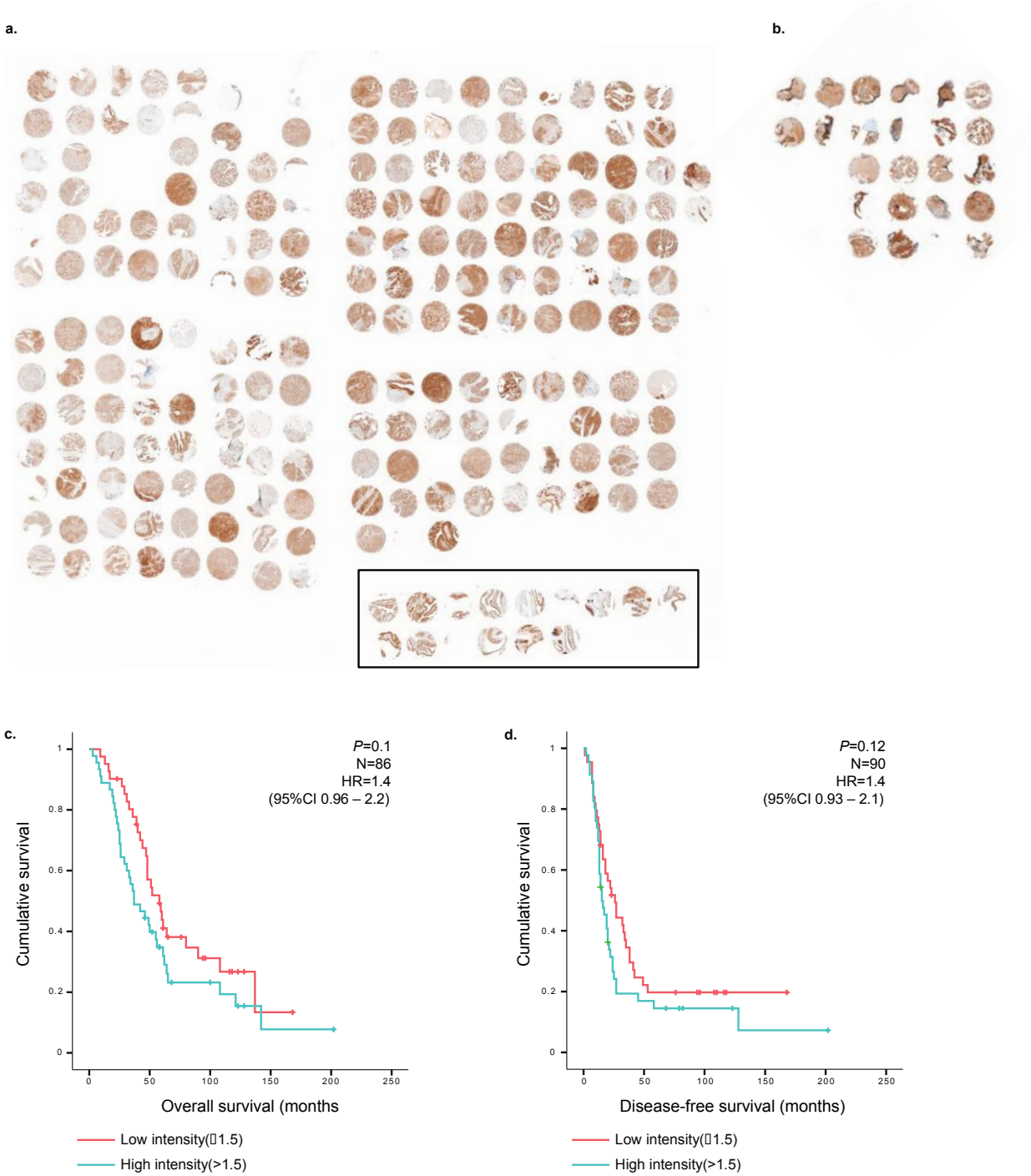

a.
b.
c.
P=0.1
N=86
HR=1.4
(95%CI 0.96 – 2.2)
Cumulative survival
Overall survival (months
Low intensity(1.5)
High intensity(>1.5)
d.
P=0.12
N=90
HR=1.4
(95%CI 0.93 – 2.1)
Cumulative survival
Disease-free survival (months)
Low intensity(1.5)
High intensity(>1.5)

Fig. 2.S4. Immunohistochemical analyses of FANCI protein expression of tissue microarrays (TMAs) from (a) HGSC cases and normal fallopian tube (a, black box) and (b) *FANCI* c.1813C>T OC carriers tissue cores. TMAs were stained with anti-FANCI antibody on the same slide. Kaplan-Meier survival curve of HGSC cases for overall (c) and disease-free (d) survival (in months) as measured in the epithelial cell component. Cases included in the analyses had received only adjuvant taxol and carboplatin chemotherapy.

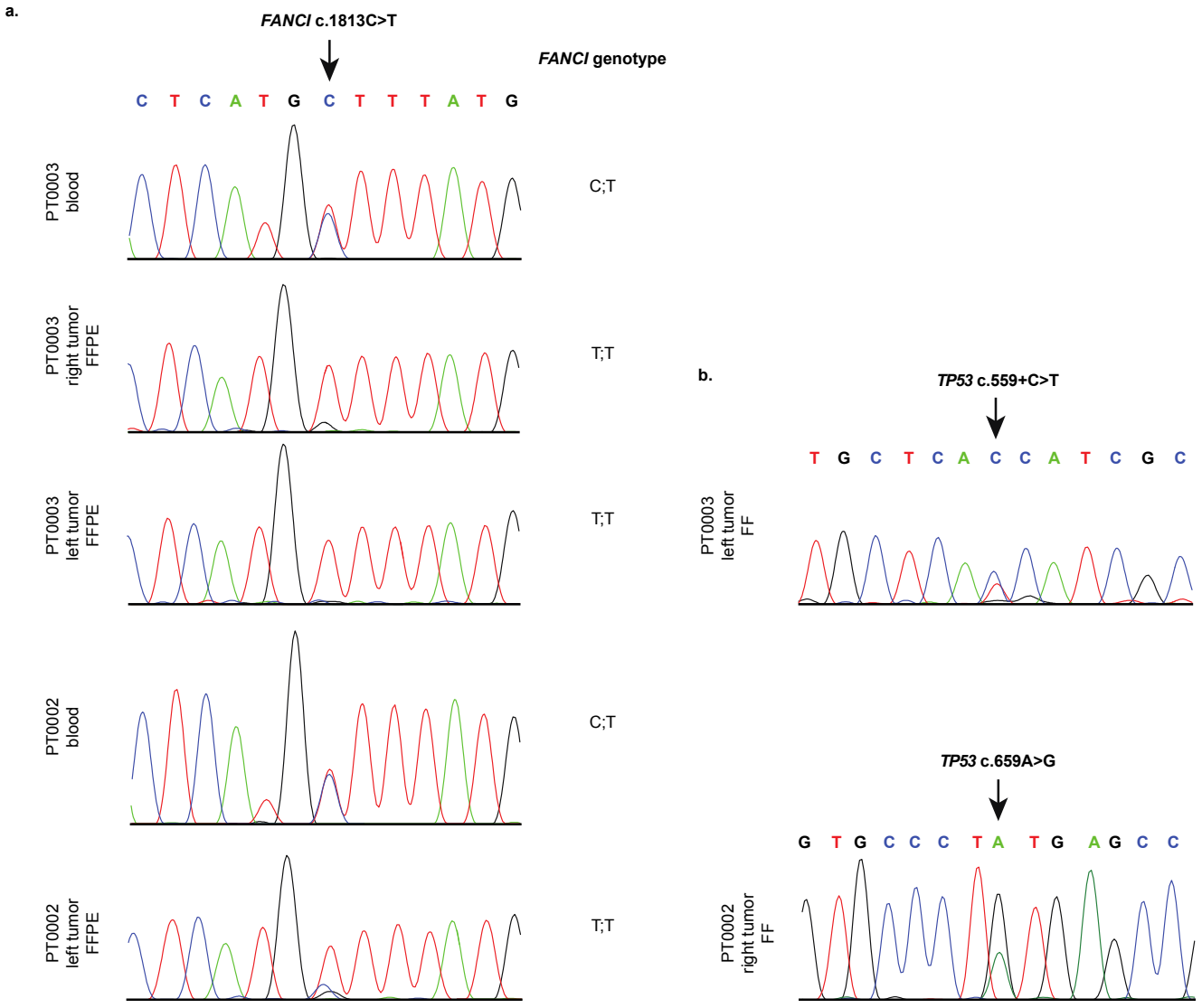

a.
FANCI c.1813C>T
FANCI genotype
C T C A T G C T T T A T G
PT0003 blood
C;T
PT0003 right tumor FFPE
T;T
PT0003 left tumor FFPE
T;T
PT0002 blood
C;T
PT0002 left tumor FFPE
T;T
b.
TP53 c.559+C>T
T G C T C A C C A T C G C
PT0003 left tumor FF
TP53 c.659A>G
G T G C C C T A T G A G C C
PT0002 right tumor FF

Fig. 2.S5. Identification of *FANCI* c.1813C>T and *TP53* variants in French Canadian HGSC cases. a) DNA sequencing chromatogram showing the region containing c.1813C>T corresponding to DNA obtained from (top to bottom): PT0003 blood showing the heterozygous variant, PT0003 FFPE right tumor showing loss of the WT allele, PT0003 FFPE left tumor showing loss of the WT allele, PT0002 blood showing the heterozygous variant, and PT0002 FFPE left tumor showing loss of the WT allele. The forward sequence is shown. B) DNA sequencing chromatogram showing the region containing identified *TP53* variants corresponding to (top to bottom): PT0003 FF left tumor showing *TP53* variant c.559+1C>T and PT0002 FF right tumor showing TP53 variant c.659A>G. The reverse sequence is shown. FFPE: formalin-fixed paraffin-embedded tumor tissue; FF: fresh frozen tumor tissue.

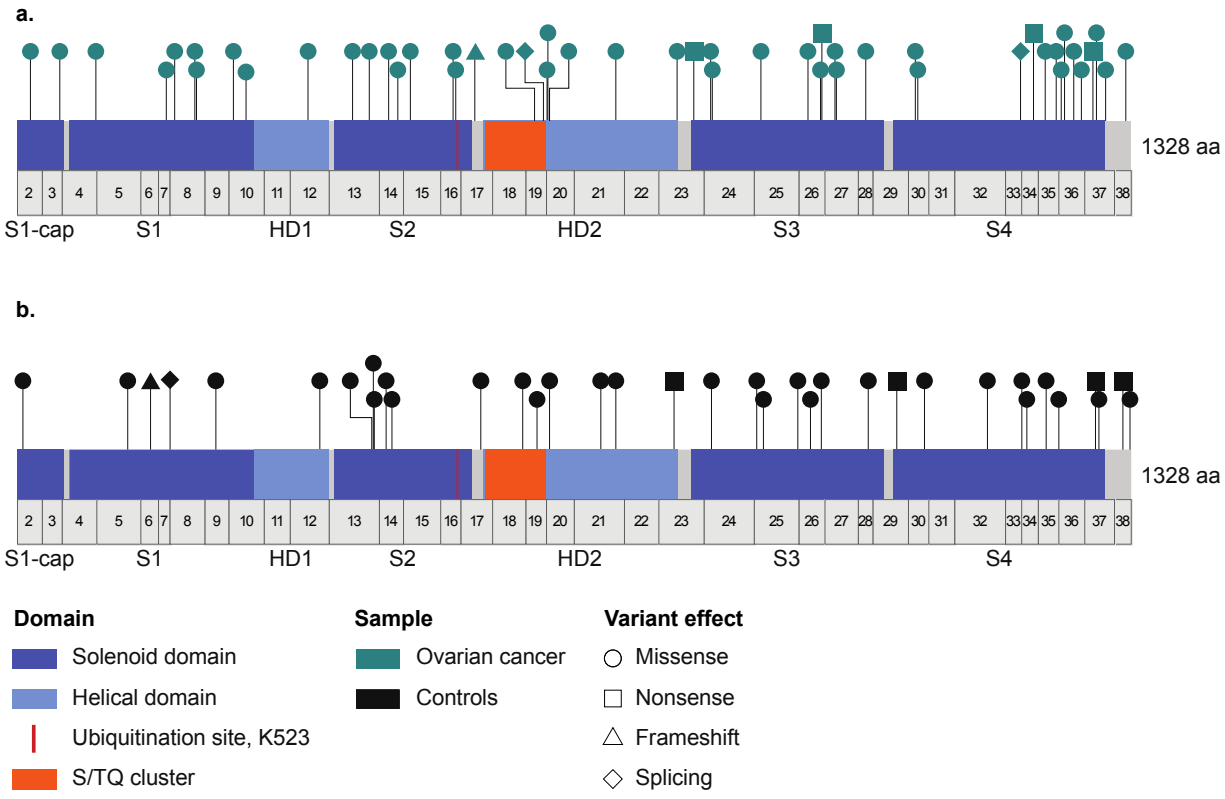


Fig. 2.S6. Schemata of *FANCI* gene showing rare variants (VAF<1%) reported in (a) OC (b) and controls from the literature. FANCI domains were adapted from pfam (https://pfam.xfam.org). *FANCI* exon locations adapted from University of California Santa Cruz Genome Browser (https://genome.ucsc.edu).

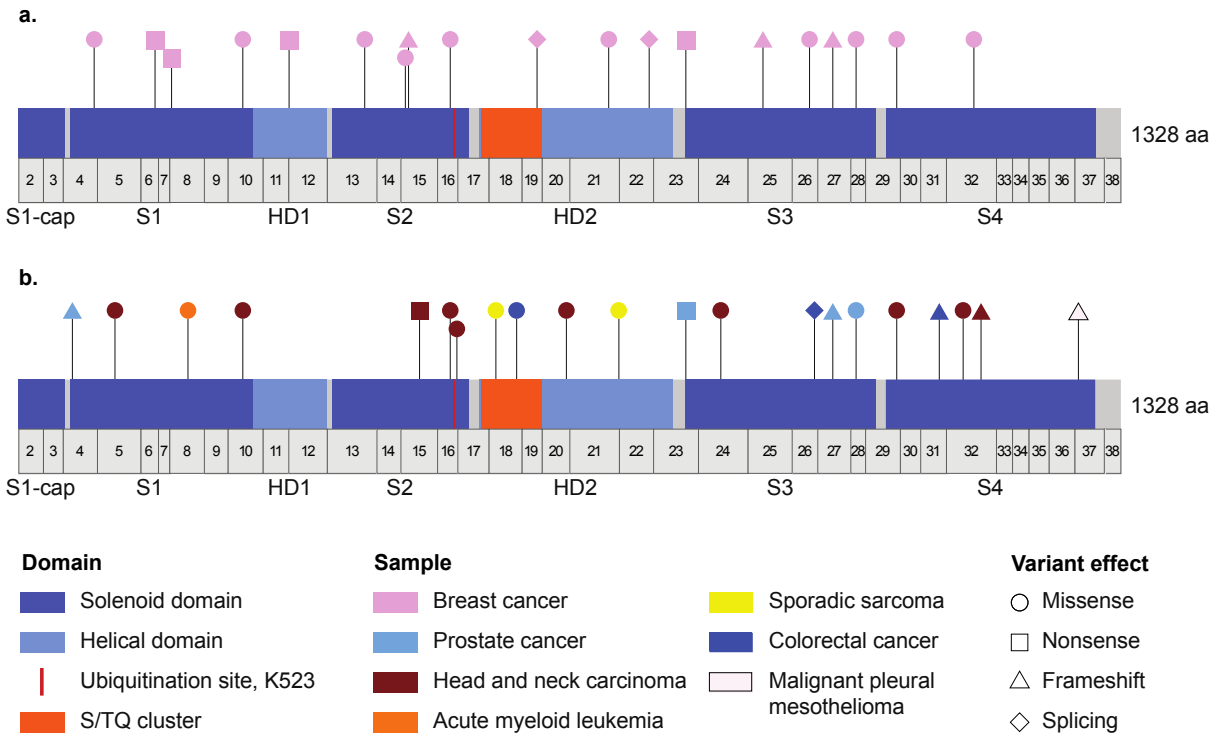


Fig. 2.S7. Schemata of *FANCI* gene showing rare variants (VAF<1%) reported in the literature in (a) BC (b) and other cancers. FANCI domains were adapted from pfam (https://pfam.xfam.org). *FANCI* exon locations adapted from University of California Santa Cruz Genome Browser (https://genome.ucsc.edu).

Bridging text

In Chapter 2.0, we investigated the possibility that *FANCI* c.1813C>T may be associated with OC risk and our results suggest this is a likely pathogenic variant. Our study was prompted by the discovery of this *FANCI* variant in the context of our discovery family (F1528) that was negative for pathogenic variants in known OC predisposing genes. Though *FANCI* c.1813C>T; p.L605F was the most plausible candidate to investigate, as it was the only gene involved in a DNA repair pathway, it was proposed that there could be a different variant exhibiting moderate to high risk that is not involved in a DNA repair pathway in this family. To address this question, I performed a global analysis to identify variants across the genetic landscape of the two sisters from family F1528. This analysis was restricted to family F1528 as it was the only family with two affected OC cases that both harboured *FANCI* c.1813C>T; p.L605F. After confirming that there were no other strong candidate OC risk variants identified in family F1528, I investigated *FANCI* to determine if it displayed characteristics exhibited by known OC predisposing genes and CPGs. This analyses compliments and expands on analyses performed in Chapter 2.0. I investigated somatic loss of heterozygosity and somatic genetic alterations in OC carriers of *FANCI* c.1813C>T. I investigated the possibility of tissue tropism for *FANCI* c.1813C>T by determining the carrier frequency across cancer types. The somatic genetic landscape of *FANCI* variants was investigated across cancer types to determine the type of variants and context in which they can be identified. Finally, as all data pointed back to *FANCI* as a candidate OC predisposing gene, I investigated other missense variants in *FANCI* with characteristics similar to c.1813C>T; p.L605F that would benefit from further

biological investigation. I also used a candidate gene approach to determine if other genes in the FANCI protein interactome may be involved in risk to OC by investigating variants in the interactome genes in OC cases negative for variants in *BRCA1*, *BRCA2*, *BRIP1*, *RAD51C*, *RAD51D*, and *FANCI*. This data presented in Chapter 3.0 has furthered our understanding of *FANCI* OC carriers and *FANCI* in the context of other cancers.

3.0 Chapter 3: Molecular genetic characteristics of *FANCI*, a proposed new ovarian cancer predisposing gene

Authors: Caitlin T Fierheller[1,2], Wejdan M Alenezi[1,2,3], Corinne Serruya[2], Timothée Revil[1,4], Setor Amuzu[1,4], Karine Bedard[5,6], Deepak N Subramanian[7], Eleanor Fewings[8], Jeffrey P Bruce[9], Stephenie Prokopec[9], Luigi Bouchard[10,11,12], Diane Provencher[13,14], William D Foulkes[1,2,15,16], Zaki El Haffaf[17], Anne-Marie Mes-Masson[13,18], Marc Tischkowitz[8], Ian G Campbell[7,19], Trevor J Pugh[9], Celia M T Greenwood[1,15,20,21], Jiannis Ragoussis[1,4], Patricia N Tonin[1,2,16]

Affiliations: [1]Department of Human Genetics, McGill University, Montreal, Quebec, Canada, [2]Cancer Research Program, The Research Institute of the McGill University Health Centre, Montreal, Quebec, Canada, [3]Department of Medical Laboratory Technology, Taibah University, Medina, Saudi Arabia, [4]McGill Genome Centre, McGill University, Montreal, Quebec, Canada, [5]Laboratoire de Diagnostic Moléculaire, Centre Hospitalier de l'Université de Montréal (CHUM), Montreal, Quebec, Canada, [6]Département de pathologie et biologie cellulaire, Université de Montréal, Montreal, Quebec, Canada, [7]Cancer Genetics Laboratory, Peter MacCallum Cancer Centre, Melbourne, Victoria, Australia, [8]Department of Medical Genetics, National Institute for Health Research Cambridge Biomedical Research Centre, University of Cambridge, Cambridge, UK, [9]Princess Margaret Cancer Centre, University Health Network, Toronto, Ontario, Canada, [10]Department of Biochemistry and Functional Genomics, Université de Sherbrooke, Sherbrooke, Quebec, Canada, [11]Department of

Medical Biology, Centres intégrés universitaires de santé et de services sociaux du Saguenay-Lac-Saint-Jean hôpital Universitaire de Chicoutimi, Saguenay, Quebec, Canada, [12]Centre de Recherche du Centre hospitalier l'Université de Sherbrooke, Sherbrooke, Quebec, Canada, [13]Centre de recherche du Centre hospitalier de l'Université de Montréal and Institut du cancer de Montréal, Montreal, Quebec, Canada, [14]Division of Gynecologic Oncology, Université de Montréal, Montreal, Quebec, Canada, [15]Lady Davis Institute for Medical Research, Jewish General Hospital, Montreal, Quebec, Canada, [16]Department of Medicine, McGill University, Montreal, Quebec, Canada, [17]Centre de recherche du Centre Hospitalier de l'Université de Montréal, Montreal, Quebec, Canada, [18]Department of Medicine, Université de Montréal, Montreal, Quebec, Canada, [19]Sir Peter MacCallum Department of Oncology, University of Melbourne, Melbourne, Victoria, Australia, [20]Gerald Bronfman Department of Oncology, McGill University, Montreal, Quebec, Canada, [21]Department of Epidemiology, Biostatistics & Occupational Health, McGill University, Montreal, Quebec, Canada

Submitted to:

Genes (MDPI) for peer review on December 1, 2022

3.1 Abstract

FANCI was recently identified as a new candidate ovarian cancer (OC) predisposing gene from the genetic analysis of carriers of *FANCI* c.1813C>T; p.L605F in OC families. Here, we aimed to investigate the molecular genetic characteristics of *FANCI* as they have not been described in the context of cancer. We first investigated the germline genetic landscape of two sisters with OC from the discovery *FANCI* c.1813C>T; p.L605F family (F1528) to re-affirm plausibility of this candidate. As we did not find other conclusive candidates, we then performed a candidate gene approach to identify other candidate variants in genes involved in the FANCI protein interactome in OC families negative for pathogenic variants in *BRCA1*, *BRCA2*, *BRIP1*, *RAD51C*, *RAD51D*, and *FANCI*, which identified four candidate variants. We then investigated *FANCI* in high grade serous ovarian carcinoma (HGSC) from *FANCI* c.1813C>T carriers and found evidence of loss of the wild type allele in tumour DNA from some of these cases. The somatic genetic landscape of OC tumours from *FANCI* c.1813C>T carriers was investigated for mutations in selected genes, copy number alterations, and mutational signatures, which determined that the profile of tumours from carriers were characteristic of features exhibited by HGSC cases. As other OC predisposing genes, such as *BRCA1* and *BRCA2*, are known to increase risk to other cancers, including breast cancer, we investigated the carrier frequency of germline *FANCI* c.1813C>T in various cancer types and found overall more carriers among cancer cases compared to cancer-free controls (p=0.007). In these different tumour types, we also identified a spectrum of somatic variants in *FANCI* that were not restricted to any specific region within the gene. Collectively, these findings expand on the characteristics described for

OC cases carrying *FANCI* c.1813C>T; p.L605F and suggest the possible involvement of *FANCI* in other cancer types at the germline and/or somatic level.

3.2 Background

Since the first reports of pathogenic variants in *BRCA1* and *BRCA2,* the breast and ovarian cancer predisposing genes (CPGs), almost 30 years ago[1,2], it is increasingly evident that there are unlikely to be other major high risk genes contributing to these cancers. For ovarian cancer (OC), pathogenic variants in *BRIP1*[3,4], *RAD51C*[5–7], *RAD51D*[8], and *PALB2*[9] have been identified but each account for less than 2% of sporadic OC cases[10]. Other genes, such as the mismatch DNA repair genes *MLH1*[11,12], *MSH2*[13], *MSH6*[14], and *PMS2*[15] featured in colorectal cancer (Lynch syndrome) families, have been associated with OC risk as well, though carriers are also rare, cumulatively less than 1% of sporadic OC cases[10]. Other genes have been proposed for OC risk such as *ATM*[16–19], *BARD1*[20,21], and *FANCM*[22,23], though penetrance is not yet established. Our strategy for identifying new risk genes has focused on individuals from an ancestrally defined population exhibiting genetic drift, which allowed us to identify frequently occurring variants as potential candidates[24]. In contrast to the general population, a few pathogenic variants in *BRCA1*[25,26], *BRCA2*[25,26], *PALB2*[27,28], *RAD51C*[29], and *RAD51D*[29,30] are found to frequently occur in French Canadians (FCs) of Quebec, Canada[24]. Using this approach and applying biological assays, our group reported the candidacy of *FANCI* c.1813C>T; p.L605F as a new OC predisposing gene[31]. Heterozygous carriers of *FANCI* c.1813C>T were identified more commonly in OC families negative for *BRCA1* and *BRCA2* pathogenic variants compared to FC cancer-free controls. *FANCI* c.1813C>T cancer-free control carriers were more likely to

have a first degree relative with OC, suggesting a role in risk for OC. In addition, this variant and other candidate variants in *FANCI* were more common in familial OC compared to sporadic OC cases from another (Australian) population. FANCI, a member of the Fanconi anemia (FA) DNA repair pathway involved in the repair of interstrand crosslinks, was shown to have abrogated function in cells expressing p.L605F. FANCI p.L605F showed reduced protein expression, was destabilized upon treatment with DNA damaging agents (mitomycin C and formaldehyde) and exhibited sensitivity to cisplatin. Though the expression of FANCI protein was variable in OC tumours, it was shown to be highly expressed in normal fallopian tube epithelium, a purported tissue of origin for high grade serous ovarian carcinoma (HGSC)[32–37], the most common histopathological subtype of OC. Although its role in conferring risk to OC remains to be determined, the number of *FANCI* c.1813C>T; p.L605F carriers identified in OC cases affords an opportunity to investigate the molecular genetic features of carriers.

