U.PORTO

MESTRADO EM ONCOLOGIA

ONCOLOGIA MOLECULAR

The Germline and Somatic Landscape of Familial Intestinal Gastric Cancer: search for a Cause

Celina Beatriz Teixeira São José

M

2017

U.PORTO

Celina Beatriz Teixeira São José. The Germline and Somatic Ladnscape of Familial Intestinal Gastric Cancer: search for a Cause

M.ICBAS **2017**

The Germline and Somatic Landcape of Famial Intestinal Gastric Cancer: search for a Cause

Celina Beatriz Teixeira São José

INSTITUTO DE CIÊNCIAS BIOMÉDICAS ABEL SALAZAR

Celina Beatriz Teixeira São José

The Germline and Somatic Landscape of Familial Intestinal Gastric Cancer: Search for a Cause

Dissertação de Candidatura ao grau de Mestre em Oncologia submetida ao Instituto de Ciências Biomédicas de Abel Salazar da Universidade do Porto.

Orientador – Doutora Carla Oliveira

Categoria – Investigadora Sénior[1], Líder do Grupo Regulação de Expressão em Cancro[1] e Professora Afiliada[2]

Afiliação – [1]Instituto de Inovação e Investigação em Saúde / Instituto de Patologia e Imunologia Molecular da Universidade do Porto, [2]Faculdade de Medicina da Universidade do Porto,

Co-orientador – Doutora Joana Carvalho

Categoria – Investigadora pós-doutoral

Afiliação – Instituto de Inovação e Investigação em Saúde / Instituto de Patologia e Imunologia Molecular da Universidade do Porto

Agradecimentos

Durante a realização desta dissertação contei com o apoio de várias pessoas, sem as quais a sua concretização não seria possível.

À minha orientadora, Carla Oliveira, que me acolheu de braços abertos, gostaria de agradecer toda a dedicação e apoio. Não posso deixar de agradecer todos os ensinamentos para a realização da tese e um pouco mais para ter na bagagem.

À Joana Carvalho, obrigada por todo o apoio, dedicação e, principalmente, por me ajudares sempre. Obrigada por tudo que me ensinaste e por estares sempre disponível.

À Patrícia Oliveira obrigada por estares sempre pronta para me ajudar e pela paciência, tens sempre razão!

À Diana e ao Necas, agradeço a companhia ao almoço, a transcrição do que conseguimos viajar para escrito apenas diminuiria o seu verdadeiro valor.

A todo o grupo ERiC (com marca registada), um enorme obrigada por me receberem tão bem, por toda a entreajuda e dinâmica incríveis. Em especial, à minha colega de bancada, Sara Teles, a entreajuda tem um significado especial contigo. Juntas no Pentas! Não podia deixar de agradecer à Sara Rocha, porque ”um título vale mais do que mil palavras”.

Um especial agradecimento aos meus pais, por tudo que fizeram por mim e, principalmente por nunca me cortarem as asas e sempre me apoiarem. Aos meus avós, que mesmo distantes, têm sempre uma palavra de apoio e consolo.

Ao João, que tive a sorte de conhecer no início deste percurso e que teve o azar de me aturar até então. Obrigada por estares comigo nas melhores e nas piores alturas, dás-me força quando mais preciso.

À Mafalda e à Raquel viro-me e digo: aturar-me durante 10 anos é obra! Obrigada por acreditarem sempre em mim e por estarem sempre a ouvir-me, por mais longa que possa ser.

Não posso deixar de agradecer aos meus conterrâneos bioquímicos e colegas do mestrado que me acompanharam em todo o meu percurso académico, obrigada por todo o apoio. Agradeço em particular ao Molezas, Sázona, Mafalda, Tero, Nuno, Diana, Jocina, Beja, Escaca e T-tus (sem qualquer ordem de preferência!) por sempre me ouvirem e ajudarem, mas principalmente por todas as gargalhadas e aulas práticas.