This study applies a bioinformatic approach to assess the genetic background in which *FANCI* carriers were identified in the context of familial OC and investigates the molecular genomic landscape of ovarian tumours from carriers using available whole exome sequencing (WES) data. We also investigate *FANCI* variant carriers in the context of other cancer types by taking advantage of The Cancer Genome Atlas (TCGA) molecular genetic data sets. We relate our findings to current knowledge concerning moderate-to-high risk OC predisposing genes and other established CPGs and the biological role of FANCI.

3.3 Methods

3.3.1 Study subjects

Study groups are described in Supplementary Table 1. Family F1528 was previously reported in a study on the histopathology of FC hereditary BC and OC families[38] and was updated with clinical data in the identification of *FANCI* as a new candidate OC predisposing gene[31]. WES analyses were previously described[31]. Briefly, DNA extracted from peripheral blood lymphocytes was exome captured followed by 100bp paired-end sequencing. Variants were aligned to human genome assembly hg19 for germline variant calling.

WES data available from familial OC cases negative for pathogenic variants in *BRCA1*, *BRCA2, BRIP1, RAD51C,* and *RAD51D* (n=13 cases [12 families]) and OC cases harbouring *FANCI* c.1813C>T; p.L605F (n=10) have been previously described[25,26,29,31,39]. Four additional HGSC cases of *FANCI* c.1813C>T were also included, which were obtained from the Réseau de recherche sur le cancer (RRCancer) Tumour and Data biobank. All cases were self-reported FC ancestry. *FANCI* c.1813C>T harbouring cases are also included in familial OC cases with (n=1) or without (n=2) *BRCA1*, *BRCA2, BRIP1, RAD51C,* or *RAD51D* pathogenic variants. There are no other cases known to be ascertained to more than one FC study group. For cancer-free FC controls, we used WES data available from CARTaGENE (n=171)[40–42] (cartagene.qc.ca) and whole genome sequencing (WGS) data from Genetics of Glucose regulation in Gestation and Growth project (Gen3G) (n=422)[43] to survey the germline genetic landscape variants.

WES data available from Australian HGSC cases (n=516)[31,44] was surveyed for genetic landscape variants.

TCGA PanCancer Atlas cancer cases from the general population were investigated for germline *FANCI* c.1813C>T carriers (n=10,389)[45] and somatic *FANCI* variants (n=10,434)[46,47]. Age at diagnosis and sex for cancer cases is available on cbioportal.org. Cancer-free cases from the general population Genome Aggregation Database (gnomAD; gnomad.broadinstitute.org)[48] were used as a comparator for *FANCI* c.1813C>T carrier frequency.

All biological samples and associated clinical information were obtained from biobanks where participants were recruited in accordance with ethical guidelines and approved Institutional Research Ethics Boards (Supplementary Table 1). FC OC samples were anonymized at source by providers and were assigned unique PT identifiers to further protect anonymity. This project received approval from and was conducted in accordance with The McGill University Health Centre Research Ethics Board (MP-37-2019-4783 and 2017-2722).

3.3.2 WES filtering and prioritization of variants identified in family F1528

Sequencing data from *FANCI* c.1813C>T carrier sisters from family F1528 was sequentially filtered (Figure 1a) for: (a) rare (minor allele frequency [MAF]≤1%) variants in the general population database gnomAD[48]; (b) variants within autosomes and the X chromosome only; (c) variant allele frequency (≥20% in at least one sister); (d) variant depth (≥10 reads in at least one sister); and (e) protein coding variants. The remaining variants were visually inspected and confirmed by Integrative Genomics Viewer (IGV)[49] and filtered as shown in Figure 1a, where variants in both sisters had variant depth ≥10

reads, variant allele frequency ≥20% for variants called heterozygous, and variant allele frequency ≥80% for variants called homozygous. The final filter applied included a survey for rare (MAF≤1%) variants in the general population from 1000 Genomes Project[50], National Heart, Lung and Blood Institute (NHLBI) Exome Sequencing Project (ESP) ESP6500SI-V2 (https://evs.gs.washington.edu/EVS/), and Exome Aggregation Consortium[51] (ExAC). These filtering steps led to a list of variants for further annotation and prioritization. The genetic landscape variants were annotated for type of variant effect (nonsense, frameshift, splice site, or missense) and results from applying in silico tools that predict: if the variant is located at a conserved locus; whether the variant is deleterious to the protein; or if the variant has the potential to affect splicing. These in silico tools were selected based on their best predictive performance[52]. The tools used to determine the predicted conservation of variants were Genomic Evolutionary Rate Profiling (GERP++)[53], Site-specific Phylogenetic analysis (SiPhy)[54], Phylogenetic P-values (PhyloP) 100 way in vertebrates[55], and Phylogenetic Analysis with Space/Time Models Conservation (PhastCons) v1.5[56]. Tools used to predict the ability of the amino acid change to affect protein function (deleterious or not) were Combined Annotation Dependent Depletion (CADD) v1.4[57], Eigen v1.1[58], Protein Variant Effect Analyzer (PROVEAN) v1.1[59], Meta Logistic Regression (MetaLR)[60], Meta Support Vector Machine (MetaSVM)[60], Rare Exome Variant Ensemble Learner (REVEL)[61], and Variant Effect Scoring Tool (VEST) v4.0[62]. Tools used to predict the potential of variants to affect splicing were database of splicing consensus regions (dbscSNV) adaptive boosting (ADA) and random forest (RF)[63], Maximum Entropy Modeling of Short Sequence Motifs (MaxEntScan)[64], and SpliceAI[65]. These tools have been applied in our

previous studies where biological function of proteins aligned with in silico tool prediction[29,31].

Variants were then prioritized if they were identified in both sisters and predicted to be inherited as an autosomal dominant (heterozygous) or autosomal recessive (homozygous or compound heterozygous) trait. Variants were then prioritized if they were (a) nonsense, frameshift, or canonical splice site variants (±1-2 nucleotides away from the exon), (b) missense variants predicted to affect protein function by ≥5/7 in silico tools and highly conserved by ≥3/4 in silico tools, or (c) non-canonical splice site variants (>±2 nucleotides away from the exon) predicted to affect splicing by ≥3/4 in silico tools as these variants either will not encode a protein product (nonsense mediated decay) or could affect protein function. Variants were further prioritized if they had a MAF<1% in cancer-free controls of FC ancestry (n=1,208 alleles) because pathogenic variants are more likely to be rare based on the rare allele hypothesis[66]. The resulting variants are henceforth referred to as genetic landscape variants. *FANCI* c.1813C>T; p.L605F met all filtering and prioritization criteria and is included in tables as a reference but is not included in total variant counts.

3.3.3 Investigation of genetic landscape variants

Genes associated with genetic landscape variants that were identified in the *FANCI* c.1813C>T; p.L605F carriers were annotated for biological function, cellular location, encoded protein function, associated disease(s), and RNA expression in the ovaries and fallopian tubes using the Human Protein Atlas[67] (proteinatlas.org). These genes were annotated using the Cancer Hallmarks Analytics Tool[68] and a list of previously identified genes associated with hallmarks of cancer[69], which are defined as

various abilities or characteristics acquired by cells in the development of cancer[70–72]. Genes were also characterized based on being catalogued as having any somatic variants regardless of location in the same gene in TCGA PanCancer OC cases[46,73] (cbioportal.org) and for their association with disease in ClinGen[74].

Genetic landscape variants were annotated for carrier or allele frequency from available WES data of familial OC cases of FC ancestry negative for pathogenic variants in *BRCA1*, *BRCA2*, *BRIP1*, *RAD51C*, and *RAD51D* (n=13). We also investigated this OC study group for other variants in genes where genetic landscape variants had been identified as there may be allelic as well as genetic heterogeneity among CPGs. The same filtering and prioritization criteria were applied to OC cases of FC ancestry.

Genetic landscape variants were annotated for carrier frequency from available WES data of Australian HGSC cases. Other variants in genes where genetic landscape variants were identified were not investigated in this study group as these samples have been previously reported using a landscape approach[44].

3.3.4 Loss of heterozygosity analyses of *FANCI* c.1813C>T in OC tumour DNA from candidate variant carriers

Loss of heterozygosity analysis of *FANCI* c.1813C>T was analyzed by Sanger sequencing of OC tumour DNA from carriers. Extracted DNA from fresh frozen tumours was provided by the RRCancer biobank. Previously reported primers were used[31]. Sequencing chromatograms were inspected using 4peaks (nucleobytes.com/4peaks/index.html) visualization software.

3.3.5 Somatic genetic landscape of *FANCI* c.1813C>T carriers

Extracted DNA from fresh frozen HGSC tumours from *FANCI* c.1813C>T carriers (n=7) was provided by the RRCancer biobank. WES was performed at the McGill Genome Centre as previously described[31]. Annotated Variant Call Format files were inspected for variants in genes most commonly altered somatically in HGSC: *TP53*; *BRCA1*; *BRCA2*; *RB1*; *NF1*; *FAT3*; *CSMD3*; *GABRA6*; and *CDK12*[75].

Somatic copy number alteration (CNA) profiles were generated from WES data from tumour samples and corresponding matched-normal samples using Fraction and Allele specific Copy number Estimate from Tumour-normal Sequencing (FACETS) version 0.61[76]. Total and allele-specific read counts were extracted from tumour and normal samples based on common, polymorphic SNV loci from dbSNP version 150[77]. The following parameters were used for copy number segmentation: minimum total sample depth of 20, critical value for segmentation of 350, and minimum number of heterozygous single nucleotide variants (SNVs) to cluster segments of 100. Focal amplification of *CCNE1* was assessed as it is identified in over 20% of HGSC cases[75] and may be a therapeutic target for cyclin dependent kinase (CDK) inhibitors[78].

DeconstructSigs version 1.8.0[79] was used to determine the contribution of known mutational signatures associated with OC in each tumour sample. Catalog of Somatic Mutations in Cancer (COSMIC) Single Base Substitution (SBS) version 3.2 signatures were used as reference (cancer.sanger.ac.uk/signatures/documents/452/COSMIC_v3.2_SBS_GRCh37.txt: accessed May 16, 2022). Mutational signatures were compared to those associated with OC[80]. Synonymous and non-synonymous SNVs with at least three alternate reads

were used for mutational signature analysis. The number of SNVs per sample ranged from 65 to 2560.

The somatic genetic landscape of *FANCI* c.1813C>T carrier tumours from TCGA (n=6) was assessed for genes most commonly altered somatically in HGSC and focal amplification of *CCNE1* using cBioPortal[46,47] (cbioportal.org).

3.3.6 *FANCI* c.1813C>T germline carrier frequency across different cancer types from TCGA PanCancer Atlas

Data from the analysis of germline pathogenic variants in TCGA PanCancer Atlas cancer cases was downloaded[45]. A Variant Call Format file was generated with all *FANCI* c.1813C>T events identified at the germline level.

Clinical and genetic characteristics, including age at diagnosis and sex, of cancer cases harbouring germline *FANCI* c.1813C>T were retrieved from cBioPortal[46,47]. These characteristics were compared to the entire TCGA PanCancer Atlas study group[45].

3.3.7 Identification of somatic *FANCI* variants in different cancer types from TCGA PanCancer Atlas

All somatic *FANCI* variants were retrieved from cBioPortal[46,47] TCGA PanCancer Atlas Studies. Clinical and genetic characteristics of cancer cases where somatic *FANCI* variants had been identified were also retrieved. These included total variant count, microsatellite instability (MSI) score from microsatellite analysis for normal tumour instability (MANTIS)[81], age at cancer diagnosis, and sex. These characteristics were compared to the entire TCGA PanCancer Atlas study group.

3.3.8 Investigation of missense variants in *FANCI* reported in public databases

All missense variants in *FANCI* were retrieved from ClinVar[82] in March 2022. Variants were investigated using bioinformatic criteria established for *FANCI* c.1813C>T including those uncommon in gnomAD non-cancer controls (MAF 0.1-1%), highly conserved by ≥3/4 in silico tools, and predicted to affect protein function by ≥5/7 in silico tools. In silico tools used are the same as mentioned above.

3.3.9 Identification of variants in the FANCI protein interactome

Familial OC cases of FC ancestry negative for pathogenic variants in *BRCA1*, *BRCA2*, *BRIP1*, *RAD51C*, *RAD51D*, and *FANCI* (n=11) were investigated for variants in genes involved in the FANCI protein interactome. To construct a list of FANCI protein interactome genes in *Homo sapiens* we searched STRING[83], BioGRID[84], Protein Interaction Network Online Tool (PINOT)[85], Signalling Network Open Resource (SIGNOR)[86], Molecular Interaction Database (MINT)[87], Protein Interaction Knowledgebase (PICKLE)[88], Database of Interacting Proteins (DIP)[89], and IntAct[90] (Supplementary table 2). The literature was also searched for proteins shown to directly interact with FANCI protein experimentally by using “FANCI” as the search term (Supplementary table 2). As germline variants in DNA repair pathway genes have already been investigated in this study group (P.N. Tonin unpublished data) we focused on genes that were part of the FANCI interactome but not involved in DNA repair pathways (n=115). Variants were filtered and prioritized using the same criteria as described above (Figure 1a, b).

3.4 Results

3.4.1 Candidate variants identified in family F1528

We reported the discovery of *FANCI* c.1813C>T; p.L605F in two siblings in family F1528 as the most plausible OC predisposing candidate based on the association of FANCI in the Fanconi anemia homologous recombination (FA-HR) DNA repair pathway[31]. The only other variant shared between these siblings and other OC carriers of *FANCI* c.1813C>T (n=14), is *POLG* c.2492A>G; p.Y831C, a marker found in linkage disequilibrium with carriers of *FANCI* c.1813C>T in all the populations that we studied, and which remains an unlikely candidate OC predisposing allele based on purported function as reported previously[31]. Using bioinformatic tools and the most recent annotation of genetic databases, we re-evaluated the WES germline data of these siblings to further investigate the genetic landscape of carrier siblings reasoning that there may be other co-occurring potentially pathogenic variants of interest that could also be investigated in other OC families of the same FC ancestry. We only applied this strategy to this OC family (F1528), as we do not have any other examples of familial cases, especially sibling pairs affected with OC where both siblings harboured the same *FANCI* variant, which would facilitate the identification of candidate variants associated with disease.

Using the same generated WES data used in the discovery of *FANCI* c.1813C>T, we performed a new bioinformatic analysis and applied best performing predictive tools to identify germline genetic landscape variants that were shared among the siblings in family F1528 (Figure 1a). From a master list of 86,061 variants identified in both sisters, we used a filtering strategy to identify rare (MAF$\leq$1%), high quality

variants that were most likely to affect protein coding regions, which generated a list of 222 variants that were shared among these siblings. These variants were identified in 214 different genes and included *POLG* c.2492A>G, as expected (Supplementary table 3). The variants were present in the same genetic state in both siblings: heterozygous (n=196 variants), compound heterozygous (n=14 variants in 7 genes) or homozygous (n=1 variant), and heterozygous X chromosome linked (n=11). The MAF of these variants varied from $4x10^{-6}$ to $9.7x10^{-3}$ in gnomAD except for 14 variants that were not found in this database. These variants were comprised of 186 missense, 14 non-canonical splice site, 9 frameshift, 6 nonsense, 3 inframe, 3 canonical splice site, and 1 stop loss.

To further refine the list of plausible candidates, we applied criteria to the list of 222 variants to select those of interest for further investigation focusing on in silico tools, type of variant, and allele frequency in FC cancer-free controls (Figure 1b). Using these criteria, we identified 18 variants with nonsense, frameshift, or canonical splice site effects. We selected 57 missense candidates predicted to be damaging based on ≥5 of 7 of our best performing in silico tools, in keeping with the rationale that the most likely biologically relevant would be predicted by the majority of in silico tools[91]. We also selected an additional two non-canonical splice site variants that were predicted to affect splicing. As expected, this excluded *POLG* c.2492A>G as a variant of interest but not *FANCI* c.1813C>T. From this list of 76 variants, we selected 66 variants for further investigation based on their allele frequency (MAF<1%) in population matched FC controls. Excluding the *FANCI* variant, the 66 variants of interest were identified in 66 different genes (Supplementary table 4**).**

A large proportion of the 66 variants were identified at least once in various cancer contexts such as hallmarks of cancer[68,69] (83%) or somatically mutated in OC cases from the TCGA[46,73] (83%) as summarized in Figure 1c. Although some (9/66 ,14%) variants have been reported in other clinical contexts, none of the variants were found associated with a cancer context in ClinGen[74] (Supplementary Table 4). As shown in Supplementary Table 4, some of these genes were found in more than one group as defined in Figure 1c. None of the 66 variants were identified in genes associated with DNA repair pathways.

3.4.2 Genetic analyses of variants identified in *FANCI* carrier siblings in FC study groups

To further characterize our 66 variants, we reviewed available WES data from 13 familial OC cases negative for pathogenic variants in *BRCA1*, *BRCA2*, *BRIP1*, *RAD51C*, and *RAD51D* for carrier status. These cases have been well characterized and are of FC ancestry[25,26,28–31,38,92–99]. We identified a total of four carriers from three families of variants in *PTPN22, GPD1,* and *SEC14L4* (Table 1). Notable is that none of three *FANCI* c.1813C>T carriers from independently ascertained cases harboured any of the 66 variants. Thus, though there may be shared FC ancestry among the carriers, none of the five independently ascertained familial OC cases in our FC study group of *FANCI* carriers harbour other potentially deleterious alleles initially identified in the index *FANCI* c.1813C>T OC cases.

Given the possibility of allelic heterogeneity even within the FC population as we have previously demonstrated with established OC predisposing genes[24], we screened the same 13 familial OC cases for other plausible deleterious variants in the 66

candidate genes. Applying our bioinformatic filtering and prioritization criteria we identified 10 different variants that were not present in our index family F1528. There were six carriers of variants in *PIWIL3*, *SCN10A*, *PCDH15*, *TEX2*, *DNAH3*, *DNAH1*, *IQCA1*, *CACNA1S*, and *MYO7A* (Table 2). Four cases were found to carry variants in two different genes: F1085-PT0134 (*CACNA1S* and *MYO7A)*; F845-PT0196 (*PCDH15* and *TEX2*); F1506-PT0136 (*PIWIL3* and *DNAH1*); and F1543-PT0137 (*DNAH1* and *IQCA1*). Both variants in F1085-PT0134 were not carried by their sibling F1085-PT0135 and thus did not segregate with the disease. Each variant was harboured by only one case; two different variants were identified in the same gene, *DNAH1*.

3.4.3 Genetic analyses of variants identified in *FANCI* carrier siblings in non-FC study groups

We screened our 66 variants in available WES data from the Australian HGSC study group regardless of *FANCI* variant carrier status. We identified 70 carriers of 21 variants among 516 HGSC cases, where the majority harboured only one variant (Table 3). None of these variants were identified in any of the previously identified 10 Australian carriers of *FANCI* c.1813C>T[31]. Two carriers of other likely pathogenic *FANCI* variants (c.1264G>A; p.G422R and c.3635T>C; p.F1212S) were found to harbour variants in *ALDH16A1* and *NBAS*, respectively. Though different variants were identified in OC cases of FC ancestry in *DNAH3*, *IQCA1*, and *PCDH15*, the same variants as those found in family F1528 sisters were identified in these genes in five Australian HGSC cases.

3.4.4 Genetic analyses of germline *FANCI* interactome variants identified in FC OC cases

Based on our previous analyses, *FANCI* c.1813C>T is the strongest candidate OC predisposing variant identified in family F1528. Therefore, we used a candidate gene approach to investigate germline variants in genes that encode proteins that are part of the FANCI interactome. We reviewed available WES data from 11 familial OC cases negative for pathogenic variants in *BRCA1*, *BRCA2*, *BRIP1*, *RAD51C*, *RAD51D,* and *FANCI* to search for other candidate predisposition variants. Variants in DNA repair pathway genes were excluded as this has been previously reported by our lab (P.N. Tonin unpublished data). We identified a total of three carriers from three families of missense variants in *EZH2*, *ANKRD55*, *MOV10*, and *LRRK2* (Table 4). Variants in *ANKRD55* and *MOV10* were identified in the same case F1506-PT0136.

3.4.5 Identification of other germline potentially deleterious variants in *FANCI*

To identify additional germline potentially pathogenic variants in *FANCI,* we used ClinVar[82] which aggregates information about genomic variation and its relationship to human health. We focused our analyses on missense variants as loss of function variants (i.e., frameshift, nonsense, and canonical splice site) have been previously reported by our group[31]. We investigated 319 missense variants using bioinformatic criteria established for selecting variants of interest: uncommon (0.1-1% MAF) in non-cancer controls from gnomAD, highly conserved, and predicted to affect protein function by in silico tools. Three missense variants met these criteria (Figure 2): *FANCI* c.286G>A; p.E96K, c.1573A>G; p.M525V, and our candidate variant c.1813C>T; p.L605F. *FANCI* c.1573A>G; p.M525V was previously reported by our group, however,

based on in cellulo assays we concluded that this variant did not affect protein function[31]. Thus, the only *FANCI* variant to investigate further was c.286G>A; p.E96K. We investigated available genetic data or genotyped FC, Australian, and TCGA study groups for carriers of *FANCI* c.286G>A; p.E96K. *FANCI* c.286G>A was not identified in OC (n=527) or BC (n=220) cases or controls (n=171) of FC ancestry, but was previously reported by our group in an Australian OC case (1/516, 0.2%) and controls (5/4878, 0.1%) and TCGA OC cases (1/412, 0.2%)[31].

3.4.6 Loss of heterozygosity analyses of *FANCI* c.1813C>T in OC tumour DNA from carriers

We previously reported loss of the wild type allele in bilateral OC tumours from *FANCI* c.1813C>T carriers, suggesting that loss of FANCI function was an early event in tumourigenesis[31]. We have extended our analysis to investigate tumour samples from other carriers though only DNA from FC cases was available for these analyses. Inspection of Sanger sequencing chromatograms from OC tumour DNA from nine carriers revealed three cases exhibiting loss of the wild type allele and retention of the *FANCI* c.1813C>T allele. One case showed loss of the variant allele and retention of the wild type allele, and the remaining cases retained heterozygosity with little evidence of allelic imbalance.

3.4.7 Somatic genetic analyses of OC tumours from *FANCI* c.1813C>T carriers

The somatic genetic landscape of HGSC cases has been well characterized, where there is a long-tail of uncommon somatic variants and extensive genome wide CNAs, with the exception of *TP53* (which harbours driver mutations that cause cells to become cancerous) being the most altered gene (>95% of cases)[75]. To determine if

HGSC cases carrying *FANCI* c.1813C>T exhibit similar somatic genetic characteristics to HGSC cases we performed WES analyses or surveyed available genetic data from seven FC cases and six TCGA cases, respectively. We focused our analyses on the most altered genes reported for HGSCs: *TP53* (96%), *BRCA1* (3.5%), *CSMD3* (6%), *NF1* (4%), *CDK12* (3%), *FAT3* (6%), *GABRA6* (2%), *BRCA2* (3%), and *RB1* (2%)[75]. Somatic variants were identified in *CDK12* (1/13, 8%), *FAT3* (3/13, 23%), *BRCA2* (3/13, 23%), and *TP53* (11/13, 85%) (Table 5), at frequencies comparable to those HGSC cases[75]. As expected, most deleterious variants identified in our HGSC cases carrying *FANCI* c.1813C>T in *TP53* were missense[75,100,101] (n=7), with the remainder being frameshift (n=2), splice (n=1), or inframe indel (n=1). Extensive and unremarkable genome-wide CNAs were evident across tumours from *FANCI* c.1813C>T carriers consistent with that seen for HGSC tumours (Supplementary Figure 1). Amplification of *CCNE1*, reported to occur in approximately 20% of HGSC cases[75], was exhibited in two *FANCI* carrier cases (2/11, 18%), PT0006 from our FC study group and sample TCGA-25-2393 from TGCA project.

DNA from HGSC tumours has been shown to exhibit global DNA mutational signatures reflecting disruptions in specific DNA repair pathways, aging, and other processes that have accumulated during tumourigenesis[80]. We performed a somatic mutational signature analysis using WES data derived from FC OC tumour DNA from *FANCI* c.1813C>T carriers using COSMIC SBS signatures as a reference. The signatures identified in tumours from *FANCI* c.1813C>T carriers were compared to those exhibited by HGSC tumours as there have been no reports attributing mutational signatures to cancers harbouring deleterious *FANCI* variants. We were able to profile

seven OC tumour samples from FC carriers due to availability of WES data for these samples (Supplementary Figure 2). The mutational profiles were indicative of the presence of extensive and complex mutational patterns typified by HGSC tumours (https://signal.mutationalsignatures.com/explore/tissueType/15). The homologous recombination deficiency signature (referred to as SBS3) was identified in 6/7 (86%) cases. The sample that did not exhibit this SBS3 signature, PT0003, exhibited signature pattern SBS8, a signature whose etiology is unknown, but it has been proposed to be associated with homologous recombination deficiency (cancer.sanger.ac.uk/signatures/sbs/sbs8/). PT0003 also exhibited the largest contribution of signature SBS6, which has been attributed to defective mismatch repair and MSI. The aging signature (SBS1) was identified in 5/7 (71%) tumours and the contribution was consistent to age at diagnosis[31]. All tumours exhibited varying proportions of SBS18, a signature indicative of damage due to reactive oxygen species. A signature with a proposed etiology associated with prior treatment with platinum chemotherapy drugs (SBS35) was evident in 5/7 (71%) tumours, though not in sample PT0007, which was from a patient who received neoadjuvant chemotherapy with the platinum compound carboplatin.

3.4.8 Germline *FANCI* c.1813C>T carriers identified in other cancer types

We previously reported carriers of *FANCI* c.1813C>T and other potentially pathogenic *FANCI* variants in BC cases, a disease associated with OC risk genes[102], and a review of the literature also indicated that there were *FANCI* carriers in other cancer types[31]. To further investigate the role of *FANCI* in other cancer types we investigated the germline carrier frequency of *FANCI* c.1813C>T in 10,389 cancer

cases from TCGA PanCancer data set[45]. We focused on this variant to further investigate its association with familial OC and our in cellulo assays demonstrating abrogated protein function[31]. The highest carrier frequency was observed in adrenocortical carcinoma cases (3.3%, 3/92); there were no carriers identified in cases with diffuse large B-cell carcinoma (n=41), kidney renal papillary cell carcinoma (n=289), thymoma (n=123), or uterine carcinosarcoma (n=57) (Table 6). The median age of diagnosis (59 ± 14.7 years) and number of females (48.5%) of *FANCI* c.1813C>T carriers were comparable to the total study group (59.2 ± 14.4[45] and 52%[45], respectively). Interestingly, the overall carrier frequency of *FANCI* c.1813C>T at 1.6% (171/10389) was significantly higher in TCGA PanCancer cases than in that of non-cancer individuals in gnomAD (1.3%, 1787/134,164; Pearson's $\chi^2 = 7.3$, $p = 0.007$).

3.4.9 A wide spectrum of somatic *FANCI* variants identified in a variety of cancer types

Approximately 40% of germline CPGs have been found with somatic variants in tumours as drivers and some of these genes have been shown to play a role in tumourigenesis, with *RB1* being the classical example[103]. From the above analysis of the germline *FANCI* variant there were four *FANCI* c.1813C>T germline carriers with different somatic *FANCI* variants, two bladder urothelial carcinomas and two lung squamous cell carcinomas. These observations prompted us to investigate the spectrum and frequency of somatic variants in *FANCI* in TCGA PanCancer tumours (n=10,434) from cBioPortal[46,47]. We identified 198 different variants in 172 tumours (1.65%, 172/10434) in 28 different cancer types comprised of a variety of genetic abnormalities: 168 missense, 11 nonsense, 10 splice, 6 frameshift, 2 stop loss, and 1 inframe deletion (Table 7 and Supplementary table 5). There appears to be no

mutational hotspot identified and variants were distributed across the gene (Figure 3). The highest total number of variants was identified in uterine corpus endometrial carcinoma tumours (8.32%, 43/517); no variants were identified in tumours from cases with cholangiocarcinoma (n=36), diffuse large B-cell carcinoma (n=41), kidney chromophobe (n=65), pancreatic adenocarcinoma (n=179), or uveal melanoma (n=80). The median age at diagnosis and sex of cases with somatic *FANCI* variants were comparable to those without somatic *FANCI* variants (60 ± 13.5 vs. 60 ± 14.4 and 58% vs. 52%, respectively). Tumours with somatic *FANCI* variants had higher mutational load ($p = 2.2x10^{-16}$) and MSI score ($p = 6.8x10^{-10}$) compared to tumours without somatic *FANCI* variants.

3.5 Discussion

Genetic analyses of the germline of two sisters with OC from family F1528 that are heterozygous for *FANCI* c.1813C>T; p.L605F revealed 222 variants of interest, of which 66 were most likely to exert a function on the encoded protein (genetic landscape variants). Of note is the number of loss-of-function (frameshift, nonsense, and canonical splice site) variants identified in both sisters after filtering and prioritization (n=15). Increasing evidence suggests that all individuals carry more potentially deleterious variants than previously suspected (approximately 24-100 heterozygous variants per individual[104,105]) although some may not impact gene function[106,107]. As adequate in silico tools with high predictive performance, such as those for missense and splice site variants, have not yet been developed for loss-of-function variants, it is difficult to further interpret these variants in the absence of laboratory experiments examining their biological effect. The investigation of the genetic landscape variants in OC cases of FC

ancestry revealed four carriers of variants in *PTPN22, GPD1,* and *SEC14L4* in three different families, though two of these families were previously identified by our group to harbour likely pathogenic variants in known or putative DNA repair pathway genes (P.N. Tonin unpublished data). When we assessed the same study group for carrier status, we identified six carriers of variants in nine different genes, where all but two families were previously identified to harbour likely pathogenic variants in DNA repair genes (P.N. Tonin unpublished data). Families F845 and F1543 were the only cases not found to harbour likely pathogenic variants in DNA repair genes. Although the previously identified DNA repair pathway variants are plausible OC predisposing variants they have yet to be verified independently. Though it is possible that genetic variants identified in this study may independently affect risk in OC, *FANCI* c.1813C>T; p.L605F remains the most likely candidate OC predisposing gene in family F1528.

We also investigated the 66 genetic landscape variants for carrier frequency in Australian OC cases (regardless of *FANCI* variant carrier status) and identified several carriers of variants in these genes, mainly of missense variants. These variants had not been reported in a previous independent analysis of WES data from these cases as the study was focused primarily on loss of function variants[44]. Thus, it is possible that these missense variants are relevant in these cases, which are notably negative for pathogenic variants in *BRCA1* and *BRCA2*. Also notable is that we did not identify any carriers of any of our genetic landscape variants among the 10 previously identified *FANCI* c.1813C>T; p.L605F carriers from this study group.

The role of these 66 genetic variants identified in *FANCI* carriers in modifying risk to OC in family F1528 which harbours *FANCI* c.1813C>T; p.L605F remains to be

determined. Interestingly, additional potentially pathogenic variants were identified in *FANCI* using the criteria established for c.1813C>T and it would be of interest to further assess their effect on gene function. The most interesting variant, *FANCI* c.286G>A; p.E96K, was identified in a BC case (diagnosed at 44 and 50 years) who had OC at age 53 years, was of Ukrainian ancestry, who reported a grandmother with OC (Supplementary figure 3). This individual did not harbour any pathogenic variants in *BRCA1*, *BRCA2*, *BRIP1*, *RAD51C*, or *RAD51D* by WES analyses. *FANCI* c.286G>A; p.E96K was reported in BC cases (2/133, 1.5%)[108] and OC cases (1/6385, 0.02% [0/6115 controls])[109] in the literature. FANCI p.E96K may affect the ubiquitination of FANCD2, the heterodimeric binding partner of FANCI, and/or the Van der Waals forces between FANCI and FANCD2[186].

HGSC tumours from cases harbouring germline *FANCI* c.1813C>T exhibited features consistent with tumours from HGSC cases. The majority (85%) of tumours had identifiable pathogenic variants in *TP53* and two cases were identified with *CCNE1* amplification (18%). The somatic mutational signatures characteristic of HGSC tumour cells were also present in HGSC samples harbouring *FANCI* c.1813C>T. We also identified signatures SBS35 (platinum associated) and SBS6 (mismatch repair deficiency) in 5/7 and 6/7 cases respectively, which are less commonly observed in HGSC cases (<10%). This may be a result of small sample size, though they could also be a result of harbouring *FANCI* c.1813C>T in the germline. Interestingly, the two cases with the highest contribution of SBS6 signature, samples PT0004 and PT0005, have germline or somatic variants in mismatch repair genes, which may contribute to the presence of this signature. It has recently been reported in a study involving *C. elegans,*

that genotoxic agents tended to have a stronger influence on the mutational signature than a DNA repair deficient background[110]. Moreover, the signature attributed to defects in HR DNA repair may be identified in the absence of an identifiable DNA repair pathway gene pathogenic variant[111]. This could be because the signature is more attributed to global defects, which is distinct from other signatures that have characteristic nucleotide changes, such as the aging signature[111]. Although the sample size is small, there appeared to be no identifiable *FANCI* specific signature from analyses of tumours from *FANCI* c.1813C>T carriers.

The contribution of cancer cases attributed to CPGs is approximately 3%, though this varies based on cancer type[103]. Some CPGs predispose to multiple primary cancer types, such as *BRCA1* with OC and BC, though there is often preferential predisposition to certain histological subtypes, such as the association of *BRCA1* with HGSC. Germline *FANCI* c.1813C>T was initially identified in HGSC cases[31], but as shown in our study can be observed across many cancer types in TGCA PanCancer Atlas. It is unknown if *FANCI* c.1813C>T is contributing to risk in these cancers as the variant is more common in the general population compared to other high risk CPGs (0.6% vs. 0.0001%)[48]. The moderately increased carrier frequency of *FANCI* c.1813C>T in TCGA cancer cases (1.6%) compared to gnomAD cancer-free controls (1.3%) is intriguing suggesting a role for this variant in risk (p = 0.007). Additionally, TCGA cancer cases were identified with somatic variants in *FANCI* across cancer types. As variants were identified in cases with significantly higher mutational load and/or MSI scores, it is possible that these somatic *FANCI* variants arose as a consequence of either of these processes, though they could be drivers of these processes through mechanisms that

remain to be elucidated. *FANCI* is not currently included among the 733 cancer-driving genes in the Cancer Gene Census[112] (cancer.sanger.ac.uk/census#cl_sub_tables), but a number of other genes involved in the FA pathway, such as *BRCA1 (FANCS)*, *BRCA2 (FANCD1)*, *BRIP1 (FANCJ)*, *FANCA*, *FANCC*, *FANCD2*, *PALB2 (FANCN),* and *RAD51C (FANCO)*, are implicated as cancer-drivers in this census. Further assessment of cancers in individuals harbouring germline *FANCI* c.1813C>T or somatic *FANCI* variants may elucidate a genetic signature associated with *FANCI*, as has been reported for FA associated squamous cell carcinomas[113].

Many CPGs are also associated with non-cancer phenotypes, the spectrum of which is broad. Homozygous or compound heterozygous variants in *FANCI* were associated with FA group I in 2007[114–116]. FA is a rare disease characterized by congenital defects, progressive bone marrow failure, and increased risk of cancers (mainly acute myeloid leukemia and squamous cell carcinomas)[117]. FA-I cases comprise approximately 1% of all FA cases and have been associated with at least three features of VACTERL-H[118], a rare disease that affects multiple body systems. Recently, an eight-year-old male with aplasia referred for a diagnosis of FA was reported to harbour germline homozygous *FANCI* c.1813C>T[119]. FANCD2 ubiquitination was not detected in peripheral blood cells from this patient and increased chromosomal breakage was observed, suggesting abrogation of the FA pathway. These data are consistent with our previous observations that *FANCI* c.1813C>T abrogates FANCI protein function[31]. FANCI is an integral member of the FA pathway and acts as the molecular switch to activate the pathway[120]. FANCI also functions outside the FA

pathway such as in dormant origin firing[121], negative regulation of Akt signaling[122], and ribosome biogenesis[123].

Risk of cancer has been assessed in heterozygous relatives of individuals with FA, and though no association with cancer risk was found, few *FANCI* families (n=4) have been investigated due to the paucity of *FANCI* carriers[124–126]. *FANCI* c.1813C>T; p.L605F was previously identified by our group in the germline and was associated with a suspected autosomal dominant mode of inheritance of OC[31]. This is consistent with more than half of CPGs which are associated with an autosomal dominant mode of inheritance[103]. The majority of CPGs also act as tumour suppressors, where many are classical tumour suppressors that require biallelic inactivation for tumour development and/or progression, though some CPGs may exert their effect through haploinsufficiency or a dominant-negative manner[103]. Here, we have shown that biallelic inactivation of *FANCI* c.1813C>T may occur through loss of the wild type allele. As we showed that not all carriers exhibited this loss, it is possible that loss of the wild type allele may not be required for OC tumourigenesis. This is consistent with the OC predisposing genes *BRCA1* and *BRCA2*, where loss of the wild type allele is not always observed in tumour cells from carriers of pathogenic variants in these CPGs[69,187]. Beyond assessing the biological effect of a variant using cell line models as we have shown with *FANCI* c.1813C>T[31], there are no suitable animal models to evaluate OC risk alleles.

Identification of CPGs has had a large clinical impact on diagnosis and management, targeted therapies, and screening and prevention. The clinical utility of *FANCI* for diagnosis and management cannot be determined until penetrance for

cancer risk is established. Currently there are no effective cancer screening methods for OC or prevention strategies to reduce OC risk in the general population, though prophylactic salpingo-oophorectomy has been proven to reduce risk in carriers of pathogenic *BRCA1* or *BRCA2* variants[128]. Though there are no targeted therapies for *FANCI*, 73 chemicals interact with this gene, including cisplatin and mitomycin C, which is concordant with our previous findings that loss of *FANCI* sensitizes cells to these drugs[124]. There are 12 cancer related drugs that interact with FANCI, 7 of which are chemotherapies and 5 of which are targeted therapies[129]. These chemicals present opportunities for future investigation for the treatment of cancer cases with *FANCI* variants.

3.6 Conclusion

This study has expanded on the molecular genetic characteristics of *FANCI* c.1813C>T; p.L605F in OC first reported in OC families of FC ancestry[31]. These data suggest *FANCI* c.1813C>T carrier HGSC tumours show characteristics known to be exhibited by HGSC cases. The identification of germline *FANCI* c.1813C>T carriers and various somatic *FANCI* variants across cancer types suggests a possible involvement of *FANCI* in other cancers and an avenue for future research.

3.7 References

1. Miki, Y. *et al.* Strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. *Science* **266**, 66–71 (1994).

2. Wooster, R. *et al.* Identification of the breast cancer susceptibility gene BRCA2. *Nature* **378**, 789–92 (1995).

3. Song, H. *et al.* Tagging single nucleotide polymorphisms in the BRIP1 gene and susceptibility to breast and ovarian cancer. *PLoS ONE* **2**, e268–e268 (2007).

4. Rafnar, T. *et al.* Mutations in BRIP1 confer high risk of ovarian cancer. *Nat. Genet.* **43**, 1104–1107 (2011).

5. Meindl, A. *et al.* Germline mutations in breast and ovarian cancer pedigrees establish RAD51C as a human cancer susceptibility gene. *Nat. Genet.* **42**, 410–414 (2010).

6. Zheng, Y. *et al.* Screening RAD51C nucleotide alterations in patients with a family history of breast and ovarian cancer. *Breast Cancer Res. Treat.* **124**, 857–861 (2010).

7. Somyajit, K., Subramanya, S. & Nagaraju, G. RAD51C: A novel cancer susceptibility gene is linked to Fanconi anemia and breast cancer. *Carcinogenesis* **31**, 2031–2038 (2010).

8. Loveday, C. *et al.* Germline mutations in RAD51D confer susceptibility to ovarian cancer. *Nat. Genet.* **43**, 879–882 (2011).

9. Yang, X. *et al.* Cancer Risks Associated With Germline PALB2 Pathogenic Variants: An International Study of 524 Families. *J. Clin. Oncol.* **38**, 674–685 (2020).

10. Pavanello, M. *et al.* Rare germline genetic variants and the risks of epithelial ovarian cancer. *Cancers* **12**, 3046–3046 (2020).

11. Papadopoulos, N. *et al.* Mutation of a mutL homolog in hereditary colon cancer. *Science* **263**, 1625–1629 (1994).

12. Bronner, C. E. *et al.* Mutation in the DNA mismatch repair gene homologue hMLH1 is associated with hereditary non-polyposis colon cancer. *Nature* **368**, 258–261 (1994).

13. Fishel, R. *et al.* The human mutator gene homolog MSH2 and its association with hereditary nonpolyposis colon cancer. *Cell* **75**, 1027–1038 (1993).

14. Miyaki, M. *et al.* Germline mutation of MSH6 as the cause of hereditary nonpolyposis colorectal cancer. *Nat. Genet.* **17**, 271–272 (1997).

15. Nicolaides, N. C. *et al.* Mutations of two PMS homologues in hereditary nonpolyposis colon cancer. *Nature* **371**, 75–80 (1994).

16. Pennington, K. P. *et al.* Germline and Somatic Mutations in Homologous REcombination Genes Predict Platinum Response and Survival in Ovarian, Fallopian Tube, and Peritoneal Carcinomas. *Clin Cancer Res* **20**, 764–775 (2014).

17. Lu, H. M. *et al.* Association of breast and ovarian cancers with predisposition genes identified by large-scale sequencing. *JAMA Oncol.* **5**, 51–57 (2019).

18. Lilyquist, J. *et al.* Frequency of mutations in a large series of clinically ascertained ovarian cancer cases tested on multi-gene panels compared to reference controls. *Gynecol. Oncol.* **147**, 375–380 (2017).

19. Norquist, B. M. *et al.* Inherited Mutations in Women With Ovarian Carcinoma. *JAMA Oncol.* **2**, 482–490 (2016).

20. Weber-Lassalle, N. *et al.* Germline loss-of-function variants in the BARD1 gene are associated with early-onset familial breast cancer but not ovarian cancer. *Breast Cancer Res. BCR* **21**, 55–55 (2019).

21. Alenezi, W. M., Fierheller, C. T., Recio, N. & Tonin, P. N. Literature review of BARD1 as a cancer predisposing gene with a focus on breast and ovarian cancers. *Genes* **11**, 856–856 (2020).

22. Kanchi, K. L. *et al.* Integrated analysis of germline and somatic variants in ovarian cancer. *Nat. Commun.* **5**, 3156–3156 (2014).

23. Dicks, E. *et al.* Germline whole exome sequencing and large-scale replication identifies FANCM as a likely high grade serous ovarian cancer susceptibility gene. *Oncotarget* **8**, 50930–50940 (2017).

24. Fierheller, C. T., Alenezi, W. M. & Tonin, P. N. The Genetic Analyses of French Canadians of Quebec Facilitate the Characterization of New Cancer Predisposing Genes Implicated in Hereditary Breast and/or Ovarian Cancer Syndrome Families. *Cancers* **13**, (2021).

25. Tonin, P. N. *et al.* Founder BRCA1 and BRCA2 mutations in French Canadian breast and ovarian cancer families. *Am. J. Hum. Genet.* **63**, 1341–1351 (1998).

26. Oros, K. K. *et al.* Significant proportion of breast and/or ovarian cancer families of French Canadian descent harbor 1 of 5 BRCA1 and BRCA2 mutations. *Int. J. Cancer* **112**, 411–419 (2004).

27. Foulkes, W. D. *et al.* Identification of a novel truncating PALB2 mutation and analysis of its contribution to early-onset breast cancer in French-Canadian women. *Breast Cancer Res.* **9**, R83–R83 (2007).

28. Tischkowitz, M. *et al.* Analysis of PALB2/FANCN-associated breast cancer families. *PNAS* **104**, 6788–6793 (2007).

29. Alenezi, W. M. *et al.* The Genetic and Molecular Analyses of RAD51C and RAD51D Identifies Rare Variants Implicated in Hereditary Ovarian Cancer from a Genetically Unique Population. *Cancers* **14**, (2022).

30. Rivera, B. *et al.* Functionally null RAD51D missense mutation associates strongly with ovarian carcinoma. *Cancer Res.* **77**, 4517–4529 (2017).

31. Fierheller, C. T. *et al.* A functionally impaired missense variant identified in French Canadian families implicates FANCI as a candidate ovarian cancer-predisposing gene. *Genome Med.* **13**, 186 (2021).

32. Callahan, M. J. *et al.* Primary fallopian tube malignancies in BRCA-positive women undergoing surgery for ovarian cancer risk reduction. *J. Clin. Oncol.* **25**, 3985–3990 (2007).

33. Carlson, J. W. *et al.* Serous tubal intraepithelial carcinoma: Its potential role in primary peritoneal serous carcinoma and serous cancer prevention. *J. Clin. Oncol.* **26**, 4160–4165 (2008).

34. Finch, A. *et al.* Clinical and pathologic findings of prophylactic salpingo-oophorectomies in 159 BRCA1 and BRCA2 carriers. *Gynecol. Oncol.* **100**, 58–64 (2006).

35. Labidi-galy, S. I. *et al.* High grade serous ovarian carcinomas originiate in the fallopian tube. *Nat. Commun.* **8**, (2017).

36. Leeper, K. *et al.* Pathologic findings in prophylactic oophorectomy specimens in high-risk women. *Gynecol. Oncol.* **87**, 52–56 (2002).

37. Powell, C. B. *et al.* Risk-reducing salpingo-oophorectomy in BRCA mutation carriers: role of seial sectioning in the setection of occult malignancy. *J. Clin. Oncol.* **23**, 127–132 (2005).

38. Tonin, P. N., Maugard, C. M., Perret, C., Mes-Masson, A. M. & Provencher, D. M. A review of histopathological subtypes of ovarian cancer in BRCA-related French Canadian cancer families. *Fam. Cancer* **6**, 491–497 (2007).

39. Tonin, P. N., Mes-Masson, A.-M., Narod, S. A., Ghadirian, P. & Provencher, D. Founder BRCA1 and BRCA2 mutations in French Canadian ovarian cancer cases unselected for family history. *Clin. Genet.* **55**, 318–324 (1999).

40. Hodgkinson, A. *et al.* High-resolution genomic analysis of human mitochondrial RNA sequence variation. *Science* **344**, 413–415 (2014).

41. Hussin, J. G. *et al.* Recombination affects accumulation of damaging and disease-associated mutations in human populations. *Nat. Genet.* **47**, 400–404 (2015).

42. Peischl, S. *et al.* Relaxed Selection During a Recent Human Expansion. *Genetics* **208**, 763–777 (2018).

43. Guillemette, L. *et al.* Genetics of Glucose regulation in Gestation and Growth (Gen3G): a prospective prebirth cohort of mother–child pairs in Sherbrooke, Canada. *BMJ Open* **6**, e010031–e010031 (2016).

44. Subramanian, D. N. *et al.* Exome sequencing of familial high-grade serous ovarian carcinoma reveals heterogeneity for rare candidate susceptibility genes. *Nat. Commun.* **11**, (2020).

45. Huang, K.-L. *et al.* Pathogenic germline variants in 10,389 adult cancers. *Cell* **173**, 355-370.e14 (2018).

46. Cerami, E. *et al.* The cBio Cancer Genomics Portal: An open platform for exploring multidimensional cancer genomics data. **2**, 401–404 (2014).

47. Gao, J. *et al.* Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci. Signal.* **6**, pl1 (2013).

48. Karczewski, K. J. *et al.* The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **581**, 434–443 (2020).

49. Robinson, J. T. *et al.* Integrative genomics viewer. *Nat. Biotechnol.* **29**, 24–26 (2011).

50. 1000 Genomes Project Consortium *et al.* A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).

51. Lek, M. *et al.* Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285–291 (2016).

52. Ghosh, R., Oak, N. & Plon, S. E. Evaluation of in silico algorithms for use with ACMG/AMP clinical variant interpretation guidelines. *Genome Biol.* **18**, (2017).

53. Davydov, E. V., Goode, D. L., Sirota, M., Cooper, G. M. & Sidow, A. Identifying a high fraction of the human genome to be under selective constraint using GERP ++. *PLoS Comput. Biol.* **6**, e1001025–e1001025 (2010).

54. Garber, M. *et al.* Identifying novel constrained elements by exploiting biased substitution patterns. *Bioinformatics* **25**, i54–i62 (2009).

55. Pollard, K. S., Hubisz, M. J., Rosenbloom, K. R. & Siepel, A. Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Res.* **20**, 110–121 (2010).

56. Siepel, A. *et al.* Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res.* **15**, 1034–1050 (2005).

57. Rentzsch, P., Witten, D., Cooper, G. M., Shendure, J. & Kircher, M. CADD: predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Res.* **47**, D886–D894 (2019).

58. Ionita-Laza, I., McCallum, K., Xu, B. & Buxbaum, J. D. A spectral approach integrating functional genomic annotations for coding and noncoding variants. *Nat. Genet.* **48**, 214–220 (2016).

59. Choi, Y., Sims, G. E., Murphy, S., Miller, J. R. & Chan, A. P. Predicting the Functional Effect of Amino Acid Substitutions and Indels. *PLoS ONE* **7**, e46688–e46688 (2012).

60. Dong, C. *et al.* Comparison and integration of deleteriousness prediction methods for nonsynonymous SNVs in whole exome sequencing studies. *Hum. Mol. Genet.* **24**, 2125–2137 (2015).

61. Ioannidis, N. M. *et al.* REVEL: An Ensemble Method for Predicting the Pathogenicity of Rare Missense Variants. *Am. J. Hum. Genet.* **99**, 877–885 (2016).

62. Douville, C. *et al.* Assessing the Pathogenicity of Insertion and Deletion Variants with the Variant Effect Scoring Tool (VEST-Indel). *Hum. Mutat.* **37**, 28–35 (2016).

63. Jian, X., Boerwinkle, E. & Liu, X. In silico prediction of splice-altering single nucleotide variants in the human genome. *Nucleic Acids Res.* **42**, 13534–13544 (2014).

64. Shamsani, J. *et al.* A plugin for the Ensembl Variant Effect Predictor that uses MaxEntScan to predict variant spliceogenicity. *Bioinforma. Oxf. Engl.* **35**, 2315–2317 (2019).