Index

Figures Index

Tables Index

Abbreviations

A	Adenine
AKAP12	A-Kinase Anchoring Protein 12
Ala	Alanine
APC	Adenomatous Polyposis Coli
Arg	Arginine
ARID1A	AT-Rich Interaction Domain 1A
Asn	Asparagine
Asp	Aspartic Acid
ATM	Ataxia Telangiectasia Mutated
BMPR1A	Bone Morphogenetic Protein Receptor Type 1A
BRCA1	Breast Cancer Type 1 Susceptibility Protein
BRCA2	Breast Cancer Type 2 Susceptibility Protein
C	Cytosine
CASP10	Caspase 10
CDH1	Cadherin 1
CDKN2A	Cyclin Dependent Kinase Inhibitor 2A
CIN	Chromosomal Instability
COSMIC	Catalogue of Somatic Mutations in Cancer
CTHRC1	Collagen Triple Helix Repeat Containing 1
CTNNA1	Catenin Alpha 1
Cys	Cysteine
DNA	Deoxyribonucleic acid
EPCAM	Epithelial Cell Adhesion Molecule
ERBB2	Erb-B2 Receptor Tyrosine Kinase 2
ExAC	Exome Aggregation Consortium

FAT4	FAT Atypical Cadherin 4
FATHMM	Functional Analysis through Hidden Markov Models
FIGC	Familial Intestinal Gastric Cancer
G	Guanine
GAPPS	Gastric and Proximal Polyposis of the Stomach
GC	Gastric Cancer
Gln	Glutamine
Glu	Glutamic Acid
Gly	Glycine
HDGC	Hereditary Diffuse Gastric Cancer
His	Histidine
HR	Homologous Recombination
HSPA5	Heat Shock Protein Family A (Hsp70) Member 5
IDH2	Isocitrate Dehydrogenase (NADP(+)) 2
IGCLC	International Gastric Cancer Linkage Consortium
Ile	Isoleucine
ITIH2	Inter-Alpha-Trypsin Inhibitor Heavy Chain 2
JAK2	Janus Kinase 2
KRAS	KRAS Proto-Oncogene
Leu	Leucine
LOH	Loss of Heterozygosity
LRP1B	LDL Receptor Related Protein 1B
Lys	Lysine
MAP3K6	Mitogen-Activated Protein Kinase Kinase Kinase 6
MAPK	Mitogen-activated Protein kinase
MCCC1	Methylcrotonoyl-CoA Carboxylase 1

Met	Methionine
MLH1	MutL Homolog 1
MLH3	MutL Homolog 3
MMR	Mismatch Repair
mRNA	Messenger Ribonucleic acid
MSH2	MutS Homolog 2
MSH3	MutS Homolog 3
MSH6	MutS Homolog 6
MSI	Microsatellite Instability
MSR1	Macrophage Scavenger Receptor 1
MTUS1	Mitochondrial Tumor Suppressor Gene 1
MYH	MutY Homolog
PALB2	Partner And Localizer Of BRCA2
PCLO	Piccolo Presynaptic Cytomatrix Protein
PCP	Planar Cell Polarity
PCR	Polymerase Chain Reaction
PD-L1	Programmed death-ligand 1
PD-L2	Programmed death-ligand 2
Phe	Phenylalanine
PI3K	Phosphoinositide-3-Kinase
PIK3CA	Phosphatidylinositol-4,5-Bisphosphate 3-Kinase Catalytic Subunit Alpha
PMS1	PMS1 Homolog 1
PMS2	PMS1 Homolog 2
Pro	Proline
PTEN	Phosphatase And Tensin Homolog
RHOA	Ras Homolog Family Member A

SDHA	Succinate Dehydrogenase Complex Subunit A
SDHB	Succinate Dehydrogenase Complex Subunit B
SDHC	Succinate Dehydrogenase Complex Subunit C
SDHD	Succinate Dehydrogenase Complex Subunit D
Ser	Serine
SMAD4	SMAD Family Member 4
SNV	Single Nucleotide Variant
STK11	Serine/Threonine Kinase 11
SYNE1	Spectrin Repeat Containing Nuclear Envelope Protein 1
T	Thymine
TCGA	The Cancer Genome Atlas
TGFβRII	Transforming Growth Factor Beta Receptor 2
Thr	Threonine
TP53	Tumor Protein P53
Trp	Tryptophan
Tyr	Tyrosine
U	Uracil
Val	Valine

Abstract

Gastric cancer is the fifth most common cancer and the third cause of cancer-related mortality worldwide. The majority of the cases have sporadic nature, nevertheless, 10% display familial aggregation. These may account for at least three syndromes: Hereditary Diffuse Gastric Cancer (HDGC), Gastric Adenocarcinoma and Proximal Polyposis of the Stomach (GAPPS) and Familial Intestinal Gastric Cancer (FIGC).