65. Jaganathan, K. *et al.* Predicting Splicing from Primary Sequence with Deep Learning. *Cell* **176**, 535-548.e24 (2019).

66. Gibson, G. Rare and common variants: Twenty arguments. *Nat. Rev. Genet.* **13**, 135–145 (2012).

67. Uhlen, M. *et al.* Tissue-based map of the human proteome. *Science* **347**, 1260419–1260419 (2015).

68. Baker, S. *et al.* Cancer Hallmarks Analytics Tool (CHAT): a text mining approach to organize and evaluate scientific literature on cancer. *Bioinformatics* **33**, 3973–3981 (2017).

69. Menyhárt, O. *et al.* Guidelines for the selection of functional assays to evaluate the hallmarks of cancer. *Biochim. Biophys. Acta* **1866**, 300–319 (2016).

70. Hanahan, D. & Weinberg, R. A. The hallmarks of cancer. *Cell* **100**, 57–70 (2000).

71. Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: The next generation. *Cell* **144**, 646–674 (2011).

72. Hanahan, D. Hallmarks of Cancer: New Dimensions. *Cancer Discov.* **12**, 31–46 (2022).

73. Gao, J. *et al.* Integrative Analysis of Complex Cancer Genomics and Clinical Profiles Using the cBioPortal. *Biophys. Chem.* **257**, 2432–2437 (2013).

74. Rehm, H. L. *et al.* ClinGen — The Clinical Genome Resource. *N. Engl. J. Med.* **372**, 2235–2242 (2015).

75. Bell, D. *et al.* Integrated genomic analyses of ovarian carcinoma. *Nature* **474**, 609–615 (2011).

76. Shen, R. & Seshan, V. E. FACETS: allele-specific copy number and clonal heterogeneity analysis tool for high-throughput DNA sequencing. *Nucleic Acids Res.* **44**, e131 (2016).

77. Sherry, S. T. *et al.* dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.* **29**, 308–311 (2001).

78. Etemadmoghadam, D. *et al.* Resistance to CDK2 inhibitors is associated with selection of polyploid cells in CCNE1-amplified ovarian cancer. *Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res.* **19**, 5960–5971 (2013).

79. Rosenthal, R., McGranahan, N., Herrero, J., Taylor, B. S. & Swanton, C. DeconstructSigs: delineating mutational processes in single tumors distinguishes DNA repair deficiencies and patterns of carcinoma evolution. *Genome Biol.* **17**, 31 (2016).

80. Degasperi, A. *et al.* Substitution mutational signatures in whole-genome-sequenced cancers in the UK population. *Science* **376**, science.abl9283 (2022).

81. Kautto, E. A. *et al.* Performance evaluation for rapid detection of pan-cancer microsatellite instability with MANTIS. *Oncotarget* **8**, 7452–7463 (2017).

82. Landrum, M. J. *et al.* ClinVar: Improving access to variant interpretations and supporting evidence. *Nucleic Acids Res.* **46**, 1062–1067 (2018).

83. Szklarczyk, D. *et al.* The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Res.* **49**, D605–D612 (2021).

84. Stark, C. *et al.* BioGRID: a general repository for interaction datasets. *Nucleic Acids Res.* **34**, D535-539 (2006).

85. Tomkins, J. E. *et al.* PINOT: an intuitive resource for integrating protein-protein interactions. *Cell Commun. Signal* **18**, 92 (2020).

86. Licata, L. *et al.* SIGNOR 2.0, the SIGnaling Network Open Resource 2.0: 2019 update. *Nucleic Acids Res.* **48**, D504–D510 (2020).

87. Calderone, A., Iannuccelli, M., Peluso, D. & Licata, L. Using the MINT Database to Search Protein Interactions. *Curr. Protoc. Bioinforma.* **69**, e93 (2020).

88. Gioutlakis, A., Klapa, M. I. & Moschonas, N. K. PICKLE 2.0: A human protein-protein interaction meta-database employing data integration via genetic information ontology. *PloS One* **12**, e0186039 (2017).

89. Salwinski, L. *et al.* The Database of Interacting Proteins: 2004 update. *Nucleic Acids Res.* **32**, D449-451 (2004).

90. Orchard, S. *et al.* The MIntAct project—IntAct as a common curation platform for 11 molecular interaction databases. *Nucleic Acids Res.* **42**, D358–D363 (2014).

91. Cubuk, C. *et al.* Clinical likelihood ratios and balanced accuracy for 44 in silico tools against multiple large-scale functional assays of cancer susceptibility genes. *Genet. Med. Off. J. Am. Coll. Med. Genet.* **23**, 2096–2104 (2021).

92. Tischkowitz, M. *et al.* Contribution of the PALB2 c.2323C>T [p. Q775X] Founder mutation in well-defined breast and/or ovarian cancer families and unselected ovarian cancer cases of French Canadian descent. *BMC Med. Genet.* **14**, 5–5 (2013).

93. Oros, K. K. *et al.* Application of BRCA1 and BRCA2 mutation carrier prediction models in breast and/or ovarian cancer families of French Canadian descent. *Clin. Genet.* **70**, 320–329 (2006).

94. Tonin, P. N. *et al.* Founder BRCA1 and BRCA2 mutations in early-onset french Canadian breast cancer cases unselected for family history. *Int. J. Cancer* **95**, 189–193 (2001).

95. Arcand, S. L. *et al.* Germline TP53 mutational spectrum in French Canadians with breast cancer. *BMC Med. Genet.* **16**, 24–24 (2015).

96. Cote, S. *et al.* The BRCA2 c.9004G>A (E2002K) [corrected] variant is likely pathogenic and recurs in breast and/or ovarian cancer families of French Canadian descent. *Breast Cancer Res. Treat.* **131**, 333–340 (2012).

97. Osher, D. J. *et al.* Mutation analysis of RAD51D in non-BRCA1/2 ovarian and breast cancer families. *Br. J. Cancer* **106**, 1460–1463 (2012).

98. Arcand, S. L. *et al.* Germline TP53 mutations in BRCA1 and BRCA2 mutation-negative French Canadian breast cancer families. *Breast Cancer Res. Treat.* **108**, 399–408 (2008).

99. Alenezi, W. M. *et al.* Case Review: Whole-Exome Sequencing Analyses Identify Carriers of a Known Likely Pathogenic Intronic BRCA1 Variant in Ovarian Cancer Cases Clinically Negative for Pathogenic BRCA1 and BRCA2 Variants. *Genes* **13**, (2022).

100. Ahmed, A. A. *et al.* Driver mutations in TP53 are ubiquitous in high grade serous carcinoma of the ovary. *J. Pathol.* **221**, 49–56 (2010).

101. Wojnarowicz, P. M. *et al.* The genomic landscape of TP53 and p53 annotated high grade ovarian serous carcinomas from a defined founder population associated with patient outcome. *PLoS ONE* **7**, e45484–e45484 (2012).

102. PDQ Cancer Genetics Editorial Board. *Genetics of breast and gynecologic cancers (PDQ®): health professional version*. *PDQ Cancer Information Summaries* (National Cancer Institute (US)).

103. Rahman, N. Realizing the promise of cancer predisposition genes. *Nature* **505**, 302–308 (2014).

104. MacArthur, D. G. *et al.* A systematic survey of loss-of-function variants in human protein-coding genes. *Science* **335**, 823–828 (2012).

105. Halldorsson, B. V. *et al.* The sequences of 150,119 genomes in the UK Biobank. *Nature* **607**, 732–740 (2022).

106. MacArthur, D. G. & Tyler-Smith, C. Loss-of-function variants in the genomes of healthy humans. *Hum. Mol. Genet.* **19**, R125-130 (2010).

107. Sulem, P. *et al.* Identification of a large set of rare complete human knockouts. *Nat. Genet.* **47**, 448–452 (2015).

108. Lin, P.-H. *et al.* Multiple gene sequencing for risk assessment in patients with early-onset or familial breast cancer. *Oncotarget* **7**, 8310–8320 (2016).

109. Song, H. *et al.* Population-based targeted sequencing of 54 candidate genes identifies PALB2 as a susceptibility gene for high-grade serous ovarian cancer. *J. Med. Genet.* **58**, 305–313 (2021).

110. Volkova, N. V. *et al.* Mutational signatures are jointly shaped by DNA damage and repair. *Nat. Commun.* **11**, 2169 (2020).

111. Degasperi, A. *et al.* A practical framework and online tool for mutational signature analyses show inter-tissue variation and driver dependencies. *Nat. Cancer* **1**, 249–263 (2020).

112. Sondka, Z. *et al.* The COSMIC Cancer Gene Census: describing genetic dysfunction across all human cancers. *Nat. Rev. Cancer* **18**, 696–705 (2018).

113. Webster, A. L. H. *et al.* Fanconi Anemia Pathway Deficiency Drives Copy Number Variation in Squamous Cell Carcinomas. *bioRxiv* 2021.08.14.456365 (2021) doi:10.1101/2021.08.14.456365.

114. Dorsman, J. C. *et al.* Identification of the Fanconi anemia complementation group I gene, FANCI. *Cell. Oncol.* **29**, 211–218 (2007).

115. Smogorzewska, A. *et al.* Identification of the Fanconi anemia (FANC) I protein, a monoubiquitinated FANCD2 paralog required for crosslink repair. *Cell* **129**, 289–301 (2007).

116. Sims, A. E. *et al.* FANCI is a second monoubiquitinated member of the Fanconi anemia pathway. *Nat. Struct. Mol. Biol.* **14**, 564–567 (2007).

117. Mehta, P. A. & Tolar, J. Fanconi Anemia. in *GeneReviews(®) [Internet]* (eds. Pagon, R., Adam, M. & Ardinger, H.) (Seattle (WA): University of Washington, Seattle, 2002).

118. Savage, S. A. *et al.* Novel FANCI Mutations in Fanconi Anemia with VACTERL Association. *Am. J. Med. Genet. A.* **170A**, 386–391 (2015).

119. George, M. *et al.* A comprehensive molecular study identified 12 complementation groups with 56 novel FANC gene variants in Indian Fanconi anemia subjects. *Hum. Mutat.* **42**, 1648–1665 (2021).

120. Ishiai, M. *et al.* FANCI phosphorylation functions as a molecular switch to turn on the Fanconi anemia pathway. *Nat. Struct. Mol. Biol.* **15**, 1138–1146 (2008).

121. Chen, Y. *et al.* ATR-mediated phosphorylation of FANCI regulates dormant origin firing in response to replication stress. *Mol. Cell* **58**, 323–338 (2016).

122. Zhang, X., Lu, X., Akhter, S., Georgescu, M. & Legerski, R. J. FANCI is a negative regulator of Akt activation. *Cell Cycle* **15**, 1134–1143 (2016).

123. Sondalle, S. B., Longerich, S., Ogawa, L. M., Sung, P. & Baserga, S. J. Fanconi anemia protein FANCI functions in ribosome biogenesis. *PNAS* **116**, 2561–2570 (2019).

124. McReynolds, L. J. *et al.* Risk of cancer in heterozygous relatives of patients with Fanconi anemia. *Genet. Med. Off. J. Am. Coll. Med. Genet.* **24**, 245–250 (2022).

125. Berwick, M. *et al.* Genetic heterogeneity among fanconi anemia heterozygotes and risk of cancer. *Cancer Res.* **67**, 9591–9596 (2007).

126. Swift, M., Caldwell, R. J. & Chase, C. Reassessment of cancer predisposition of Fanconi anemia heterozygotes. *J. Natl. Cancer Inst.* **65**, 863–867 (1980).

127. Maxwell, K. N. *et al.* BRCA locus-specific loss of heterozygosity in germline BRCA1 and BRCA2 carriers. *Nat. Commun.* **8**, 319–319 (2017).

128. Ovarian, Fallopian Tube, and Primary Peritoneal Cancer Prevention (PDQ®): Health Professional Version. in *PDQ Cancer Information Summaries* (National Cancer Institute (US), 2002).

129. Davis, A. P. *et al.* The Comparative Toxicogenomics Database: update 2019. *Nucleic Acids Res.* **47**, D948–D954 (2019).

130. Sims, A. E. *et al.* FANCI is a second monoubiquitinated member of the Fanconi anemia pathway. *Nat. Struct. Mol. Biol.* **14**, 564–567 (2007).

131. Smogorzewska, A. *et al.* Identification of the Fanconi anemia (FANC) I protein, a monoubiquitinated FANCD2 paralog required for crosslink repair. *Cell* **129**, 289–301 (2007).

132. Ishiai, M. *et al.* FANCI phosphorylation functions as a molecular switch to turn on the Fanconi anemia pathway. *Nat. Struct. Mol. Biol.* **15**, 1138–1146 (2008).

133. Tan, W., van Twest, S., Murphy, V. J. & Deans, A. J. ATR-Mediated FANCI Phosphorylation Regulates Both Ubiquitination and Deubiquitination of FANCD2. *Front. Cell Dev. Biol.* **8**, 2 (2020).

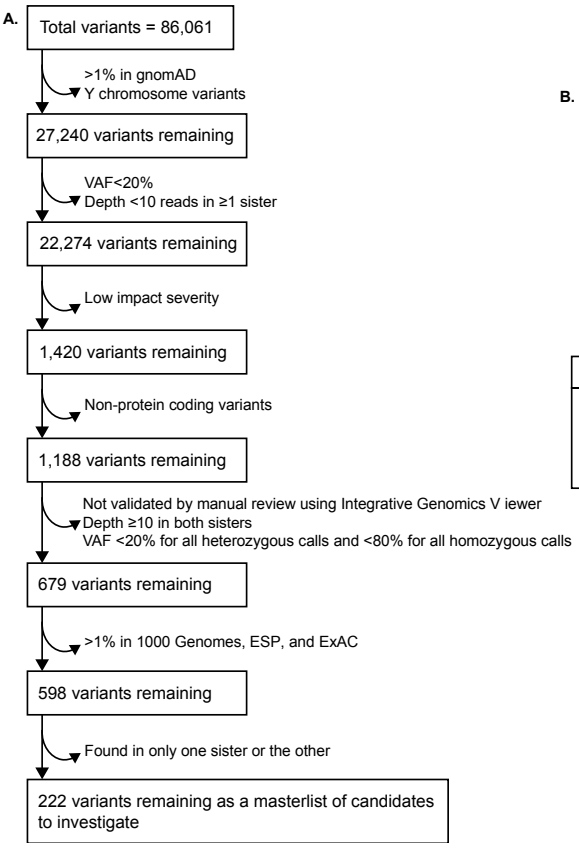
134. Yuan, F., El Hokayem, J., Zhou, W. & Zhang, Y. FANCI protein binds to DNA and interacts with FANCD2 to recognize branched structures. *J. Biol. Chem.* **284**, 24443–24452 (2009).

135. Wang, R., Wang, S., Dhar, A., Peralta, C. & Pavletich, N. P. DNA clamp function of the monoubiquitinated Fanconi anaemia ID complex. *Nature* **580**, 278–282 (2020).

136. Yang, K. *et al.* Regulation of the Fanconi anemia pathway by a SUMO-like delivery network. *Genes Dev.* **25**, 1847–1858 (2011).

137. Colnaghi, L. *et al.* Patient-derived C-terminal mutation of FANCI causes protein mislocalization and reveals putative EDGE motif function in DNA repair. *Blood* **117**, 2247–2256 (2011).

3.8 Main figures and tables



222 candidates

B. Prioritization criteria	**No. of variants**
1.a) Nonsense, frameshift, or canonical splice site variants (≤2 nucleotides from the exon)	18
b) Missense variants predicted damaging by ≥5/7 tools and conserved by ≥3/4 conservation tools	56
c) Non-canonical splice site variants (>±2 nucleotides from the exon) predicted to affect splicing by ≥3/4 tools	2
	TOTAL: 76
2. Allele frequency <1% in FC cancer-free controls (n=1,208)	66

C. Characterization of 66 candidates most likely to exert an effect on the encoded protein			
Hallmark of cancer	Somatic variants in TCGA HGSC	Found in FC OC	ClinGen cancer associated disease
50	50	3	0

Figure 3.1. Criteria used for filtering and prioritizing variants identified across the genetic landscape of sisters from family F1528. (A) Filtering strategy to identify genetic landscape variants; (B) prioritization of variants to identify those most likely to exert an effect on the encoded protein; and (C) characterization of variants most likely to affect protein function using various characteristics of cancer associated genes.

gnomAD: Genome Aggregation Database; VAF: variant allele frequency; ESP: Exome Sequencing Project; ExAC: Exome Aggregation Consortium; FC: French Canadian; TCGA: The Cancer Genome Atlas; HGSC: high grade serous ovarian carcinoma; OC: ovarian cancer

Table 3.1. Genetic landscape variants identified in OC cases of FC ancestry negative for pathogenic variants in *BRCA1*, *BRCA2*, *BRIP1*, *RAD51C*, and *RAD51D*, including *FANCI* c.1813C>T carrier status.

Gene	Coding change	Protein change	F1528-PT0056[1]	F1528-PT0057[1]	F1085-PT0134[2]	F1085-PT0135[2]	F1601-PT0138	F845-PT0196	F1490-PT0047	F1620-PT0100	F1506-PT0136	F1543-PT0137	F1288-PT0158	F1617-PT0090	F694-PT0128	F439-PT0184	F1650-PT0142
FANCI	c.1813C>T	p.L605F	x	x					x	x							
PTPN22	c.993-1G>A	NA	x	x	x	x											
GPD1	c.431T>C	p.M144T	x	x			x										
SEC14L4	c.364C>T	p.R122W	x	x				x									

x: heterozygous; NA: not applicable
[1]From the same family F1528
[2]From the same family F1085

Table 3.2. Other variants identified in genes where genetic landscape variants were identified in OC cases of FC ancestry negative for pathogenic variants in *BRCA1*, *BRCA2*, *BRIP1*, *RAD51C*, and *RAD51D*.

Gene	Coding change	Protein change	F1528-PT0056[1]	F1528-PT0057[1]	F1085-PT0134[2]	F1085-PT0135[2]	F1601-PT0138	F845-PT0196	F1490-PT0047	F1620-PT0100	F1506-PT0136	F1543-PT0137	F1288-PT0158	F1617-PT0090	F694-PT0128	F439-PT0184	F1650-PT0142
PIWIL3	c.2023T>G	p.C675R	x	x													
PIWIL3	c.1932+1G>A	NA									x						
CACNA1S	c.4340G>A	p.R1447Q	x	x													
CACNA1S	c.773G>A	p.G258D			x												
MYO7A	c.5866G>A	p.V1956I	x	x													
MYO7A	c.1078G>T	p.E360*			x												
SCN10A	c.3776G>A	p.R1259Q	x	x													
SCN10A	c.2972C>T	p.P991L											x				
PCDH15	c.3127C>T	p.P1043S	x	x													
PCDH15	c.2581G>A	p.V861M						x									
TEX2	c.73G>T	p.V25L	x	x													
TEX2	c.3040G>A	p.E1014K						x									
DNAH3	c.10382C>G	p.P3461R	x	x													
DNAH3	c.5368A>T	p.I1790F								x							
DNAH1	c.1941_1944del	p.N648Afs*36	x	x													
DNAH1	c.2717A>G	p.D906G										x					
DNAH1	c.10216G>A	p.V3406I									x						
IQCA1	c.29G>A	p.W10*	x	x													
IQCA1	c.979G>C	p.A327P										x					

x: heterozygous; NA: not applicable

[1]From the same family F1528

[2]From the same family F1085

Table 3.3. Genetic landscape variants identified in Australian HGSC cases (n=516).

Gene	Coding change	Protein change	No. of carriers (%)
CACNA1S	c.4340G>A	p.R1447Q	1 (0.2)
NBAS	c.3217C>T	p.R1073C	6 (1.2)
ANKAR	c.3815G>A	p.R1272H	2 (0.4)
PARD3B	c.365T>C	p.I122T	1 (0.2)
TNS1	c.1333G>C	p.G445R	4 (0.8)
IQCA1	c.29G>A	p.W10*	3 (0.6)
CXCL6	c.239dup	p.V81Gfs*44	9 (1.7)
CEP120	c.2134C>T	p.L712F	8 (1.6)
KCNU1	c.2731G>A	p.A911T	1 (0.2)
NUP188	c.3974G>A	p.R1325H	4 (0.8)
CREM	c.677C>T	p.S226L	5 (1)
PCDH15	c.3127C>T	p.P1043S	1 (0.2)
NPFFR1	c.8-2A>G	NA	6 (1.2)
MYO7A	c.5866G>A	p.V1956I	3 (0.6)
PWP1	c.1402G>A	p.E468K	7 (1.4)
PAQR5	c.20C>G	p.P7R	3 (0.6)
DNAH3	c.10382C>G	p.P3461R	1 (0.2)
PLIN4	c.3260_3263dup	p.F1089Pfs*32	3 (0.6)
CYP2A6	c.289G>A	p.E97K	1 (0.2)
ALDH16A1	c.1376A>T	p.D459V	5 (1)
MYH9	c.4396C>T	p.R1466W	2 (0.4)

Table 3.4. FANCI interactome candidate variants identified in FC OC cases negative for pathogenic variants in *BRCA1*, *BRCA2*, *BRIP1*, *RAD51C*, *RAD51D*, and *FANCI*.

Gene	Coding change	Protein change	F1601-PT0138	F1506-PT0136	F845-PT0196	F1085-PT0134[1]	F1085-PT0135[1]	F1617-PT0090	F694-PT0128	F1288-PT0158	F1543-PT0137	F439-PT0184	F1650-PT0142
EZH2	c.1786G>A	p.Ala596Thr	x										
ANKRD55	c.1126T>C	p.Ser376Pro		x									
MOV10	c.2501G>A	p.Arg834Gln		x									
LRRK2	c.356T>C	p.Leu119Pro			x								

[1]From the same family F1085

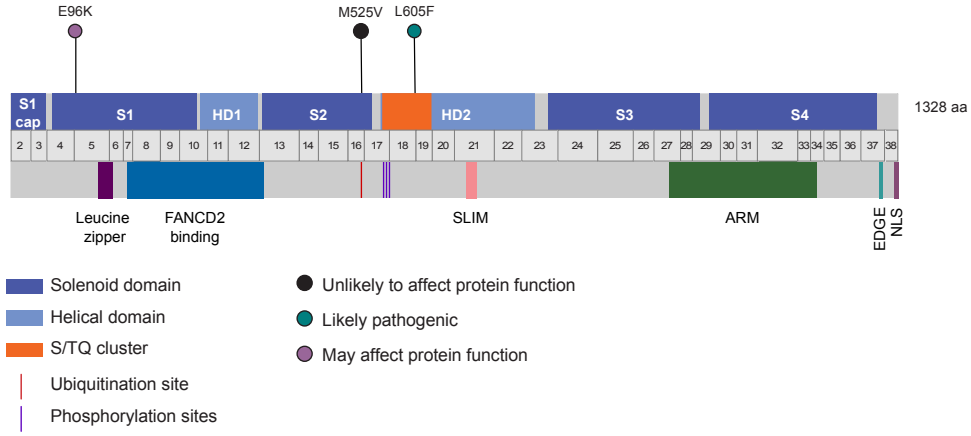

E96K
M525V
L605F
S1 cap
S1
HD1
S2
HD2
S3
S4
1328 aa
2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38
Leucine zipper
FANCD2 binding
SLIM
ARM
EDGE
NLS
Solenoid domain
Helical domain
S/TQ cluster
Ubiquitination site
Phosphorylation sites
Unlikely to affect protein function
Likely pathogenic
May affect protein function

Figure 3.2. Schema showing location of variants in *FANCI* gene and protein identified from ClinVar using the established criteria for c.1813C>T; p.L605F.
FANCI domains were adapted from pfam (https://pfam.xfam.org). *FANCI* exon locations were adapted from University of California Santa Cruz Genome Browser (https://genome.ucsc.edu). Solenoid domain: antiparallel pairs of α-helices that form α-α superhelix segments; Helical domain: α-helices; Ubiquitination site: site of monoubiquitination by the FA core complex to allow downstream FA pathway function, located at K523[130,131]; S/TQ cluster: location of conserved phosphorylation sites[132]. Phosphorylation sites (556, 559, and 565aa): sites of phosphorylation that stabilize the association of FANCI with DNA and FANCD2[133]. Leucine zipper (130-151aa): may be related to protein-protein interactions, DNA binding, or RNA binding, but the leucine zipper found at the N-terminus of FANCI has been shown not to bind to DNA[134]. Ubiquitin binding (175-377aa): this region binds to the ubiquitin on FANCD2[135]. SUMO-like domain-interacting motif (SLIM; 682-696aa): binds to the SUMO-like domain 2 (SLD2) of UAF1 promoting FANCD2 deubiqutination which is required for FA pathway function[136]. Armadillo repeat (ARM; 985-1207aa): forms a super helix of helices, which can also be found in FANCD2[131]. EDGE motif (1300-1303aa): this motif consists of Glutamic acid (E) - Aspartic acid (D) - Glycine (G) - Glutamic acid (E) and is required for DNA crosslink repair function[130,131,137]. Nuclear localization site (NLS; 1323-1238aa): required for localization to the nucleus where subsequent function in the FA pathway can occur[137].

Table 3.5. Somatic variants in the nine most frequently altered genes in HGSC identified in cases harbouring *FANCI* c.1813C>T (n=13).

Sample ID	*TP53*	*BRCA1*	*CSMD3*	*NF1*	*CDK12*	*FAT3*	*GABRA6*	*BRCA2*	*RB1*
PT0001									
PT0002									
PT0003									
PT0004									
PT0006									
PT0005									
PT0007									
TCGA-04-1336									
TCGA-24-1603									
TCGA-25-2393									
TCGA-29-2431									
TCGA-61-1903									
TCGA-61-2009									
Total (%)	**11 (85%)**	**0**	**0**	**0**	**1 (8%)**	**3 (23%)**	**0**	**3 (23%)**	**0**

Missense	
Splice	
Frameshift	
In-frame	

Table 3.6. Carrier frequency of *FANCI* c.1813C>T in TCGA PanCancer cases (n=10,389).

Cancer Type (TCGA acronym)	Total no. cases	No. of *FANCI* c.1813C>T carriers	Carrier frequency of *FANCI* c.1813C>T (%)
Adrenocortical carcinoma (ACC)	92	3	3.3
Kidney chromophobe (KICH)	66	2	3
Lung squamous cell carcinoma (LUSC)	499	14	2.8
Skin cutaneous melanoma (SKCM)	470	13	2.8
Kidney renal clear cell carcinoma (KIRC)	387	10	2.6
Colon adenocarcinoma (COAD)	419	10	2.4
Cholangiocarcinoma (CHOL)	45	1	2.2
Esophageal carcinoma (ESCA)	184	4	2.2
Brain lower grade glioma (LGG)	515	11	2.1
Liver hepatocellular carcinoma (LIHC)	375	8	2.1
Head & neck squamous cell carcinoma (HNSC)	526	9	1.7
Uterine corpus endometrial carcinoma (UCEC)	543	9	1.7
Breast invasive carcinoma (BRCA)[1]	1076	17	1.6
Cervical squamous cell carcinoma (CESC)	305	5	1.6
Sarcoma (SARC)	255	4	1.6
Stomach adenocarcinoma (STAD)	443	7	1.6
Ovarian serous cystadenocarcinoma (OV)	412	6	1.5
Testicular germ cell tumours (TGCT)	134	2	1.5
Lung adenocarcinoma (LUAD)	518	7	1.4
Rectum adenocarcinoma (READ)	145	2	1.4
Uveal melanoma (UVM)	80	1	1.3
Bladder urothelial carcinoma (BLCA)	412	5	1.2
Mesothelioma (MESO)	82	1	1.2
Thyroid carcinoma (THCA)	499	6	1.2
Pancreatic adenocarcinoma (PAAD)	185	2	1.1
Pheochromocytoma & paraganglioma (PCPG)	179	2	1.1
Glioblastoma multiforme (GBM)	393	4	1
Prostate adenocarcinoma (PRAD)	498	5	1
Acute myeloid leukemia (LAML)	142	1	0.7
Diffuse large B-cell carcinoma (DLBC)	41	0	0
Kidney renal papillary cell carcinoma (KIRP)	289	0	0
Thymoma (THYM)	123	0	0
Uterine carcinosarcoma (UCS)	57	0	0
Total	**10389**	**171**	**1.6**
gnomAD non-cancer overall[2]	**134164**	**1787**	**1.3**

[1] 1 homozygous carrier

[2] 17 homozygous carriers

Table 3.7. Frequency of somatic *FANCI* variants identified in TCGA PanCancer tumours.

Cancer type (TCGA acronym)	Total no. of cases	No. of tumours harbouring *FANCI* variants	Frequency of tumours harbouring *FANCI* variants (%)
Uterine corpus endometrial carcinoma (UCEC)	517	43	8.32
Skin cutaneous melanoma (SKCM)	438	20	4.57
Bladder urothelial carcinoma (BLCA)	410	13	3.17
Colon adenocarcinoma (COAD)/ Rectum adenocarcinoma (READ)	534	14	2.62
Stomach adenocarcinoma (STAD)	436	9	2.06
Cervical squamous cell carcinoma (CESC)	291	6	2.06
Uterine carcinosarcoma (UCS)	57	1	1.75
Lung squamous cell carcinoma (LUSC)	484	8	1.65
Lung adenocarcinoma (LUAD)	566	9	1.59
Head & neck squamous cell carcinoma (HNSC)	515	8	1.55
Acute myeloid leukemia (LAML)	200	3	1.50
Mesothelioma (MESO)	86	1	1.16
Adrenocortical carcinoma (ACC)	91	1	1.10
Esophageal carcinoma (ESCA)	182	2	1.10
Glioblastoma multiforme (GBM)	391	4	1.02
Breast invasive carcinoma (BRCA)	1066	10	0.94
Thymoma (THYM)	123	1	0.81
Sarcoma (SARC)	255	2	0.78
Ovarian serous cystadenocarcinoma (OV)	523	4	0.76
Kidney renal clear cell carcinoma (KIRC)	402	3	0.75
Testicular germ cell tumours (TGCT)	149	1	0.67
Pheochromocytoma & paraganglioma (PCPG)	178	1	0.56
Liver hepatocellular carcinoma (LIHC)	366	2	0.55
Prostate adenocarcinoma (PRAD)	494	2	0.40
Brain lower grade glioma (LGG)	514	2	0.39
Kidney renal papillary cell carcinoma (KIRP)	276	1	0.36
Thyroid carcinoma (THCA)	489	1	0.20
Cholangiocarcinoma (CHOL)	36	0	0
Diffuse large B-cell carcinoma (DLBC)	41	0	0
Kidney chromophobe (KICH)	65	0	0
Pancreatic adenocarcinoma (PAAD)	179	0	0
Uveal melanoma (UVM)	80	0	0
Total	**10,434**	**172**	**1.65**

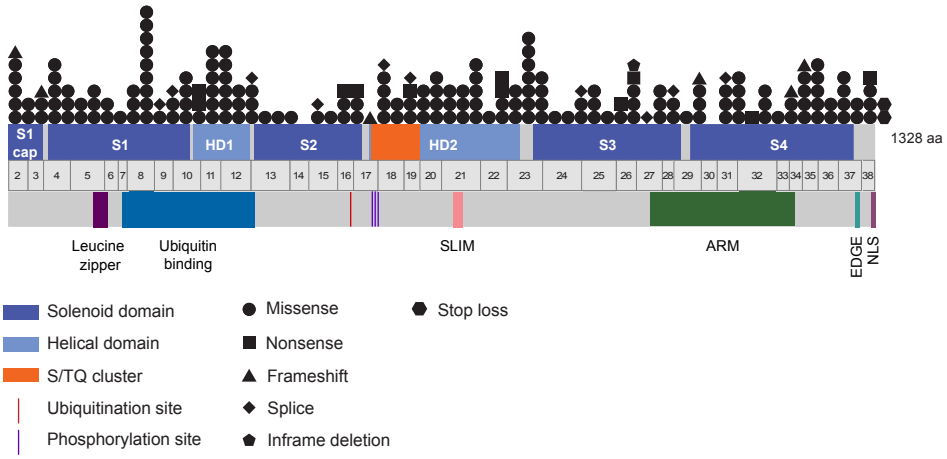
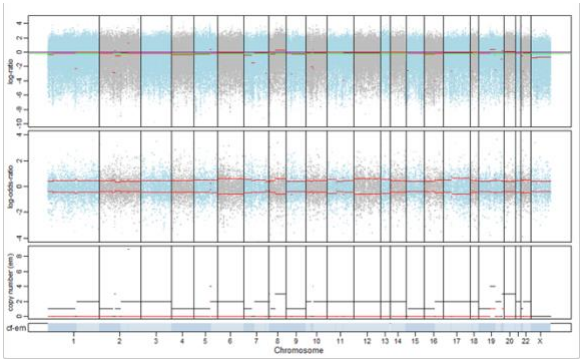


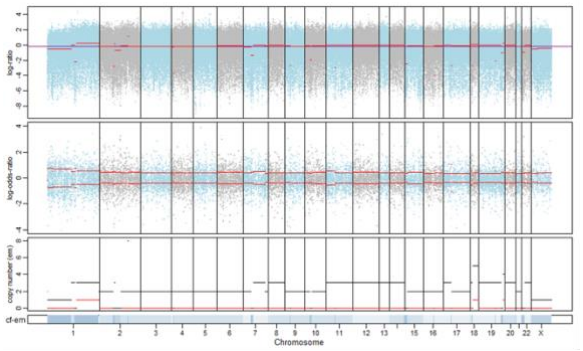
Figure 3.3. Schema showing the location of all somatic variants in FANCI gene and protein identified in tumours from TCGA PanCancer Atlas.

3.9 Supplementary figures

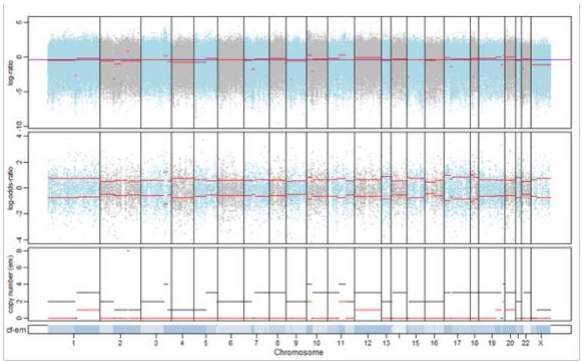
PT0001



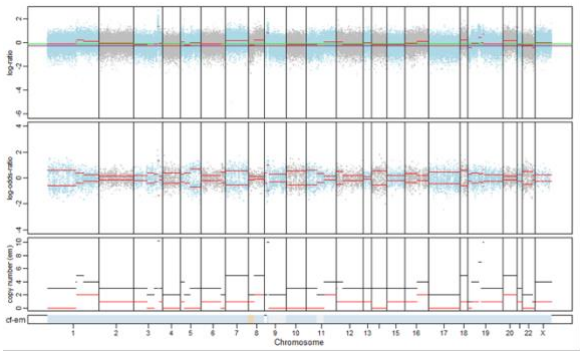
PT0006



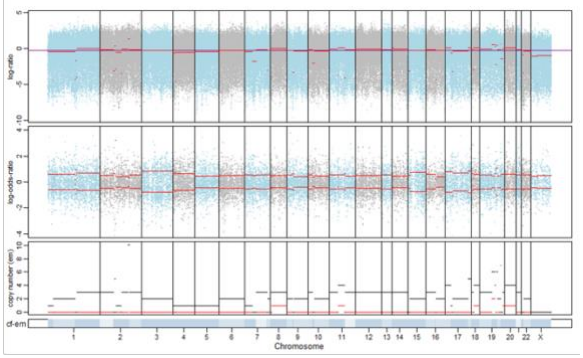
PT0002



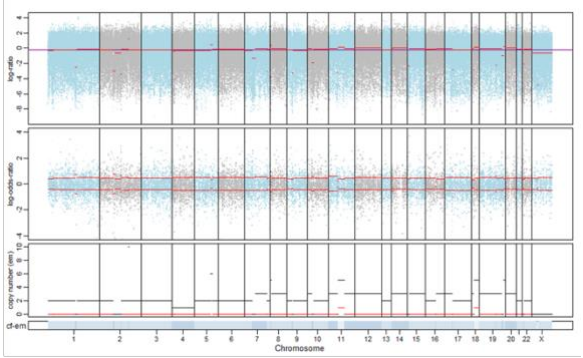
PT0005



PT0003



PT0007



PT0004

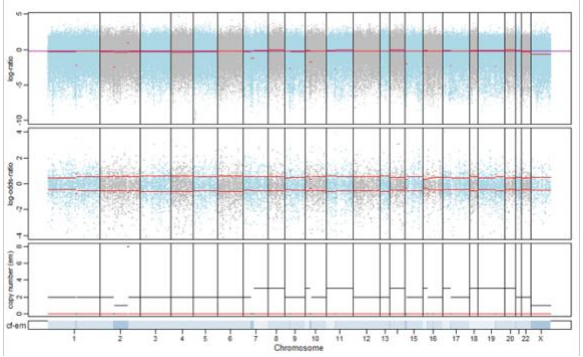


Fig 3.S1. Copy number alterations per chromosome across the genome of tumours from French Canadian cases harbouring germline *FANCI* c.1813C>T. The top panel shows total copy number log-ratio, the middle panel shows allele-specific copy number log-ratio, and the bottom panel shows the corresponding integer copy number calls (red is the minor allele).

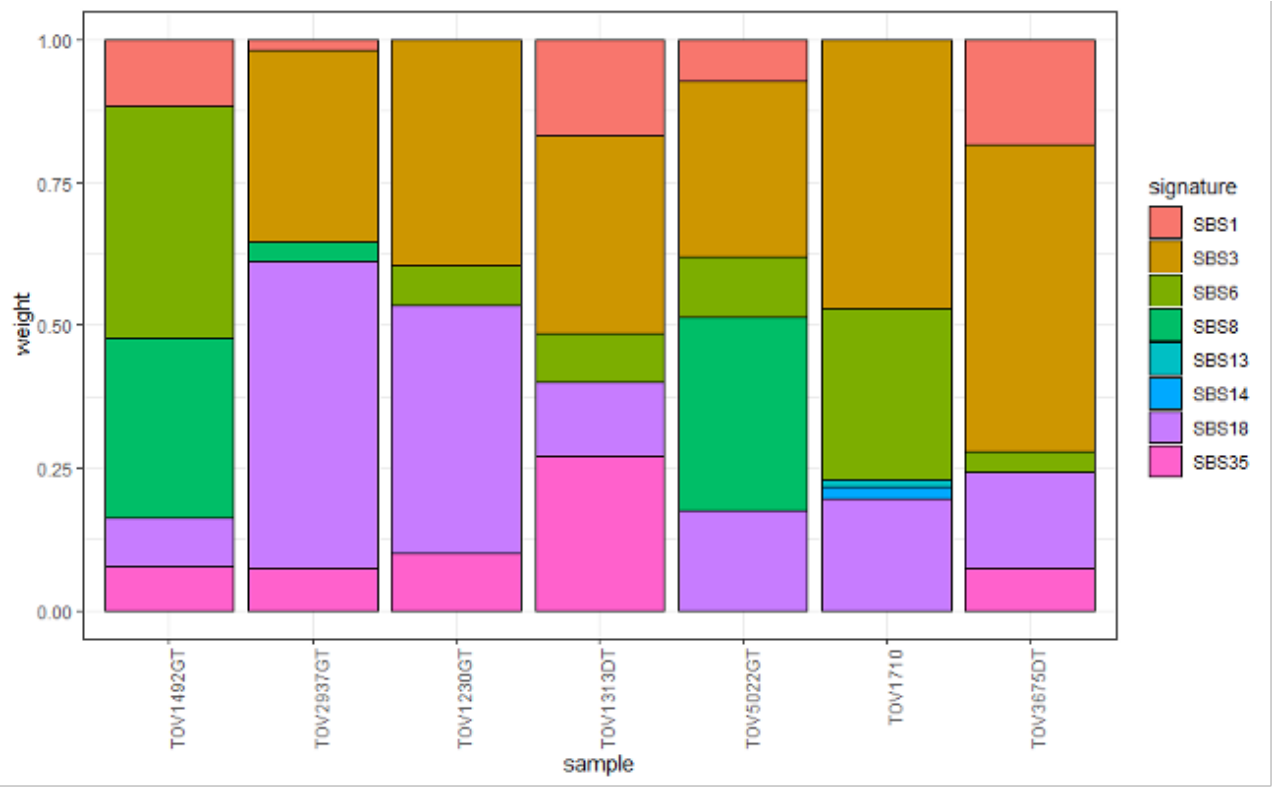


Fig 3.S2. Proportional representation of defined somatic mutational signatures observed in each ovarian tumour from French Canadian cases harbouring germline *FANCI* c.1813C>T.

SBS1: Ageing

SBS3: Homologous recombination deficiency

SBS6: Mismatch repair deficiency

SBS8: Unknown (putative homologous recombination deficiency)

SBS13: Activation of APOBEC family

SBS14: Polymerase epsilon mutation and mismatch repair deficiency

SBS18: Damage due to reactive oxygen species

SBS35: Platinum chemotherapy treatment

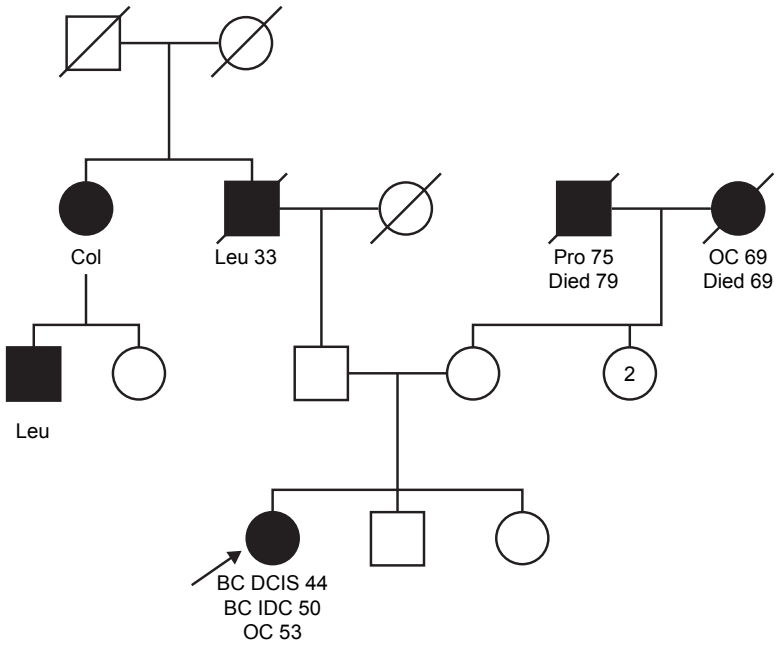


Fig 3.S3. Pedigree of a Ukrainian breast and ovarian cancer family negative for *BRCA1* and *BRCA2* pathogenic variants where *FANCI* c.286G>A; p.E96K was identified by WES in the proband, indicated by the arrow. Age at diagnosis and death are shown where known. Cancer type (OC: ovarian, BC: breast, Pro: prostate, Col: colon, and Leu: leukemia) and age of diagnosis are shown.

DCIS: ductal carcinoma in situ

IDC: infiltrating ductal carcinoma

4.0 Discussion

4.1 Identification of *FANCI* c.1813C>T in familial OC

The identification of *FANCI* c.1813C>T; p.L605F in a family with multiple cases of OC and no identifiable *BRCA1* or *BRCA2* pathogenic variant prompted the investigation of this gene as a new candidate OC predisposing gene. Based on the established method of estimating the frequency of a candidate variant in different subgroups to find evidence for a risk allele, I showed that *FANCI* c.1813C>T was more frequent in FC OC families negative for pathogenic variants in *BRCA1* and *BRCA2* (3/23, 13%) compared to sporadic FC OC cases (7/439, 1.6%) and cancer-free controls of FC ancestry (74/2950, 2.5%), which supports a role for this variant in OC risk. I was unable to estimate the penetrance of this variant as it would take many more carriers to determine. Given the relatively high frequency of *FANCI* c.1813C>T in my cancer free controls and that this variant is predicted to encode a missense variant, it is plausible that it exhibits a low to moderate risk for OC, that is, there is incomplete penetrance of this variant.

Knowledge of family history of cancer has been integral in the identification of CPGs, as evidenced in the identification of *BRCA1*[16] and *BRCA2*[18], where carriers of pathogenic variants were more readily identified in families with multiple reported cases of BC or OC. Similarly, a higher carrier frequency of pathogenic variants *RAD51C*[188] or *RAD51D*[37] has been found in families with at least two cases of OC, compared to controls. As mentioned in Chapter 1.0, the initial report of *CHEK2* c.1100del not only suggested a higher than expected carrier frequency in population controls (1.4%) but also showed a significantly higher number of carriers in cases who had a family history

of BC or BC and OC, but not in those unselected for family history of cancer[133]. Since then, *CHEK2* pathogenic or likely pathogenic variants have been shown to exhibit an absolute risk of 20-40% for BC, which is moderate compared to the >60% absolute risk for BC for carriers of pathogenic *BRCA1* or *BRCA2* variants[19]. It is therefore possible that *FANCI* c.1813C>T, a likely pathogenic variant, exhibits a moderate risk for OC.

The possibility that *FANCI* c.1813C>T could affect risk in non-FC populations was confirmed as I identified carriers of this variant in Australian OC cases[124]. My identification of other rare *FANCI* variants predicted to affect protein function by in silico tools in this Australian OC study group suggest the possibility that these variants could affect OC risk[124]. Particularly, *FANCI* c.286G>A; p.E96K would be of interest to investigate using functional biology experiments as this variant was identified in a sporadic Australian OC case[124] and an OC family of Ukrainian ancestry (Chapter 3.0).

A major challenge in the identification of *FANCI* c.1813C>T as a new OC predisposing variant included the inability to fully investigate segregation of the variant with disease, due to a lack of available DNA from family members of carrier index cases[124]. The identification of a candidate variant in affected individuals but not unaffected individuals from the same family can suggest a role in risk for this variant, assuming that the trait being investigated is associated with an autosomal dominant inheritance pattern, such as for *BRCA1* and *BRCA2*[189]. Therefore, the expectation would be that I would identify *FANCI* c.1813C>T carriers in OC, and possibly BC, cases within a family but not in unaffected relatives. In the OC cases with available family history of disease (n=4 families; F1528 [discovery family], F1023, F1490, and F1620) segregation analysis was only able to be partially performed in F1528, where both

sisters were carriers of *FANCI* c.1813C>T[124]. The only other family member available from F1490 was the mother of the index case, who was diagnosed with colon cancer at age 63 years and was found to be a carrier of *FANCI* c.1813C>T. There was no available DNA from family members in families F1023 and F1620. Although segregation could not be determined for *FANCI* c.1813C>T in my study of FC OC families, this has been a challenge in other studies as well, such as those involving *RAD51C* and *RAD51D,* which have required large studies and statistical modeling using complex segregation analyses to estimate associated risks[190]. It is however important to note that the penetrance for *BRCA1*, *BRCA2*, and other OC predisposing genes is not 100%, and few CPGs approach complete penetrance[191].

As the possibility of an association between FANCI protein or mRNA expression with overall survival in HGSC was observed in Chapter 2.0, we investigated the protein expression of FANCI in a larger study group of HGSC cases (n=1159) from the Canadian Ovarian Experimental Unified Resource (COEUR) (https://www.tfri.ca/coeur). There was no association between FANCI protein expression and overall survival in this study group (P.N. Tonin unpublished data). As DNA samples were available from this study group, I was able to investigate the carrier frequency of *FANCI* c.1813C>T in Canadian HGSC cases. I identified 18 carriers in 822 available HGSC cases (2.2%; P.N. Tonin unpublished data), which is comparable to the 7 carriers in 341 HGSC cases of FC ancestry (2.1%; p=1)[124], though the carrier frequency of *FANCI* c.1813C>T across the Canadian population is unknown.

4.2 Role of *FANCI* in other cancers

I identified carriers of *FANCI* c.1813C>T in individuals diagnosed with 29 different types of cancer in TCGA Pan Cancer data at a higher frequency than in population-based controls (Chapter 3.0). This provides further support for *FANCI* as a CPG not only for OC but also for cancer in general. Further investigation is required to determine if *FANCI* increases risk to other cancer types or may alter the progression of cancer. It is interesting to note the ear, nose, and throat cancer diagnosed in a sister from family F1528[124] as FA genes are associated with an increased risk of head and neck carcinomas[117,192], though the genotype of this individual was unable to be determined due to unavailability of genetic material. Other CPGs have been shown to exhibit risk to various cancers, such as *BRCA1* predisposing to OC, BC (female and male), pancreatic, and prostate[19]. The absolute risk for these cancers differs in carriers of pathogenic variants in *BRCA1*, the highest for female BC (>60%) and lowest for male BC (0.2-1.2% by age 70 years)[19]. There is therefore precedence for CPGs to predispose to different cancer types with varying penetrance.

As discussed in our report of *FANCI* as a candidate OC predisposing gene[124] heterozygous germline *FANCI* variants were reported in the literature in other cancer types including BC, prostate cancer, sarcoma, malignant pleural mesothelioma, acute myeloid leukemia, head and neck carcinoma, and colorectal cancer. There have been reports of heterozygous germline *FANCI* variants in individuals diagnosed with OC[193], BC[194], advanced melanoma[195], acute myeloid leukemia[196], acute lymphoblastic leukemia[197], clear cell papillary renal cell carcinoma[198], pediatric brain tumor[199], pancreatic cancer[200], Kaposi sarcoma[201], prostate cancer[202], and synchronous

sialoblastoma and hepatoblastoma[203]. This, along with the identification of *FANCI* c.1813C>T; p.L605F variant carriers across different cancer types from TCGA PanCancer Atlas, suggests that risk to cancers other than OC in carriers of heterozygous pathogenic or likely pathogenic *FANCI* variants should be further investigated.