FIGC has been thought of as an autosomal dominant inheritance pattern and is characterised by intestinal type adenocarcinoma without gastric polyposis, displaying common macroscopic features observed in sporadic intestinal gastric cancer. However, no underlying genetic defects have so far been described, and no genetic screening is available for FIGC families. Moreover, affected individuals generally display late onset intestinal gastric cancer and the pedigrees of FIGC families often exhibit generations without affected individuals. Underlying these characteristics may be the presence of low or moderate risk alleles that, when occurring in combination, may increase FIGC susceptibility. Therefore, the hypothesis of this master thesis was that co-occurrence of germline low or moderate risk alleles in one or more cancer-related genes may be the underlying genetic defect of FIGC. The general aim of this thesis was to identify genetic defects that could increase susceptibility to develop FIGC.

To prove this hypothesis, normal and tumour DNA from 52 FIGC probands were screened for 67 candidate gastrointestinal cancer-associated genes with Illumina's MiSeq platform, and classified using two distinct softwares: Illumina's Variant Interpreter and our own Annotator57 software. Somatic second hits, such as second mutation and promoter methylation were searched for in FIGC tumours at potentially causative genes by PCR-Sequencing. In addition, the mRNA expression of specific genes was evaluated, by quantitative real-time PCR, in order to determine the impact of germline and somatic variants found in one family.

In the 52 FIGC families fulfilling IGCLC criteria, 36 germline variants were found in 24 FIGC families, affecting 17 genes. These variants include: one Likely Pathogenic, 11 Conflicting, three Likely Benign and 21 Variants of Unknown Significance according to Annotator57. Moreover, the most frequently mutated genes were *MSH6* (17%), *SDHD* (11%), *ATM* (8%) and *MTUS1* (8%).

Interestingly, 10 out of the 24 FIGC families (42%) carried co-occurrence of germline variants, with potential impairment of specific pathways: Mismatch Repair, Homologous

Recombination, Mitogen-activated Protein kinase and Planar Cell Polarity. Of notice, DNA repair genes were frequently altered in co-occurrence with metabolism associated genes (e.g. *MSH2* and *SDHD*).

Further, only *MSH6* and *FAT4* genes were somatically inactivated, through a second mutation, in two families.

The somatic landscape revealed 115 variants, affecting 23 genes, found in 36 families. These variants were classified as: Pathogenic (12), Likely Pathogenic (42), Conflicting (1) and Variants of Unknown Significance (60). The most frequently mutated genes were: *TP53* (18%), *MSH3* (11%), *ARID1A* (10%), *FAT4* (7%), *APC* (8%), *MSH6* (7%), *ATM* (7%), *BRCA2* (6%) and *CTNNA1* (6%). In addition, the majority of families with germline variants (63%) harbored somatic variants in more than one gene, comparing with families without germline variants (35%). These results reflect the high frequency of MSI phenotype in families with germline variants.

In conclusion, the work described in this thesis pinpointed FIGC as a likely polygenic rather than a monogenic disease in 42% of families, where co-occurrence of low or moderate risk alleles that interact with family history and other non-genetic factors (environmental) can affect the risk of cancer of each individual.

Resumo

O cancro gástrico é o quinto cancro mais incidente e o terceiro cancro mais mortal no mundo. A maioria dos casos tem uma natureza esporádica, contudo, 10% demonstram agregação familiar. Estes últimos incluem, pelo menos, três síndromes: “Hereditary Diffuse Gastric Cancer”, “Gastric Adenocarcinoma and Proximal Polyposis of the Stomach” e “Familial Intestinal Gastric Cancer” (FIGC).

FIGC tem sido considerado uma síndrome com padrão de hereditariedade autossómico dominante e é caracterizado por adenocarcinoma do tipo intestinal sem polipose gástrica, revelando características macroscópicas observadas em cancro gástrico esporádico do tipo intestinal. No entanto, até agora não são conhecidas alterações genéticas causais e nenhum rastreio está em vigor para famílias com FIGC. Para além destes factos, os indivíduos afetados apresentam, geralmente, cancro gástrico tardiamente e os pedigrees das famílias com FIGC exibem gerações sem indivíduos afetados. A presença de alelos de risco baixo ou moderado podem justificar estas características quando ocorrem em combinação, podendo aumentar a suscetibilidade para desenvolver FIGC. Desta forma, o estudo aqui desenvolvido, assenta na hipótese que a co-ocorrência de alelos de risco baixo ou moderado em genes associados a cancro pode ser o defeito genético subjacente ao FIGC. Assim, o objetivo desta tese foi identificar os defeitos genéticos que poderão aumentar a suscetibilidade para desenvolver FIGC.