FANCI has also been identified as a hub gene, that is a gene which is part of gene network and has many connections with other genes, in multiple myeloma[204], hepatitis B virus related hepatocellular carcinoma[205], thymoma[206], colorectal cancer[207–209], nasopharyngeal carcinoma[210], prostate cancer[211], lung adenocarcinoma[212], non-small cell lung cancer[213], cutaneous melanoma[214], hepatoblastoma[215], Ewing sarcoma[216], and retinoblastoma[217]. *FANCI* has also been implicated as a possible oncogene in lung adenocarcinoma[218] and non-small cell lung cancer[219] and as a prognostic biomarker for cervical cancer[220]. An integrative analysis of data on methylation, gene expression, and somatic mutations for 13 different types of cancers in TGCA showed that *FANCI* is predictive of both cancer diagnosis and prognosis[221]. Hub genes are often suspected of playing integral roles in a biological system due to the interconnectedness through the gene network. These gene networks can be constructed using microarray data measuring mRNA expression, which provides information about how genes are related to one another[222]. The information used to build these gene networks can include: known interactions from curated databases or experiments; predicted interactions using gene neighbourhoods, gene fusions, or gene co-occurrence; and other possible interactions from text mining, co-expression, and protein homology (Figure 4.1). *FANCI* has also been implicated as a possible oncogene

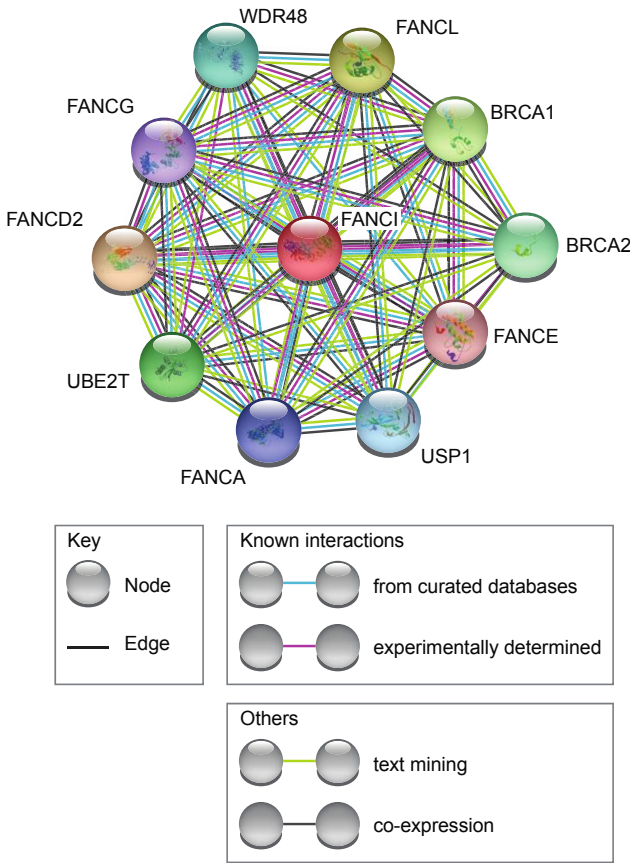


Figure 4.1. Representative gene network for *FANCI* constructed from known and other interactions as indicated by the edges. Nodes represent genes and edges represent protein-protein interactions.

Adapted from STRING[223].

in lung adenocarcinoma[218] and non-small cell lung cancer[219] and as a prognostic biomarker for cervical cancer[220]. It is therefore possible that the biological role of *FANCI* across cancer types may differ, similar to that of *TP53* which can affect tumourigenesis with either gain-of-function or loss-of-function pathogenic variants[224]. It is also possible that the cell lines used in the studies of lung adenocarcinoma and non-small cell lung cancer may be reliant on the FA-HR pathway and therefore loss of *FANCI* leads to a phenotype akin to that exhibited by an oncogene.

4.3 Biological perspectives

In collaboration with Dr. Jean-Yves Masson, functional assays were performed to determine the biological function of FANCI p.L605F (Chapter 2.0)[124]. These assays have been established for the investigation of proteins in the FA pathway and based on knowledge of the function of FANCI within this pathway[120,143]. Western blot analysis of FANCI p.L605F complemented cells showed that this protein was expressed at decreased levels compared to WT complemented cells when treated with MMC, a DNA damaging agent. This decrease was confirmed in a time-course experiment, where FANCI p.L605F protein levels decreased over time, thereby suggesting instability of the protein, which is exacerbated in the presence of DNA damaging agents. As the function of FANCI is dependent on FANCD2 and ubiquitination of both proteins, investigation of the interaction between these proteins and ubiquitination can be used to assess the functionality of the FA pathway[225]. Western blot analysis showed decreased ubiquitination of both FANCI and FANCD2 in cells complemented with p.L605F after MMC treatment. Immunoprecipitation of FANCI p.L605F and FANCD2 showed that the two proteins were still able to interact but the amount of ubiquitinated FANCD2 was

severely diminished. Ubiquitination of FANCD2 is also required to form MMC-induced foci, that is localization of FANCD2 to sites of DNA damage, and we found that FANCD2 foci were significantly reduced in FANCI p.L605F expressing cells. From these functional assays, I have gleaned that *FANCI* c.1813C>T; p.L605F encodes a protein that is unlikely to function normally, though it may exhibit hypomorphic function[124]. The results from our in cellulo and in vitro models have not yet been confirmed in vivo, though loss of heterozygosity analyses from Chapter 3.0 suggest loss of the wild type allele in carriers of FANCI p.L605F may be an early event in tumourigenesis. There is thus far no evidence to suggest *FANCI* does not play a role in tumourigenesis.

As mentioned in the introduction of this thesis, platinum chemotherapy is the first-line treatment option for patients with HGSC, and only recently has a targeted therapy (PARPi) become available for patients[226]. For this reason, FANCI p.L605F expressing cells were treated with cisplatin (a platinum chemotherapy) or olaparib (a PARPi) to determine if cells were sensitive to these therapies. FANCI p.L605F expressing cells were sensitive to cisplatin but not to olaparib[124], which is not unexpected as it has been reported that loss of proteins upstream of the HR pathway, mainly members of the FA pathway, do not exhibit synthetic lethality with PARPi[175]. Results by my collaborators from the in cellulo analysis of FANCI and FANCI p.L605F presented in Chapter 2.0 are compatible with these findings, that is, this variant does not exhibit a strong response to the PARPi olaparib[124]. This is consistent with the a previous observation that genes upstream in the FA pathway are not sensitive to PARPi[175]. Another FA gene, *FANCJ* (*BRIP1*), that is a moderate penetrance OC predisposing gene, was also not sensitive to a PARPi in this study. In collaboration with Dr. Jean-Yves Masson, we have shown

that *BRIP1* deficient cells are sensitive to cisplatin but not to the PARPi olaparib (P.N. Tonin and J.-Y. Masson unpublished data), similar to what was seen for *FANCI* deficient cells in Chapter 2.0. Currently, no targeted therapies exist for *FANCI*, though patients with germline c.1813C>T; p.L605F may benefit from treatment with cisplatin, a first line chemotherapy for OC[124]. Pathway inhibition experiments have shown that treatment of cells with ouabain or pristimerin lead to a decrease in the mRNA expression of *FANCI*, and *FANCD2*, and sensitized cells to MMC or cisplatin, respectively[227,228]. These results suggest that the FA pathway may be inhibited to serve as a chemosensitizer to ICL-inducing chemotherapies. Avenues for future research will include the investigation of drug combinations that may benefit OC patients with variants in *FANCI*, such as c.1813C>T; p.L605F, or those with intact FA pathways to sensitize them to chemotherapeutics. FA complementation group I patient fibroblast cells have been used to generate induced pluripotent stem cells that were then able to be corrected for *FANCI* variation using Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)/Cas9 gene editing, which may be a potential future avenue for treatment of FA patients[229].

The biological role of the FA pathway and of FANCI have been further elucidated in recent years including the formation of the FA core complex which exists as a dimeric catalytic molecule for the monoubiquitination of the ID2 complex[230]. The protein stability of FANCI and FANCD2 are highly interdependent; in the absence of one of the heterodimeric binding partners, approximately 50% of the other partner is lost[231]. The structure of the ID2 complex in the mouse (*Mus musculus*) was reported in 2011[232], and recently cryo-electron microscopy has revealed the structures of the FA core and ID2

complexes[155,233]. FANCI has been shown to function as a switch between repair and apoptosis in response to DNA crosslinks dependent on the ubiquitination state of FANCI[234]. This is the first report of such a direct mechanism enabling cells to choose between repair and apoptosis. Future investigation of the biochemical role of FANCI within and outside the FA DNA repair pathway will aid in determining possible mechanisms of action in tumourigenesis. It would be important to know how FANCI c.1813C>T; p.L605F could affect the function of FANCI within these pathways.

Further studies aimed at elucidating the role of *FANCI* in cancer initiation and/or progression may be conducted using animal models. For example, a recent mouse (*Mus musculus*) model for HGSC developed by Dr. Yojiro Yamanaka's group at McGill University uses a CRISPR-Cas9 system and electroporation of the oviduct, equivalent to the fallopian tube in humans, to KO genes specifically in the oviductal cells[235]. Tumour suppressor genes (*PTEN*, *BRCA1*, and *TP53*) with or without *LKB1* were targeted for KO and led to the successful generation of immunocompetent mice with HGSC[235]. *FANCI* KO or *FANCI* c.1813C>T; p.L605F could be introduced into this system to determine if time to tumour initiation differs compared to mice without an altered *FANCI* gene. A *FANCI* KO mouse was established by Dr. Jean-Yves Masson's group at Laval University and though it was not embryonic lethal the expected Mendelian ratio from crossing heterozygous mice yielded only 4% KO mice instead of 25%[236]. Low Mendelian ratios suggest that the majority of KO mice do not survive gestation, which makes this model particularly difficult to study due to the low yield of offspring. Of the 22 FA genes, there are mouse models for 21 (excluding *FANCT* [*UBE2T*]): eight have normal Mendelian ratios (*FANCA*, -*B*, -*C*, -*E*, -*F*, -*G*, -*J*, and -*Q*),

six have sub-Mendelian ratios (*FANCD2*, -*I*, -*M*, -*P*, -*V*, and -*W*) and seven are embryonic lethal or not viable (*FANCD1* [*BRCA2*], -*L*, -*N* [*PALB2*], -*O* [*RAD51C*], -*R* [*RAD51*], -*S* [*BRCA1*], and -*U* [*XRCC2*]), though this can differ based on the mouse strain[237]. *FANCI* KO mice exhibited some phenotypic characteristics shown by FA patients including hematologic deficiency and limb abnormalities in some animals. Cancer incidence was not reported in this study and could be an avenue for future investigation. A different animal model for investigating the role of *FANCI* is zebrafish (*Danio rerio*) which has largely been used to study developmental alterations observed in FA patients. Zebrafish *fanci* is the most conserved FA protein sequence with 56% overall amino acid identity[238], and the motif surrounding the ubiquitination site is identical to the human residues (residues 519-523). The region surrounding human *FANCI* is highly conserved with the downstream genes *POLG* and *RHCG* also being neighbours in the same order in zebrafish on linkage group 25 (Dre25), suggesting conserved synteny that has remained intact for over 450 million years[238]. KO of *fanci* and double KO of *fanci* and *fancd2* (ID2 complex), along with all other homozygous or multi-gene KOs in the FA pathway were viable in zebrafish, suggesting that there is not a requirement for these genes early in development[239]. There were no gross developmental abnormalities observed in *fanci* KO zebrafish, and no female KOs were born, similar to other FA KOs. This suggests a possible female-to-male sex reversal phenotype. Further investigation of KO zebrafish to adult stages (4-6 months) revealed that *fanci* KO zebrafish are male-biased (<5% females), though these male zebrafish are fertile[240]. Upon treatment with the DNA damaging agent 1,2:3,4-Diepoxybutane (DEB), *fanci* KO zebrafish exhibited strong developmental defects[238].

Other animal models to study FA have been developed including *Caenorhabditis elegans* and *Drosophila melanogaster* but most invertebrates have a minimal FA pathway that includes *FANCD1* (*BRCA2*), *-D2*, *-I*, *-M*, *-O* (*RAD51C*), and *-P* (*SLX4*). *FANCI* shares over 18% identity with *C. elegans* fnci-1 surrounding the ubiquitination site and treatment of germ cells with an ICL agent (photoactivated psoralen) led to a significant reduction (45%) in the hatching rate of fnci-1 mutant embryos[241]. Fnci-1 was further demonstrated to be involved in ICL repair in *C. elegans*, recapitulating what is known for *FANCI* in humans[241]. *FANCI* has not been directly studied in *D. melanogaster* but orthologs have been identified in silkworms (*Bombyx mori*)[242] and *Ciona intestinalis*[243] (an invertebrate closely related to vertebrates). In silkworms *FANCD2*, *-I*, and *-L* are the only FA genes conserved and both BmFancI and BmFancL are required for ubiquitination of BmFancD2 upon treatment with a DNA damaging agent (MMC), suggesting a similar function of the FA pathway in silkworms[242]. Finally, a large homozygous 3.3kb deletion of *FANCI* in Holstein cattle (*Bos taurus*) is the cause of brachyspina syndrome[244–246], which is a rare congenital defect characterized by reduced body weight, growth retardation, skeletal abnormalities, and malformation of the inner organs (kidney, heart, and gonads) (OMIA 000151-9913). This overlap of the Brachyspina syndrome phenotype with the human FA phenotypes, the only two known naturally occurring autosomal recessive syndromes caused by loss of *FANCI,* is interesting.

4.4 Clinical perspectives

Though there are many biological models available to study the function of proteins and their ability to promote tumourigenesis as described above, there are no

suitable models to study risk in humans. Determining the risk associated with genetic variation in humans relies on integration of genetic and epidemiological data[247]. This has been difficult even in major genes *BRCA1* and *BRCA2* where absolute risks for OC are presented as estimates within a range (*BRCA1* 39-58% and *BRCA2* 13-29%)[19]. As pointed out early on in the attempt to estimate penetrance for *BRCA1* and *BRCA2*, epidemiological factors need to be integrated with familial aggregation data so as not to inflate risk estimates[247]. Some of these epidemiological factors for OC could include risk for other cancers (BC), oral contraceptive pill use, age, and family history of cancers[9,248].

There is no effective screening strategy for OC and it is the fifth leading cause of cancer death in Canada[9]. The identification of *BRCA1* and *BRCA2* in 1994[16] and 1995[18], respectively, as major risk genes for OC was integral in changing the management and preventing OC in carriers of pathogenic variants as penetrance was established to be high. Penetrance is the relationship between genotype and phenotype, where the proportion of individuals in a population have a pathogenic variant and exhibit the associated phenotypic expression[191]. *FANCI* is unlikely to be highly penetrant (nearing 100% penetrance) similar to *RB1*[191], and perhaps also not as penetrant as *BRCA1* and *BRCA2* (up to 85%)[191].

As penetrance is high enough, asymptomatic carriers of pathogenic variants in *BRCA1* or *BRCA2* are now offered prophylactic bilateral salpingo-oophorectomy, removal of the fallopian tubes and ovaries, to significantly decrease the risk of OC[60,61]. It is not yet clear where *FANCI* may be integrated into clinical care, and it is premature to suggest risk reducing surgery for carriers of *FANCI* c.1813C>T; p.L605F or other likely

pathogenic/pathogenic variants, similar to other genes with unknown risks[19]. Though there is currently no evidence that carriers should be counselled based on an identified *FANCI* variant, I believe the inclusion of *FANCI* on multi-gene panels for clinical testing will help elucidate the penetrance for OC and possibly other cancers. This has been similarly argued for *BARD1* (a proposed BC predisposing gene), suggesting that the continued testing for pathogenic variants in these genes would allow for readily available translation of risk estimates and recommendations for screening or surgery once they become available[249]. A barrier to estimating the risk for *FANCI* c.1813C>T; p.L605F variant carriers has been the small number of cases identified[124]. Inclusion on clinical multi-gene panels would allow for the collection of more carriers to estimate penetrance. Estimating the penetrance of a moderate penetrance gene, *PALB2*, required the accumulation of over 500 families to provide estimates for OC and BC[40]. The inclusion of *FANCI* on research-based multi-gene panels has identified carriers of *FANCI* variants in BC cases[186], suggesting the inclusion on clinical multi-gene panel tests would be useful.

The identification of *BRCA1* and *BRCA2* led to the discovery of the mechanisms of action of the FA-HR DNA repair pathway, and development of targeted therapies (PARPi) that are now being implemented clinically[250–252]. The PARPi olaparib and niraparib have been approved for first-line maintenance therapy in Canada[253,254], with 20 clinical trials for PARPi and ovarian cancer not yet recruiting, recruiting, or active (not recruiting) in Canada[255]. Though our in cellulo data suggested FANCI p.L605F expressing cells are not responsive to olaparib[124], it is possible that the FA-HR pathway is deficient in HGSC tumours from carriers of *FANCI* variants and responsive to PARPi.

Therefore, *FANCI* variant carrier status may not be a biomarker for this targeted therapy. However, it is possible that *FANCI* variant carrier status may be associated with response to cisplatin based on the observation that FANCI p.L605F expressing cells exhibited sensitivity to this platinum chemotherapy[124]. Further clinical investigation of this observation could determine if *FANCI* carrier status can predict response to standard of care chemotherapy (cisplatin).

4.5 *FANCI* in other diseases

FANCI has been implicated in other non-cancer diseases, suggesting alternative functions for *FANCI* and/or roles of DNA repair pathways. *FANCI* has been reported as a hub gene in thoracic aortic dissection[256] (a tear in the wall of the aorta) and focal cortical dysplasia type II (a congenital abnormality affecting the organization of the layers of the brain) and was overexpressed in focal cortical dysplasia tissue compared to normal brain tissues[257]. FANCI has been related to BLM deficiency, where *FANCI* is differentially expressed in *BLM* deficient cell lines (from Bloom syndrome patients; a rare autosomal recessive disorder caused by pathogenic variants in *BLM*) compared to WT cell lines[258]. DNA methylation of *FANCI* was increased in lymphoblastoid cell lines from family members affected by bipolar disorder and major depressive disorder compared to married-in controls[259]. Immune biomarkers in cerebrospinal fluid of patients with bipolar disorder were associated with intronic SNVs in *FANCI* in a GWAS[260]. It is interesting to note that other members of the FA pathway, such as *FAN1*[261–264] and *FANCL*[265], have been associated with psychiatric disorders. Further investigation of these reports, which have yet to be validated, will be important to understand the pleiotropic effects of *FANCI* in a diverse spectrum of diseases.

4.6 Proposed models of *FANCI* as an OC predisposing gene

Based on the data presented in this thesis I propose two alternative models for *FANCI*: Model 1: *FANCI* is an OC risk gene, or Model 2: *FANCI* affects the development or progression of OC (Figure 4.2). The first model focuses on risk to an individual with a pathogenic *FANCI* variant who has an increased risk of OC compared to an individual with unknown pathogenic variant status having a population level risk of OC. This model was addressed in Chapter 2.0, though as mentioned the penetrance of *FANCI* is currently unknown. Over 100 genes exemplify this model as CPGs[6,7,64], though as mentioned throughout this thesis OC predisposing genes include *BRCA1*, *BRCA2*, *MLH1*, *MSH2*, *PMS2*, *MSH6*, *BRIP1*, *RAD51C*, *RAD51D*, *ATM*, and *PALB2*.

In Model 2 an individual with a pathogenic *FANCI* variant could develop the disease earlier and/or it may progress more rapidly compared to an individual with unknown pathogenic variant status who could have a stochastic development of OC and natural disease progression. An example of this model is the stepwise evolution of normal epithelium to adenoma to a chromosomally unstable colorectal cancer where alterations in key genes such as *KRAS*, *CDC4*, or *TP53* could influence progression to tumourigenesis[266].

These models could be applied to other cancer types and provides an opportunity to design future studies to better understand the molecular pathogenesis of CPGs with the ultimate goal of discovering new therapies to treat or prevent cancer.

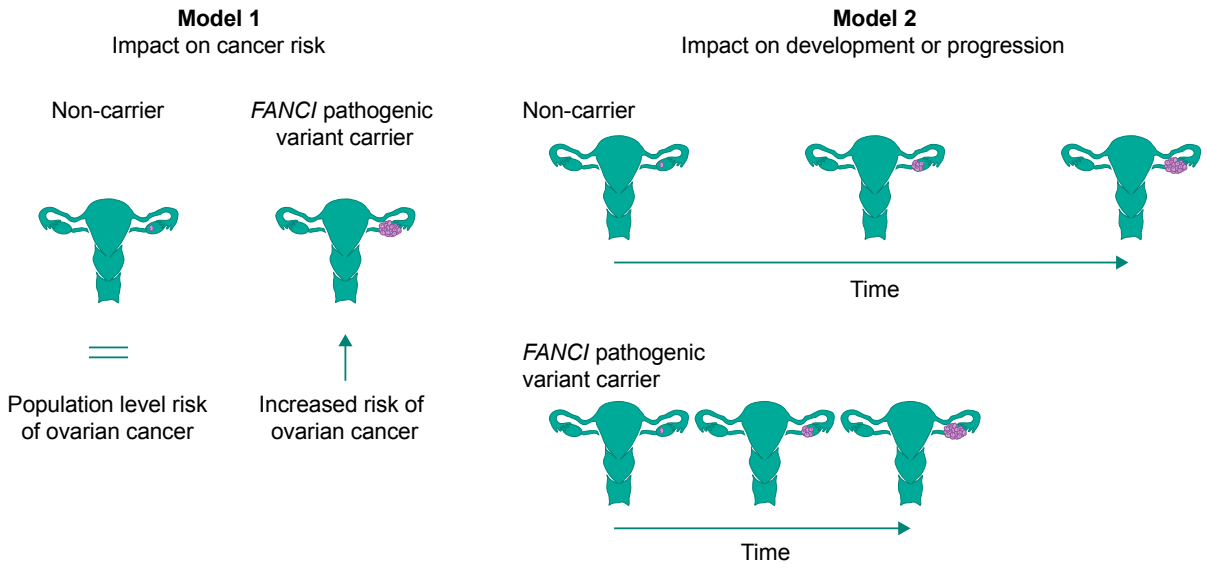


Figure 4.2. Models for *FANCI* involvement in risk (Model 1) or development or progression (Model 2).

5.0 Conclusion and future directions

5.1 Conclusions

Identification of new OC predisposing genes and CPGs for other cancer types has become complicated, as we expect each new predisposing gene identified will account for a small proportion of cases. I identified *FANCI* c.1813C>T; p.L605F in familial OC cases and my data suggest this is a likely pathogenic variant associated with familial OC. My strategy focusing on families with multiple cases of OC from a genetically unique population and including missense variants as candidates allowed me to identify this *FANCI* variant. In cellulo and in vitro analyses complemented the genetic analyses and provided evidence that FANCI p.L605F encodes a protein that is unstable and not fully functional compared to the WT. Investigation of tumours from *FANCI* c.1813C>T; p.L605F carriers showed that the somatic genetic characteristics (commonly mutated genes, CNAs, and mutational signatures) of these cases was consistent with those known for HGSC cases. Germline *FANCI* c.1813C>T; p.L605F and somatic *FANCI* variants can be identified in many different cancer types suggesting a possible role in these diseases. This thesis has provided evidence for *FANCI* as a new candidate OC predisposing gene.

5.2 Future directions

There are many possible avenues for future research pertaining to the role of *FANCI* not only in OC, but the possibility that it may play a role in other cancer types. In FCs we identified only *FANCI* c.1813C>T; p.L605F as a likely pathogenic variant, but it is evident in different populations that other candidates are identifiable and permit further investigation. The identification of other likely pathogenic variant carriers will assist in

determining the penetrance associated with such variants, though as mentioned this will require the identification of many more carriers. The possibility of *FANCI* c.1813C>T; p.L605F carriers responding well to some cancer related drugs (cisplatin), but not to others (PARPi)[124], suggests that further investigation into the drug response of *FANCI* carriers is warranted. We are interested in the role *FANCI* may play in OC tumour progression and/or initiation, which will be investigated by means of a novel mouse model of HGSC using a CRISPR/Cas9 system and electroporation of the fallopian tube[235]. Other missense variants identified, especially those found in OC families, such as *FANCI* c.286G>A; p.E96K, can be modelled using similar in cellulo and in vitro analyses as those used to investigate the function of FANCI p.L605F[124]. A *FANCI* KO mouse model has been developed and viable progeny were reported, though with sub Mendelian ratios[236], I suggest the investigation of cancer incidence in these mice as it was not reported previously. This work has raised many questions and offers a breadth of opportunities for future work for the new candidate OC predisposing gene, *FANCI*.

6.0 Reference list

1. Carrell, R. W. *et al.* Structure and variation of human alpha 1-antitrypsin. *Nature* **298**, 329–334 (1982).
2. de Verneuil, H., Grandchamp, B., Beaumont, C., Picat, C. & Nordmann, Y. Uroporphyrinogen decarboxylase structural mutant (Gly281----Glu) in a case of porphyria. *Science* **234**, 732–734 (1986).
3. Fung, Y. K. *et al.* Structural evidence for the authenticity of the human retinoblastoma gene. *Science* **236**, 1657–1661 (1987).
4. Tsuji, S. *et al.* A mutation in the human glucocerebrosidase gene in neuronopathic Gaucher's disease. *N. Engl. J. Med.* **316**, 570–575 (1987).
5. Grandchamp, B. *et al.* Tissue-specific splicing mutation in acute intermittent porphyria. *Proc. Natl. Acad. Sci. U. S. A.* **86**, 661–664 (1989).
6. Huang, K.-L. *et al.* Pathogenic germline variants in 10,389 adult cancers. *Cell* **173**, 355-370.e14 (2018).
7. Rasnic, R., Linial, N. & Linial, M. Expanding cancer predisposition genes with ultra-rare cancer-exclusive human variations. *Sci. Rep.* **10**, 13462 (2020).
8. Siegel, R. L., Miller, K. D., Fuchs, H. E. & Jemal, A. Cancer Statistics, 2021. *CA. Cancer J. Clin.* **71**, 7–33 (2021).
9. Canadian Cancer Statistics Advisory Committee in collaboration with the Canadian Cancer Society, Statistics Canada and the Public Health Agency of Canada. Canadian Cancer Statistics 2021. cancer.ca/Canadian-Cancer-Statistics-2021-EN (2021).

10. Brenner, D. R. *et al.* Projected estimates of cancer in Canada in 2022. *CMAJ Can. Med. Assoc. J. J. Assoc. Medicale Can.* **194**, E601–E607 (2022).
11. Lichtenstein, P. *et al.* Environmental and heritable factors in the causation of cancer: Analyses of cohorts of twins from Sweden, Denmark, and Finland. *N. Engl. J. Med.* **343**, 78–85 (2000).
12. Mucci, L. A. *et al.* Familial risk and heritability of cancer among twins in nordic countries. *JAMA* **315**, 68–76 (2016).
13. Stratton, J. F., Pharoah, P., Smith, S. K., Easton, D. & Ponder, B. a. A systematic review and meta-analysis of family history and risk of ovarian cancer. *Br. J. Obstet. Gynaecol.* **105**, 493–499 (1998).
14. Hall, J. M. *et al.* Linkage of early-onset familial breast cancer to chromosome 17q21. *Science* **250**, 1684–1689 (1990).
15. Narod, S. *et al.* Familial breast-ovarian cancer locus on chromosome 17q12-q23. *The Lancet* **338**, 82–83 (1991).
16. Miki, Y. *et al.* Strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. *Science* **266**, 66–71 (1994).
17. Wooster, R. *et al.* Localization of a breast cancer susceptibility gene, BRCA2, to chromosome 13q12-13. *Science* **265**, 2088–2090 (1994).
18. Wooster, R. *et al.* Identification of the breast cancer susceptibility gene BRCA2. *Nature* **378**, 789–92 (1995).
19. Daly, M. B. *et al.* NCCN guidelines insights: Genetic/familial high-risk assessment: Breast, ovarian, and pancreatic, version 1.2020. *J. Natl. Compr. Canc. Netw.* **18**, 380–391 (2020).

20. Risch, H. A. *et al.* Prevalence and penetrance of germline BRCA1 and BRCA2 mutations in a population series of 649 women with ovarian cancer. *Am. J. Hum. Genet.* **68**, 700–710 (2001).
21. Rubin, S. C. *et al.* BRCA1, BRCA2, and hereditary nonpolyposis colorectal cancer gene mutations in an unselected ovarian cancer population: relationship to family history and implications for genetic testing. *Am. J. Obstet. Gynecol.* **178**, 670–677 (1998).
22. Pal, T. *et al.* BRCA1 and BRCA2 mutations account for a large proportion of ovarian carcinoma cases. *Cancer* **104**, 2807–2816 (2005).
23. Lynch, H. T., Snyder, C. L., Shaw, T. G., Heinen, C. D. & Hitchins, M. P. Milestones of Lynch syndrome: 1895-2015. *Nat. Rev. Cancer* (2015) doi:10.1038/nrc3878.
24. Papadopoulos, N. *et al.* Mutation of a mutL homolog in hereditary colon cancer. *Science* **263**, 1625–1629 (1994).
25. Bronner, C. E. *et al.* Mutation in the DNA mismatch repair gene homologue hMLH1 is associated with hereditary non-polyposis colon cancer. *Nature* **368**, 258–261 (1994).
26. Fishel, R. *et al.* The human mutator gene homolog MSH2 and its association with hereditary nonpolyposis colon cancer. *Cell* **75**, 1027–1038 (1993).
27. Miyaki, M. *et al.* Germline mutation of MSH6 as the cause of hereditary nonpolyposis colorectal cancer. *Nat. Genet.* **17**, 271–272 (1997).
28. Nicolaides, N. C. *et al.* Mutations of two PMS homologues in hereditary nonpolyposis colon cancer. *Nature* **371**, 75–80 (1994).

29. Idos, G. & Valle, L. Lynch Syndrome. in *GeneReviews(®)* (eds. Adam, M. P. et al.) (University of Washington, Seattle, 1993).
30. Pavanello, M. *et al.* Rare germline genetic variants and the risks of epithelial ovarian cancer. *Cancers* **12**, 3046–3046 (2020).
31. Flaum, N. *et al.* High detection rate from genetic testing in BRCA-negative women with familial epithelial ovarian cancer. *Genet. Med. Off. J. Am. Coll. Med. Genet.* S1098-3600(22)00913–3 (2022) doi:10.1016/j.gim.2022.08.022.
32. Song, H. *et al.* Tagging single nucleotide polymorphisms in the BRIP1 gene and susceptibility to breast and ovarian cancer. *PLoS ONE* **2**, e268–e268 (2007).
33. Rafnar, T. *et al.* Mutations in BRIP1 confer high risk of ovarian cancer. *Nat. Genet.* **43**, 1104–1107 (2011).
34. Meindl, A. *et al.* Germline mutations in breast and ovarian cancer pedigrees establish RAD51C as a human cancer susceptibility gene. *Nat. Genet.* **42**, 410–414 (2010).
35. Zheng, Y. *et al.* Screening RAD51C nucleotide alterations in patients with a family history of breast and ovarian cancer. *Breast Cancer Res. Treat.* **124**, 857–861 (2010).
36. Somyajit, K., Subramanya, S. & Nagaraju, G. RAD51C: A novel cancer susceptibility gene is linked to Fanconi anemia and breast cancer. *Carcinogenesis* **31**, 2031–2038 (2010).
37. Loveday, C. *et al.* Germline mutations in RAD51D confer susceptibility to ovarian cancer. *Nat. Genet.* **43**, 879–882 (2011).

38. Rahman, N. *et al.* PALB2, which encodes a BRCA2-interacting protein, is a breast cancer susceptibility gene. *Nat. Genet.* **39**, 165–167 (2007).
39. Erkko, H. *et al.* A recurrent mutation in PALB2 in Finnish cancer families. *Nature* **446**, 316–319 (2007).
40. Yang, X. *et al.* Cancer Risks Associated With Germline PALB2 Pathogenic Variants: An International Study of 524 Families. *J. Clin. Oncol.* **38**, 674–685 (2020).
41. Athma, P., Rappaport, R. & Swift, M. Molecular genotyping shows that ataxia-telangiectasia heterozygotes are predisposed to breast cancer. *Cancer Genet. Cytogenet.* **92**, 130–134 (1996).
42. FitzGerald, M. G. *et al.* Heterozygous ATM mutations do not contribute to early onset of breast cancer. *Nat. Genet.* **15**, 307–310 (1997).
43. Chen, J., Birkholtz, G. G., Lindblom, P., Rubio, C. & Lindblom, A. The role of ataxia-telangiectasia heterozygotes in familial breast cancer. *Cancer Res.* **58**, 1376–1379 (1998).
44. Bay, J. O. *et al.* No evidence for constitutional ATM mutation in breast/gastric cancer families. *Int. J. Oncol.* **12**, 1385–1390 (1998).
45. Pennington, K. P. *et al.* Germline and Somatic Mutations in Homologous REcombination Genes Predict Platinum Response and Survival in Ovarian, Fallopian Tube, and Peritoneal Carcinomas. *Clin Cancer Res* **20**, 764–775 (2014).
46. Lu, H. M. *et al.* Association of breast and ovarian cancers with predisposition genes identified by large-scale sequencing. *JAMA Oncol.* **5**, 51–57 (2019).

47. Lilyquist, J. *et al.* Frequency of mutations in a large series of clinically ascertained ovarian cancer cases tested on multi-gene panels compared to reference controls. *Gynecol. Oncol.* **147**, 375–380 (2017).
48. Norquist, B. M. *et al.* Inherited Mutations in Women With Ovarian Carcinoma. *JAMA Oncol.* **2**, 482–490 (2016).
49. Phelan, C. M. *et al.* Identification of 12 new susceptibility loci for different histotypes of epithelial ovarian cancer. *Nat. Genet.* **49**, 680–691 (2017).
50. Kuchenbaecker, K. B. *et al.* Identification of six new susceptibility loci for invasive epithelial ovarian cancer. *Nat. Genet.* **47**, 164–171 (2015).
51. Yang, X. *et al.* Evaluation of polygenic risk scores for ovarian cancer risk prediction in a prospective cohort study. *J. Med. Genet.* **55**, 546–554 (2018).
52. Dareng, E. O. *et al.* Polygenic risk modeling for prediction of epithelial ovarian cancer risk. *Eur. J. Hum. Genet. EJHG* **30**, 349–362 (2022).
53. Reid, B. M., Permuth, J. B. & Sellers, T. A. Epidemiology of ovarian cancer: A summary review. *Cancer Biol Med* **14**, 9–32 (2017).
54. Song, H. *et al.* Contribution of germline mutations in the RAD51B, RAD51C, and RAD51D genes to ovarian cancer in the population. *J. Clin. Oncol.* **33**, 2901–2907 (2015).
55. Ramus, S. J. *et al.* Germline mutations in the BRIP1, BARD1, PALB2, and NBN genes in women with ovarian cancer. *J. Natl. Cancer Inst.* **107**, djv214–djv214 (2015).
56. Lakhani, S. R. *et al.* Pathology of ovarian cancers in BRCA1 and BRCA2 carriers. *Clin. Cancer Res.* **10**, 2473–2481 (2004).

57. Helder-Woolderink, J. M. *et al.* Ovarian cancer in Lynch syndrome; a systematic review. *Eur. J. Cancer Oxf. Engl. 1990* **55**, 65–73 (2016).

58. Cummings, S. *et al.* Age of ovarian cancer diagnosis among BRIP1, RAD51C, and RAD51D mutation carriers identified through multi-gene panel testing. *J. Ovarian Res.* **14**, 61 (2021).

59. Menon, U. *et al.* Ovarian cancer population screening and mortality after long-term follow-up in the UK Collaborative Trial of Ovarian Cancer Screening (UKCTOCS): a randomised controlled trial. *Lancet Lond. Engl.* **397**, 2182–2193 (2021).

60. Finch, A. *et al.* Salpingo-oophorectomy and the risk of ovarian, fallopian tube, and peritoneal cancers in women with a BRCA1 or BRCA2 mutation. *JAMA* **296**, 185–192 (2006).

61. Finch, A. P. M. *et al.* Impact of oophorectomy on cancer incidence and mortality in women with a BRCA1 or BRCA2 mutation. *J. Clin. Oncol.* **32**, 1547–1554 (2014).

62. Lheureux, S., Gourley, C., Vergote, I. & Oza, A. M. Epithelial ovarian cancer. *Lancet Lond. Engl.* **393**, 1240–1253 (2019).

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

64. Rahman, N. Realizing the promise of cancer predisposition genes. *Nature* **505**, 302–308 (2014).

65. Levine, A. J. p53: 800 million years of evolution and 40 years of discovery. *Nat. Rev. Cancer* **20**, 471–480 (2020).

66. Sanchez-Vega, F. *et al.* Oncogenic Signaling Pathways in The Cancer Genome Atlas. *Cell* **173**, 321-337.e10 (2018).

67. Campbell, P. J. *et al.* Pan-cancer analysis of whole genomes. *Nature* **578**, 82–93 (2020).

68. Bell, D. *et al.* Integrated genomic analyses of ovarian carcinoma. *Nature* **474**, 609–615 (2011).

69. Kanchi, K. L. *et al.* Integrated analysis of germline and somatic variants in ovarian cancer. *Nat. Commun.* **5**, 3156–3156 (2014).

70. Nakayama, N. *et al.* Gene amplification CCNE1 is related to poor survival and potential therapeutic target in ovarian cancer. *Cancer* **116**, 2621–2634 (2010).

71. Ramus, S. J. & Gayther, S. A. The contribution of BRCA1 and BRCA2 to ovarian cancer. *Mol. Oncol.* **3**, 138–150 (2009).

72. Ott, J., Wang, J. & Leal, S. M. Genetic linkage analysis in the age of whole-genome sequencing. *Nat. Rev. Genet.* **16**, 275–284 (2015).

73. Fierheller, C. T., Alenezi, W. M. & Tonin, P. N. The Genetic Analyses of French Canadians of Quebec Facilitate the Characterization of New Cancer Predisposing Genes Implicated in Hereditary Breast and/or Ovarian Cancer Syndrome Families. *Cancers* **13**, (2021).

74. Collins, F. S., Morgan, M. & Patrinos, A. The Human Genome Project: lessons from large-scale biology. *Science* **300**, 286–290 (2003).

75. Pulst, S. M. Genetic Linkage Analysis. *Arch. Neurol.* **56**, 667–672 (1999).

76. Kwon, J. M. & Goate, A. M. The candidate gene approach. *Alcohol Res. Health J. Natl. Inst. Alcohol Abuse Alcohol.* **24**, 164–168 (2000).

77. Prakash, R., Zhang, Y., Feng, W. & Jasin, M. Homologous recombination and human health: the roles of BRCA1, BRCA2, and associated proteins. *Cold Spring Harb. Perspect. Biol.* **7**, a016600 (2015).

78. Clark, M. J. *et al.* Performance comparison of exome DNA sequencing technologies. **29**, 908–914 (2014).

79. Gilissen, C., Hoischen, A., Brunner, H. G. & Veltman, J. A. Unlocking Mendelian disease using exome sequencing. *Genome Biol.* **12**, 228 (2011).

80. Davydov, E. V., Goode, D. L., Sirota, M., Cooper, G. M. & Sidow, A. Identifying a high fraction of the human genome to be under selective constraint using GERP ++. *PLoS Comput. Biol.* **6**, e1001025–e1001025 (2010).

81. Sim, N. *et al.* SIFT web server: predicting effects of amino acid substitutions on proteins. *Nucleic Acids Res.* **40**, 452–457 (2012).

82. Adzhubei, I. A. *et al.* A method and server for predicting damaging missense mutations. *Nat. Methods* **7**, 248–249 (2010).

83. Ghosh, R., Oak, N. & Plon, S. E. Evaluation of in silico algorithms for use with ACMG/AMP clinical variant interpretation guidelines. *Genome Biol.* **18**, (2017).

84. Niroula, A. & Vihinen, M. How good are pathogenicity predictors in detecting benign variants? *PLoS Comput. Biol.* **15**, e1006481 (2019).

85. Cubuk, C. *et al.* Clinical likelihood ratios and balanced accuracy for 44 in silico tools against multiple large-scale functional assays of cancer susceptibility genes. *Genet. Med. Off. J. Am. Coll. Med. Genet.* **23**, 2096–2104 (2021).

86. Gunning, A. C. *et al.* Assessing performance of pathogenicity predictors using clinically relevant variant datasets. *J. Med. Genet.* **58**, 547–555 (2021).

87. Ramsey, S. D., Yoon, P., Moonesinghe, R. & Khoury, M. J. Population-based study of the prevalence of family history of cancer: Implications for cancer screening and prevention. *Genet. Med.* **8**, 571–575 (2006).

88. Bray, S. M. *et al.* Signatures of founder effects, admixture, and selection in the Ashkenazi Jewish population. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 16222–16227 (2010).

89. Kääriäinen, H., Muilu, J., Perola, M. & Kristiansson, K. Genetics in an isolated population like Finland: a different basis for genomic medicine? *J. Community Genet.* **8**, 319–326 (2017).

90. Sunna Ebenesersdóttir, S. *et al.* Ancient genomes from Iceland reveal the making of a human population. *Science* **360**, 1028–1032 (2018).

91. Moreau, C., Vézina, H. & Labuda, D. Founder effects and genetic variability in Quebec. *Medecine/Sciences* **23**, 1008–1013 (2007).

92. Henn, B. M., Cavalli-Sforza, L. L. & Feldman, M. W. The great human expansion. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 17758–17764 (2012).

93. Wright, S. Isolation by Distance. *Genetics* **28**, 114–138 (1943).

94. Ramachandran, S. *et al.* Support from the relationship of genetic and geographic distance in human populations for a serial founder effect originating in Africa. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 15942–15947 (2005).

95. Ostrer, H. A genetic profile of contemporary Jewish populations. *Nat. Rev. Genet.* **2**, 891–898 (2001).

96. Scriver, C. R. Human genetics: Lessons from Quebec populations. *Annu. Rev. Genomics Hum. Genet.* **2**, 69–101 (2001).

97. Laberge, A.-M. *et al.* Population history and its impact on medical genetics in Quebec. *Clin Genet* **68**, 287–301 (2005).

98. Landrum, M. J. *et al.* ClinVar: Improving access to variant interpretations and supporting evidence. *Nucleic Acids Res.* **46**, 1062–1067 (2018).

99. Alenezi, W. M. *et al.* The Genetic and Molecular Analyses of RAD51C and RAD51D Identifies Rare Variants Implicated in Hereditary Ovarian Cancer from a Genetically Unique Population. *Cancers* **14**, (2022).

100. Chen, C. F. *et al.* The nuclear localization sequences of the BRCA1 protein interact with the importin-alpha subunit of the nuclear transport signal receptor. *J. Biol. Chem.* **271**, 32863–32868 (1996).

101. Spain, B. H., Larson, C. J., Shihabuddin, L. S., Gage, F. H. & Verma, I. M. Truncated BRCA2 is cytoplasmic: implications for cancer-linked mutations. *Proc. Natl. Acad. Sci. U. S. A.* **96**, 13920–13925 (1999).

102. Tibbetts, R. S. *et al.* Functional interactions between BRCA1 and the checkpoint kinase ATR during genotoxic stress. *Genes Dev.* **14**, 2989–3002 (2000).

103. Yang, H. *et al.* BRCA2 function in DNA binding and recombination from a BRCA2-DSS1-ssDNA structure. *Science* **297**, 1837–1848 (2002).

104. Pavlicek, A. *et al.* Evolution of the tumor suppressor BRCA1 locus in primates: implications for cancer predisposition. *Hum. Mol. Genet.* **13**, 2737–2751 (2004).

105. Teng, D. H. *et al.* Low incidence of BRCA2 mutations in breast carcinoma and other cancers. *Nat. Genet.* **13**, 241–244 (1996).

106. Paull, T. T., Cortez, D., Bowers, B., Elledge, S. J. & Gellert, M. Direct DNA binding by Brca1. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 6086–6091 (2001).

107. Zhao, W., Wiese, C., Kwon, Y., Hromas, R. & Sung, P. The BRCA Tumor Suppressor Network in Chromosome Damage Repair by Homologous Recombination. *Annu. Rev. Biochem.* **88**, 221–245 (2019).
108. Takata, M. *et al.* Homologous recombination and non-homologous end-joining pathways of DNA double-strand break repair have overlapping roles in the maintenance of chromosomal integrity in vertebrate cells. *EMBO J.* **17**, 5497–5508 (1998).
109. Scully, R., Panday, A., Elango, R. & Willis, N. A. DNA double-strand break repair-pathway choice in somatic mammalian cells. *Nat. Rev. Mol. Cell Biol.* **20**, 698–714 (2019).
110. Li, G.-M. Mechanisms and functions of DNA mismatch repair. *Cell Res.* **18**, 85–98 (2008).
111. Swisher, E. M. *et al.* Molecular and clinical determinants of response and resistance to rucaparib for recurrent ovarian cancer treatment in ARIEL2 (Parts 1 and 2). *Nat. Commun.* **12**, 2487 (2021).
112. Bruand, M. *et al.* Cell-autonomous inflammation of BRCA1-deficient ovarian cancers drives both tumor-intrinsic immunoreactivity and immune resistance via STING. *Cell Rep.* **36**, 109412 (2021).
113. Lindahl, T. & Barnes, D. E. Repair of endogenous DNA damage. *Cold Spring Harb. Symp. Quant. Biol.* **65**, 127–133 (2000).
114. Grillari, J., Katinger, H. & Voglauer, R. Contributions of DNA interstrand cross-links to aging of cells and organisms. *Nucleic Acids Res.* **35**, 7566–7576 (2007).

115. Noll, D. M., Mason, T. M. & Miller, P. S. Formation and repair of interstrand cross-links in DNA. *Chem. Rev.* **106**, 277–301 (2006).

116. Lobitz, S. & Velleuer, E. Guido Fanconi (1892-1979): a jack of all trades. *Nat. Rev. Cancer* **6**, 893–898 (2006).

117. Mehta, P. A. & Tolar, J. Fanconi Anemia. in *GeneReviews(®) [Internet]* (eds. Pagon, R., Adam, M. & Ardinger, H.) (Seattle (WA): University of Washington, Seattle, 2002).

118. Dayhoff, M. O. The origin and evolution of protein superfamilies. *Fed. Proc.* **35**, 2132–2138 (1976).

119. Wang, A. T. *et al.* A Dominant Mutation in Human RAD51 Reveals Its Function in DNA Interstrand Crosslink Repair Independent of Homologous Recombination. *Mol. Cell* **59**, 478–490 (2015).

120. Nalepa, G. & Clapp, D. W. Fanconi anaemia and cancer: an intricate relationship. *Nat. Rev. Cancer* **18**, 168–185 (2018).

121. Alter, B. P. & Rosenberg, P. S. VACTERL-H Association and Fanconi Anemia. *Mol. Syndromol.* **4**, 87–93 (2013).

122. McReynolds, L. J. *et al.* Risk of cancer in heterozygous relatives of patients with Fanconi anemia. *Genet. Med. Off. J. Am. Coll. Med. Genet.* **24**, 245–250 (2022).

123. Tonin, P. N., Maugard, C. M., Perret, C., Mes-Masson, A. M. & Provencher, D. M. A review of histopathological subtypes of ovarian cancer in BRCA-related French Canadian cancer families. *Fam. Cancer* **6**, 491–497 (2007).

124. Fierheller, C. T. *et al.* A functionally impaired missense variant identified in French Canadian families implicates FANCI as a candidate ovarian cancer-predisposing gene. *Genome Med.* **13**, 186 (2021).
125. Poteete, A. R., Rennell, D. & Bouvier, S. E. Functional significance of conserved amino acid residues. *Proteins* **13**, 38–40 (1992).
126. Garner, E. & Smogorzewska, A. Ubiquitylation and the Fanconi anemia pathway. *FEBS Lett.* **585**, 2853–2860 (2011).
127. Moldovan, G.-L. & D'Andrea, A. D. How the fanconi anemia pathway guards the genome. *Annu. Rev. Genet.* **43**, 223–49 (2009).
128. Hinds, P. W. & Weinberg, R. A. Tumor suppressor genes. *Curr. Opin. Genet. Dev.* **4**, 135–141 (1994).
129. 1000 Genomes Project Consortium *et al.* A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
130. EVS. Exome Variant Server. *NHLBI GO Exome Seq. Proj. ESP* (2014).
131. Gibson, G. Rare and common variants: Twenty arguments. *Nat. Rev. Genet.* **13**, 135–145 (2012).
132. Karczewski, K. J. *et al.* The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **581**, 434–443 (2020).
133. Vahteristo, P. *et al.* A CHEK2 genetic variant contributing to a substantial fraction of familial breast cancer. *Am. J. Hum. Genet.* **71**, 432–438 (2002).
134. Levitus, M. *et al.* Heterogeneity in Fanconi anemia : evidence for 2 new genetic subtypes. **103**, 2498–2503 (2004).

135. Sims, A. E. *et al.* FANCI is a second monoubiquitinated member of the Fanconi anemia pathway. *Nat. Struct. Mol. Biol.* **14**, 564–567 (2007).

136. Dorsman, J. C. *et al.* Identification of the Fanconi anemia complementation group I gene, FANCI. *Cell. Oncol.* **29**, 211–218 (2007).

137. Smogorzewska, A. *et al.* Identification of the Fanconi anemia (FANC) I protein, a monoubiquitinated FANCD2 paralog required for crosslink repair. *Cell* **129**, 289–301 (2007).

138. Yuan, F., El Hokayem, J., Zhou, W. & Zhang, Y. FANCI protein binds to DNA and interacts with FANCD2 to recognize branched structures. *J. Biol. Chem.* **284**, 24443–24452 (2009).

139. Yang, K. *et al.* Regulation of the Fanconi anemia pathway by a SUMO-like delivery network. *Genes Dev.* **25**, 1847–1858 (2011).

140. Siddiqui, M. Q. *et al.* Structural and biophysical properties of h-FANCI ARM repeat protein. *J. Biomol. Struct. Dyn.* **35**, 3032–3042 (2017).

141. Colnaghi, L. *et al.* Patient-derived C-terminal mutation of FANCI causes protein mislocalization and reveals putative EDGE motif function in DNA repair. *Blood* **117**, 2247–2256 (2011).

142. Ishiai, M. *et al.* FANCI phosphorylation functions as a molecular switch to turn on the Fanconi anemia pathway. *Nat. Struct. Mol. Biol.* **15**, 1138–1146 (2008).

143. Ishiai, M. *et al.* Activation of the FA pathway mediated by phosphorylation and ubiquitination. *Mutat. Res.* **803–805**, 89–95 (2017).

144. Castella, M. *et al.* FANCI regulates recruitment of the FA core complex at sites of DNA damage independently of FANCD2. *PLoS Genet.* **11**, e1005563–e1005563 (2015).

145. Renaudin, X., Guervilly, J.-H., Aoufouchi, S. & Rosselli, F. Proteomic analysis reveals a FANCA-modulated neddylation pathway involved in CXCR5 membrane targeting and cell mobility. *J. Cell Sci.* **127**, 3546–3554 (2014).

146. Shigechi, T. *et al.* ATR-ATRIP kinase complex triggers activation of the Fanconi anemia DNA repair pathway. *Cancer Res.* **72**, 1149–1156 (2012).

147. Tomida, J. *et al.* A novel interplay between the Fanconi anemia core complex and ATR-ATRIP kinase during DNA cross-link repair. *Nucleic Acids Res.* **41**, 6930–6941 (2013).

148. Tan, W., van Twest, S., Murphy, V. J. & Deans, A. J. ATR-Mediated FANCI Phosphorylation Regulates Both Ubiquitination and Deubiquitination of FANCD2. *Front. Cell Dev. Biol.* **8**, 2 (2020).