Para provar a nossa hipótese, DNA normal e tumoral de 52 probandos com FIGC foram sequenciados usando um painel de 67 genes, previamente associados a cancro gastrointestinal, e a plataforma MiSeq da Illumina. As variantes obtidas foram classificadas com dois softwares distintos: Illumina’s Variant Interpreter e um software desenvolvido no nosso grupo, designado Annotator57. Os tumores de probandos com FIGC foram também avaliados usando, PCR e sequenciação, para a presença de mecanismo somáticos de inactivação, tais como: segunda mutação e metilação do promotor, em genes potencialmente causadores. Adicionalmente, a expressão de mRNA de genes específicos foi avaliada por PCR quantitativo em tempo real, com o objetivo de determinar o impacto das variantes germinativas e somáticas numa família.

Em 52 famílias que cumprem os critérios IGCLC, foram encontradas 36 variantes germinativas em 24 famílias, afetando 17 genes. Estas variantes incluem: uma “Likely Pathogenic”, onze “Conflicting”, três “Likely Benign” e 21 “Variants of Unknown Significance”, de acordo com o software Annotator57. Os genes mais frequentemente alterados foram: *MSH6* (17%), *SDHD* (11%), *ATM* (8%) e *MTUS1* (8%).

De especial interesse foi o facto de 10 das 24 famílias (42%) revelarem co-ocorrência de variantes germinativas, com potencial desregulação de vias específicas: “Mismatch Repair”, “Homologous Recombination”, “Mitogen-activated Protein kinase” e “Planar Cell Polarity”. De notar, que genes de reparação de DNA estavam frequentemente alterados em co-ocorrência com genes associados a metabolismo (por exemplo *MSH2* e *SDHD*).

Além disso, observamos que apenas os genes *MSH6* e *FAT4* foram inativados somaticamente, através de uma segunda mutação, em duas famílias.

Ao nível somático, 115 variantes, foram encontradas em 36 famílias, afetando 23 genes. Estas variantes foram classificadas como: “Pathogenic” (12), “Likely Pathogenic” (42), “Conflicting” (1) e "Variants of Unknown Significance" (60). Os genes mais frequentemente mutados foram: *TP53* (18%), *MSH3* (11%), *ARID1A* (10%), *FAT4* (7%), *APC* (8%), *MSH6* (7%), *ATM* (7%), *BRCA2* (6%) e *CTNNA1* (6%). Por outro lado, observamos que a maioria das famílias com variantes germinativas (63%) continha variantes somáticas em mais de um gene, comparando com famílias sem variantes germinativas (35%). Estes resultados refletem a alta frequência do fenótipo de instabilidade de microssatélites em famílias com variantes germinativas.

Em conclusão, o trabalho descrito nesta tese revelou que o FIGC é, possivelmente, uma doença poligénica em vez de uma doença monogénica em 42% das famílias, onde a co-ocorrência de alelos de risco baixo ou moderado interagem com a história familiar e outros fatores não genéticos (ambientais), podendo afetar o risco de cancro de cada indivíduo.

I. Introduction

1. Cancer

Normal cells divide through mitosis to form new cells in a tightly controlled manner [1]. When cells acquire genetic changes that disrupt proliferation and/or apoptosis pathways, cells start to grow and divide in an uncontrolled manner, and a **tumor** may arise [2-4].

Tumors can be classified as benign and malignant [2]. **Benign tumors** are characterized by a mass of cells that have a relatively slow rate of growth and are surrounded by a capsule or adjacent tissues. This mass of cells do not invade neighboring tissues or metastasize [5]. **Malignant tumors** (or **cancer**), on the other hand, are characterized by the ability of cells to invade and infiltrate the surrounding tissues, and metastasize or travel to distant parts of the body. The growth rate of malignant tumors is erratic and atypical mitosis are commonly present [5].

Cancer can thus result from the abnormal proliferation of different types of cells, affecting different tissues, such as stomach and colon [2]. Indeed, cancer has been one of the major diseases of the 21st century. In 2012, there were 14.1 million new cancer cases, 8.2 million deaths and 32.5 million people lived with cancer within 5 years of diagnosis, worldwide [6].

1.1. Carcinogenesis

The process of transformation of a normal cell into a cancer cell is named **Carcinogenesis.** This multi-step process is characterized by the accumulation of successive genetic and epigenetic changes that lead to an imbalance in proliferation and apoptosis, providing a selective advantage to cancer cells [7].

The transformation process commonly requires the action of carcinogenic agents, which can be chemical, physical, and biological (e.g. alkylating agents [8], UV radiation [9], *Helicobacter pylori* [10], respectively [2,3]). These carcinogenic agents act by inducing DNA damage and mutations in **inductor genes** (i.e. proto-oncogenes) and **inhibitory genes** (i.e. tumor suppressor genes) [2,11].