149. Sijacki, T. *et al.* The DNA-damage kinase ATR activates the FANCD2-FANCI clamp by priming it for ubiquitination. *Nat. Struct. Mol. Biol.* (2022) doi:10.1038/s41594-022-00820-9.

150. Williams, S. A. *et al.* Functional and physical interaction between the mismatch repair and FA-BRCA pathways. *Hum. Mol. Genet.* **20**, 4395–4410 (2011).

151. Clark, D. W., Tripathi, K., Dorsman, J. C. & Palle, K. FANCJ protein is important for the stability of FANCD2/FANCI proteins and protects them from proteasome and caspase-3 dependent degradation. *Oncotarget* **6**, 28816–28832 (2015).

152. Sims, A. E. *et al.* FANCI is a second monoubiquitinated member of the Fanconi anemia pathway. *Nat. Struct. Mol. Biol.* **14**, 564–567 (2007).
153. Smogorzewska, A. *et al.* Identification of the Fanconi anemia (FANC) I protein, a monoubiquitinated FANCD2 paralog required for crosslink repair. *Cell* **129**, 289–301 (2007).
154. Ishiai, M. *et al.* FANCI phosphorylation functions as a molecular switch to turn on the Fanconi anemia pathway. *Nat. Struct. Mol. Biol.* **15**, 1138–1146 (2008).
155. Wang, R., Wang, S., Dhar, A., Peralta, C. & Pavletich, N. P. DNA clamp function of the monoubiquitinated Fanconi anaemia ID complex. *Nature* **580**, 278–282 (2020).
156. Boisvert, R. A. & Howlett, N. G. The Fanconi anemia ID2 complex: dueling saxes at the crossroads. *Cell Cycle Georget. Tex* **13**, 2999–3015 (2014).
157. Tan, W. *et al.* Monoubiquitination by the human Fanconi anemia core complex clamps FANCI:FANCD2 on DNA in filamentous arrays. *eLife* **9**, (2020).
158. Gibbs-Seymour, I. *et al.* Ubiquitin-SUMO circuitry controls activated fanconi anemia ID complex dosage in response to DNA damage. *Mol. Cell* **57**, 150–164 (2015).
159. Sareen, A., Chaudhury, I., Adams, N. & Sobeck, A. Fanconi anemia proteins FANCD2 and FANCI exhibit different DNA damage responses during S-phase. *Nucleic Acids Res.* **40**, 8425–8439 (2012).
160. Longerich, S. *et al.* Regulation of FANCD2 and FANCI monoubiquitination by their interaction and by DNA. *Nucleic Acids Res.* **42**, 5657–5670 (2014).
161. Chaudhury, I., Sareen, A., Raghunandan, M. & Sobeck, A. FANCD2 regulates BLM complex functions independently of FANCI to promote replication fork recovery. *Nucleic Acids Res.* **41**, 6444–6459 (2013).

162. Oka, Y., Bekker-Jensen, S. & Mailand, N. Ubiquitin-like protein UBL5 promotes the functional integrity of the Fanconi anemia pathway. *EMBO J.* **34**, 1385–1398 (2015).

163. Moriel-Carretero, M., Ovejero, S., Gérus-Durand, M., Vryzas, D. & Constantinou, A. Fanconi anemia FANCD2 and FANCI proteins regulate the nuclear dynamics of splicing factors. *J. Cell Biol.* **216**, 4007–4026 (2017).

164. Sato, K. *et al.* Histone chaperone activity of Fanconi anemia proteins, FANCD2 and FANCI, is required for DNA crosslink repair. *EMBO J.* **31**, 3524–3536 (2012).

165. Liang, Z. *et al.* Binding of FANCI-FANCD2 Complex to RNA and R-Loops Stimulates Robust FANCD2 Monoubiquitination. *Cell Rep.* **26**, 564-572.e5 (2019).

166. Sato, K. *et al.* FANCI-FANCD2 stabilizes the RAD51-DNA complex by binding RAD51 and protects the 5'-DNA end. *Nucleic Acids Res.* **44**, 10758–10771 (2016).

167. Chen, Y. *et al.* ATR-mediated phosphorylation of FANCI regulates dormant origin firing in response to replication stress. *Mol. Cell* **58**, 323–338 (2016).

168. Zhang, X., Lu, X., Akhter, S., Georgescu, M. & Legerski, R. J. FANCI is a negative regulator of Akt activation. *Cell Cycle* **15**, 1134–1143 (2016).

169. Sondalle, S. B., Longerich, S., Ogawa, L. M., Sung, P. & Baserga, S. J. Fanconi anemia protein FANCI functions in ribosome biogenesis. *PNAS* **116**, 2561–2570 (2019).

170. Savage, S. A. *et al.* Novel FANCI Mutations in Fanconi Anemia with VACTERL Association. *Am. J. Med. Genet. A.* **170A**, 386–391 (2015).

171. Mori, M. *et al.* Pathogenic mutations identified by a multimodality approach in 117 Japanese Fanconi anemia patients. *Haematologica* **104**, 1962–1973 (2019).

172. George, M. *et al.* A comprehensive molecular study identified 12 complementation groups with 56 novel FANC gene variants in Indian Fanconi anemia subjects. *Hum. Mutat.* **42**, 1648–1665 (2021).
173. Mantere, T. *et al.* Finnish Fanconi anemia mutations and hereditary predisposition to breast and prostate cancer. *Clin Genet* **88**, 68–73 (2015).
174. Chandrasekharappa, S. C. *et al.* Massively parallel sequencing , aCGH , and RNA-Seq technologies provide a comprehensive molecular diagnosis of Fanconi anemia. *Blood* **121**, (2013).
175. Kim, Y. *et al.* Regulation of multiple DNA repair pathways by the Fanconi anemia protein SLX4. *Blood* **121**, 54–64 (2013).
176. Scheckenbach, K. *et al.* Treatment of the bone marrow failure in Fanconi anemia patients with danazol. *Blood Cells. Mol. Dis.* **48**, 128–131 (2012).
177. Ameziane, N. *et al.* Diagnosis of Fanconi Anemia : Mutation Analysis by Next-Generation Sequencing. *Anemia* **2012**, (2012).
178. Shimelis, H. *et al.* BRCA2 Hypomorphic Missense Variants Confer Moderate Risks of Breast Cancer. *Cancer Res.* **77**, 2789–2799 (2017).
179. Bottega, R. *et al.* Hypomorphic FANCA mutations correlate with mild mitochondrial and clinical phenotype in Fanconi anemia. *Haematologica* **103**, 417–426 (2018).
180. Casadei, S. *et al.* Characterization of splice-altering mutations in inherited predisposition to cancer. *Proc. Natl. Acad. Sci.* **116**, 26798–26807 (2019).
181. The Rockefeller University. Fanconi anemia Mutation Database. https://www2.rockefeller.edu/fanconi/genes/.

182. Richards, S. *et al.* 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**, 405–424 (2015).

183. Kopanos, C. *et al.* VarSome: the human genomic variant search engine. *Bioinformatics* **35**, 1978–1980 (2018).

184. Cote, S. *et al.* The BRCA2 c.9004G>A (E2002K) [corrected] variant is likely pathogenic and recurs in breast and/or ovarian cancer families of French Canadian descent. *Breast Cancer Res. Treat.* **131**, 333–340 (2012).

185. Rivera, B. *et al.* Functionally null RAD51D missense mutation associates strongly with ovarian carcinoma. *Cancer Res.* **77**, 4517–4529 (2017).

186. Lin, P.-H. *et al.* Multiple gene sequencing for risk assessment in patients with early-onset or familial breast cancer. *Oncotarget* **7**, 8310–8320 (2016).

187. Maxwell, K. N. *et al.* BRCA locus-specific loss of heterozygosity in germline BRCA1 and BRCA2 carriers. *Nat. Commun.* **8**, 319–319 (2017).

188. Loveday, C. *et al.* Germline RAD51C mutations confer susceptibility to ovarian cancer. *Nat. Genet.* **44**, 475–476; author reply 476 (2012).

189. Petrucelli, N., Daly, M. B. & Pal, T. BRCA1- and BRCA2-Associated Hereditary Breast and Ovarian Cancer. in *GeneReviews(®)* (eds. Adam, M. P. et al.) (University of Washington, Seattle, 1993).

190. Yang, X. *et al.* Ovarian and breast cancer risks associated with pathogenic variants in RAD51C and RAD51D. *J. Natl. Cancer Inst.* **112**, 1242–1250 (2020).

191. Taeubner, J. *et al.* Penetrance and Expressivity in Inherited Cancer Predisposing Syndromes. *Trends Cancer* **4**, 718–728 (2018).
192. Chandrasekharappa, S. C. *et al.* Assessing the spectrum of germline variation in Fanconi anemia genes among patients with head and neck carcinoma before age 50. *Cancer* **123**, 3943–3954 (2017).
193. Yao, Q. *et al.* Mutation Landscape of Homologous Recombination Repair Genes in Epithelial Ovarian Cancer in China and Its Relationship With Clinicopathlological Characteristics. *Front. Oncol.* **12**, 709645 (2022).
194. Del Valle, J. *et al.* Exploring the Role of Mutations in Fanconi Anemia Genes in Hereditary Cancer Patients. *Cancers* **12**, (2020).
195. Amaral, T. *et al.* Are Pathogenic Germline Variants in Metastatic Melanoma Associated with Resistance to Combined Immunotherapy? *Cancers* **12**, (2020).
196. Kim, B. *et al.* Prevalence and clinical implications of germline predisposition gene mutations in patients with acute myeloid leukemia. *Sci. Rep.* **10**, 14297 (2020).
197. Kitsera, N., Dorosh, O. & Makukh, H. Woman with Turner syndrome and her child with acute leukemia (a case report). *Exp. Oncol.* **42**, 333–336 (2020).
198. Tian, X. *et al.* Clear Cell Papillary Renal Cell Carcinoma Shares Distinct Molecular Characteristics and may be Significantly Associated With Higher Risk of Developing Second Primary Malignancy. *Pathol. Oncol. Res. POR* **27**, 1609809 (2021).
199. Fukushima, H. *et al.* Cancer-Predisposition Genetic Analysis in Children with Brain Tumors Treated at a Single Institution in Japan. *Oncology* **100**, 163–172 (2022).

200. Emelyanova, M. *et al.* Platinum-based chemotherapy for pancreatic cancer: impact of mutations in the homologous recombination repair and Fanconi anemia genes. *Ther. Adv. Med. Oncol.* **14**, 17588359221083050 (2022).
201. Zhou, K. *et al.* Case Report: Metagenomic Next-Generation Sequencing Can Contribute to the Diagnosis and Treatment of Disseminated Visceral Kaposi Sarcoma Following Allogeneic Haematopoietic Stem Cell Transplantation. *Front. Oncol.* **12**, 848976 (2022).
202. Liang, Y. *et al.* Whole-exome sequencing reveals a comprehensive germline mutation landscape and identifies twelve novel predisposition genes in Chinese prostate cancer patients. *PLoS Genet.* **18**, e1010373 (2022).
203. Yang, R. *et al.* The Cellular and Molecular Landscape of Synchronous Pediatric Sialoblastoma and Hepatoblastoma. *Front. Oncol.* **12**, 893206 (2022).
204. Liu, X.-P. *et al.* Development and Validation of a 9-Gene Prognostic Signature in Patients With Multiple Myeloma. *Front. Oncol.* **8**, 615 (2018).
205. Xie, S. *et al.* Identification of significant gene and pathways involved in HBV-related hepatocellular carcinoma by bioinformatics analysis. *PeerJ* **7**, e7408 (2019).
206. Yu, L., Ke, J., Du, X., Yu, Z. & Gao, D. Genetic characterization of thymoma. *Sci. Rep.* **9**, 2369 (2019).
207. Chen, W. *et al.* Bioinformatics Analysis of Prognostic miRNA Signature and Potential Critical Genes in Colon Cancer. *Front. Genet.* **11**, 478 (2020).
208. Xu, P. *et al.* Germline mutations in a DNA repair pathway are associated with familial colorectal cancer. *JCI Insight* **6**, (2021).

209. Yang, M., He, H., Peng, T., Lu, Y. & Yu, J. Identification of 9 Gene Signatures by WGCNA to Predict Prognosis for Colon Adenocarcinoma. *Comput. Intell. Neurosci.* **2022**, 8598046 (2022).

210. Han, B. *et al.* DNA methylation biomarkers for nasopharyngeal carcinoma. *PloS One* **15**, e0230524 (2020).

211. Wu, X. *et al.* A new risk stratification system of prostate cancer to identify high-risk biochemical recurrence patients. *Transl. Androl. Urol.* **9**, 2572–2586 (2020).

212. Ye, G. *et al.* miRNA-218/FANCI is associated with metastasis and poor prognosis in lung adenocarcinoma: a bioinformatics analysis. *Ann. Transl. Med.* **9**, 1298 (2021).

213. Wang, T. *et al.* Identifying the hub genes in non-small cell lung cancer by integrated bioinformatics methods and analyzing the prognostic values. *Pathol. Res. Pract.* **228**, 153654 (2021).

214. Ma, J., Cai, X., Kang, L., Chen, S. & Liu, H. Identification of novel biomarkers and candidate small-molecule drugs in cutaneous melanoma by comprehensive gene microarrays analysis. *J. Cancer* **12**, 1307–1317 (2021).

215. Sun, R., Li, S., Zhao, K., Diao, M. & Li, L. Identification of Ten Core Hub Genes as Potential Biomarkers and Treatment Target for Hepatoblastoma. *Front. Oncol.* **11**, 591507 (2021).

216. Zhong, D. *et al.* Screening of Potential Key Biomarkers for Ewing Sarcoma: Evidence from Gene Array Analysis. *Int. J. Gen. Med.* **15**, 2575–2588 (2022).

217. Wang, Z. *et al.* UBE2T is a prognostic biomarker and correlated with Th2 cell infiltrates in retinoblastoma. *Biochem. Biophys. Res. Commun.* **614**, 138–144 (2022).

218. Zheng, P. & Li, L. FANCI Cooperates with IMPDH2 to Promote Lung Adenocarcinoma Tumor Growth via a MEK/ERK/MMPs Pathway. *OncoTargets Ther.* **13**, 451–463 (2020).

219. Zhang, J. *et al.* UBE2T regulates FANCI monoubiquitination to promote NSCLC progression by activating EMT. *Oncol. Rep.* **48**, 139 (2022).

220. Liu, X., Liu, X. & Han, X. FANCI may serve as a prognostic biomarker for cervical cancer: A systematic review and meta-analysis. *Medicine (Baltimore)* **100**, e27690 (2021).

221. Fan, S. *et al.* Integrative analysis with expanded DNA methylation data reveals common key regulators and pathways in cancers. *NPJ Genomic Med.* **4**, 2 (2019).

222. Wang, E., Lenferink, A. & O'Connor-McCourt, M. Cancer systems biology: exploring cancer-associated genes on cellular networks. *Cell. Mol. Life Sci. CMLS* **64**, 1752–1762 (2007).

223. Szklarczyk, D. *et al.* The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Res.* **49**, D605–D612 (2021).

224. Olivier, M., Hollstein, M. & Hainaut, P. TP53 mutations in human cancers: origins, consequences, and clinical use. *Cold Spring Harb. Perspect. Biol.* **2**, a001008–a001008 (2010).

225. Lemonidis, K., Arkinson, C., Rennie, M. L. & Walden, H. Mechanism, specificity, and function of FANCD2-FANCI ubiquitination and deubiquitination. *FEBS J.* **289**, 4811–4829 (2022).

226. Le Page, C. *et al.* Lessons learned from understanding chemotherapy resistance in epithelial tubo-ovarian carcinoma from BRCA1and BRCA2mutation carriers. *Semin. Cancer Biol.* **77**, 110–126 (2021).

227. Jun, D. W. *et al.* Ouabain, a cardiac glycoside, inhibits the Fanconi anemia/BRCA pathway activated by DNA interstrand cross-linking agents. *PloS One* **8**, e75905 (2013).

228. Bi, J. *et al.* MTDH/AEG-1 downregulation using pristimerin-loaded nanoparticles inhibits Fanconi anemia proteins and increases sensitivity to platinum-based chemotherapy. *Gynecol. Oncol.* **155**, 349–358 (2019).

229. Osborn, M. *et al.* CRISPR/Cas9 Targeted Gene Editing and Cellular Engineering in Fanconi Anemia. *Stem Cells Dev.* **25**, 1591–1603 (2016).

230. Swuec, P. *et al.* The FA Core Complex Contains a Homo-dimeric Catalytic Module for the Symmetric Mono-ubiquitination of FANCI-FANCD2. *Cell Rep.* **18**, 611–623 (2017).

231. Thompson, E. L. *et al.* FANCI and FANCD2 have common as well as independent functions during the cellular replication stress response. *Nucleic Acids Res.* **45**, 11837–11857 (2017).

232. Joo, W. *et al.* Structure of the FANCI-FANCD2 Complex: Insights into the Fanconi Anemia DNA Repair Pathway. *Science* **333**, 312–316 (2011).

233. Wang, S., Wang, R., Peralta, C., Yaseen, A. & Pavletich, N. P. Structure of the FA core ubiquitin ligase closing the ID clamp on DNA. *Nat. Struct. Mol. Biol.* **28**, 300–309 (2021).

234. Shah, R. B. *et al.* FANCI functions as a repair/apoptosis switch in response to DNA crosslinks. *Dev. Cell* **56**, 2207-2222.e7 (2021).

235. Teng, K. *et al.* Modeling High-Grade Serous Ovarian Carcinoma Using a Combination of In Vivo Fallopian Tube Electroporation and CRISPR-Cas9-Mediated Genome Editing. *Cancer Res.* **81**, 5147–5160 (2021).

236. Dubois, E. L. *et al.* A Fanci knockout mouse model reveals common and distinct functions for FANCI and FANCD2. *Nucleic Acids Res.* **47**, 7532–7547 (2019).

237. Guitton-Sert, L., Gao, Y. & Masson, J.-Y. Animal models of Fanconi anemia: A developmental and therapeutic perspective on a multifaceted disease. *Semin. Cell Dev. Biol.* **113**, 113–131 (2021).

238. Titus, T. A. *et al.* The Fanconi anemia/BRCA gene network in zebrafish: embryonic expression and comparative genomics. *Mutat. Res.* **668**, 117–132 (2009).

239. Ramanagoudr-Bhojappa, R. *et al.* Multiplexed CRISPR/Cas9-mediated knockout of 19 Fanconi anemia pathway genes in zebrafish revealed their roles in growth, sexual development and fertility. *PLoS Genet.* **14**, e1007821 (2018).

240. Shin, U. *et al.* Large-scale generation and phenotypic characterization of zebrafish CRISPR mutants of DNA repair genes. *DNA Repair* **107**, 103173 (2021).

241. Lee, K. Y., Chung, K. Y. & Koo, H.-S. The involvement of FANCM, FANCI, and checkpoint proteins in the interstrand DNA crosslink repair pathway is conserved in C. elegans. *DNA Repair* **9**, 374–382 (2010).

242. Sugahara, R., Mon, H., Lee, J. M. & Kusakabe, T. Monoubiquitination-dependent chromatin loading of FancD2 in silkworms, a species lacking the FA core complex. *Gene* **501**, 180–187 (2012).
243. Stanley, E. C., Azzinaro, P. A., Vierra, D. A., Howlett, N. G. & Irvine, S. Q. The Simple Chordate Ciona intestinalis Has a Reduced Complement of Genes Associated with Fanconi Anemia. *Evol. Bioinforma. Online* **12**, 133–148 (2016).
244. Charlier, C. *et al.* A deletion in the bovine FANCI gene compromises fertility by causing fetal death and brachyspina. *PloS One* **7**, e43085 (2012).
245. Fang, L. *et al.* Identification of brachyspina syndrome carriers in Chinese Holstein cattle. *J. Vet. Diagn. Investig. Off. Publ. Am. Assoc. Vet. Lab. Diagn. Inc* **25**, 508–510 (2013).
246. Ruść, A. & Kamiński, S. Detection of Brachyspina carriers within Polish Holstein-Friesian bulls. *Pol. J. Vet. Sci.* **18**, 453–454 (2015).
247. Begg, C. B. On the use of familial aggregation in population-based case probands for calculating penetrance. *J. Natl. Cancer Inst.* **94**, 1221–1226 (2002).
248. Ovarian, Fallopian Tube, and Primary Peritoneal Cancer Prevention (PDQ®): Health Professional Version. in *PDQ Cancer Information Summaries* (National Cancer Institute (US), 2002).
249. Alenezi, W. M., Fierheller, C. T., Recio, N. & Tonin, P. N. Literature review of BARD1 as a cancer predisposing gene with a focus on breast and ovarian cancers. *Genes* **11**, 856–856 (2020).
250. Fong, P. C. *et al.* Inhibition of Poly(ADP-Ribose) Polymerase in Tumors from BRCA Mutation Carriers. *N. Engl. J. Med.* **361**, 123–134 (2009).

251. Audeh, M. W. *et al.* Oral poly (ADP-ribose) polymerase inhibitor olaparib in patients with BRCA1 or BRCA2 mutations and recurrent ovarian cancer: a proof-of-concept trial. *Lancet Oncol.* **376**, 245–251 (2010).

252. Kaufman, B. *et al.* Olaparib monotherapy in patients with advanced cancer and a germline BRCA1/2 mutation. *J. Clin. Oncol.* **33**, 244–250 (2015).

253. AstraZeneca. First and only PARP inhibitor, LYNPARZA® (OLAPARIB) approved as a first-line maintenance therapy treatement in BRCA-mutated advanced ovarian cancer. https://www.astrazeneca.ca/en/media/press-releases/2019/first-and-only-parp-inhibitor--lynparza---olaparib--approved-as-.html# (2022).

254. GlaxoSmithKline Inc. ZEJULA is approved in Canada for first-line maintenance treatment of women with advanced ovarian cancer. https://www.newswire.ca/news-releases/zejula-is-approved-in-canada-for-first-line-maintenance-treatment-of-women-with-advanced-ovarian-cancer-849289412.html.

255. National Library of Medicine. ClinicalTrials.gov. https://clinicaltrials.gov/ct2/results?term=parp&cond=Ovarian+Cancer&cntry=CA&Search=Apply&recrs=b&recrs=a&recrs=f&recrs=d&age_v=&gndr=&type=&rslt=.

256. Wang, W. *et al.* Integration of Gene Expression Profile Data to Verify Hub Genes of Patients with Stanford A Aortic Dissection. *BioMed Res. Int.* **2019**, 3629751 (2019).

257. Wang, J.-Y., Li, Y., Lv, Y.-Y. & Jiang, L. Screening and identification of novel candidate biomarkers of focal cortical dysplasia type II via bioinformatics analysis. *Childs Nerv. Syst. ChNS Off. J. Int. Soc. Pediatr. Neurosurg.* **38**, 953–960 (2022).

258. Gönenc, I. I. *et al.* Single-cell transcription profiles in Bloom syndrome patients link BLM deficiency with altered condensin complex expression signatures. *Hum. Mol. Genet.* **31**, 2185–2193 (2022).

259. Walker, R. M. *et al.* DNA methylation in a Scottish family multiply affected by bipolar disorder and major depressive disorder. *Clin. Epigenetics* **8**, 5 (2016).

260. Zhang, R. *et al.* Genome-wide study of immune biomarkers in cerebrospinal fluid and serum from patients with bipolar disorder and controls. *Transl. Psychiatry* **10**, 58 (2020).

261. International Schizophrenia Consortium. Rare chromosomal deletions and duplications increase risk of schizophrenia. *Nature* **455**, 237–241 (2008).

262. Stefansson, H. *et al.* Large recurrent microdeletions associated with schizophrenia. *Nature* **455**, 232–236 (2008).

263. Malhotra, D. & Sebat, J. CNVs: harbingers of a rare variant revolution in psychiatric genetics. *Cell* **148**, 1223–1241 (2012).

264. Ionita-Laza, I., McCallum, K., Xu, B. & Buxbaum, J. D. A spectral approach integrating functional genomic annotations for coding and noncoding variants. *Nat. Genet.* **48**, 214–220 (2016).

265. Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511**, 421–427 (2014).

266. Walther, A. *et al.* Genetic prognostic and predictive markers in colorectal cancer. *Nat. Rev. Cancer* **9**, 489–499 (2009).

Copyright

SPRINGER NATURE LICENSE TERMS AND CONDITIONS

Nov 25, 2022

This Agreement between Caitlin Fierheller ("You") and Springer Nature ("Springer Nature") consists of your license details and the terms and conditions provided by Springer Nature and Copyright Clearance Center.

License Number	5420360763537
License date	Nov 01, 2022
Licensed Content Publisher	Springer Nature
Licensed Content Publication	Familial Cancer
Licensed Content Title	A review of histopathological subtypes of ovarian cancer in BRCA-related French Canadian cancer families
Licensed Content Author	Patricia N. Tonin et al
Licensed Content Date	Jul 17, 2007
Type of Use	Thesis/Dissertation
Requestor type	academic/university or research institute
Format	electronic
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	1
Will you be translating?	no
Circulation/distribution	1 - 29
Author of this Springer Nature content	no
Title	A molecular genetic investigation of FANCI as a new candidate ovarian cancer predisposing gene
Institution name	McGill University
Expected presentation date	Jan 2023
Portions	Figure 1
Requestor Location	Caitlin Fierheller 1001 Boul Decarie Montreal, QC H4A 3J1 Canada Attn: Caitlin Fierheller
Total	**0.00 CAD**
Terms and Conditions	

Springer Nature Customer Service Centre GmbH
Terms and Conditions