Proto-oncogenes are normal cellular genes that participate in cell survival and proliferation pathways. These genes encode several types of proteins, such as: growth factors and their receptors (*e.g. EGFR* [12]), signal transducers (*e.g. RAS* [13]), transcription factors (*e.g. MYC* [14]) and cyclins and cyclin dependent kinases (*e.g. CDKs* [15]) [2]. After DNA damage, proto-oncogenes are activated into **oncogenes**, and promote

cellular proliferation in an autonomous way [2,3]. At least, three mechanisms have been described to explain oncogene activation [16]: point mutations that leads to increased gene products (*e.g.* G12C mutation in *K-RAS* in 40% of lung cancer cases [17]), gene amplification that results in increased gene copy number (*e.g. HER2* in breast cancer [18]) and chromosomal rearrangement in which a promoter of a gene expressed constitutively is rearranged to overproduce a normal protein, (*e.g.* Ig promoter, which is constitutively expressed, become rearranged with *MYC*, leading to its constitutive expression in Burkitt lymphoma [14] or the overproduction of a fusion protein as the case of BCR-ABL in chronic myeloid leukemia [19]). These gain-of-function alterations that convert proto-oncogenes into oncogenes are dominant [20]. Therefore, a mutation in only one of the two alleles is sufficient to alter the cellular phenotype (one hit) and induce cancer, being this gene haplosufficient [2,20].

Tumor suppressor genes encode proteins that inhibit cell survival and proliferation [11]. These proteins can function as **gatekeeper** (*e.g. TP53* [21,22]) which regulates cellular growth or apoptosis, or **caretaker** (*e.g. MSH2* [23]) ensuring DNA fidelity, and participating in DNA repair and genomic stability processes [2,3].

Tumor suppressor genes generally encode cellular growth regulation factors (*e.g. TGFβRII* in colon cancer [24,25]), cellular adhesion proteins (*e.g. CDH1* in Hereditary diffuse gastric cancer [26]), intracellular signal transducers (*e.g. APC* in colon cancer [27]), nuclear transcription factors (*e.g. STAT3* in colon cancer [28]) and proteins involved in apoptosis (*e.g. TP53* in Li-Fraumeni Syndrome [29]). Consequently, inactivation of these genes leads to uncontrolled apoptosis signaling pathways and defects in DNA damage repair [2,3]. As a result, a proto-oncogene could be activated, due to the unsuccessful repair of DNA, leading to uncontrolled proliferation and cell survival [11].

The loss-of-function alterations that inactivate Tumor suppressor genes act recessively [20]. Thus, both alleles must be affected to alter the cellular phenotype (two hits) and enable tumor progression [20].

Genomic alterations can occur at the germline and somatic levels. Germline alterations are present in the gametes, being inheritable and the underlying cause of hereditary syndromes. Somatic alterations are acquired during lifetime and are not transmitted to the offspring [2].

In the context of hereditary syndromes, a germline alteration in a tumor suppressor gene is already present in all cells of the individual. When a somatic second hit appears in the wild-type allele, the expression of the gene is decreased or completely absent, leading ultimately to the formation of a tumor [20]. This hypothesis is designated as "**Two-hit**

hypothesis", from Knudson and was discovered in Retinoblastoma, which is caused by germline mutations in *Rb* gene (the first characterized tumor suppressor gene) [20]. In fact, Hereditary Retinoblastoma commonly arises from *de novo* mutations in *Rb* and have an increased probability of suffering a second hit in the wild-type allele, such as loss of heterozygosity. Further, these cases are usually presented with bilateral retinoblastoma [20]. In the case of sporadic Retinoblastoma, a germline mutation in *Rb* gene is not present. Thus, the individuals affected have a decreased probability of suffering two hits in both eyes, being, usually, presented as unilateral retinoblastoma [20].

At least three mechanisms have been described to explain inactivation of the second, and wild-type allele at the somatic level: point mutation, loss of heterozygosity (LOH) and DNA promoter hypermethylation [16,30]. Patients that already harbor a germline mutation, may acquire a second somatic mutation in the wild-type allele, that leads to complete loss of protein function, as observed in ta fraction of tumors with from Hereditary Diffuse Gastric Cancer patients bearing germline CDH1 mutations [31]. LOH is a long deletion, resulting in partial or complete loss of a gene. In a heterozygous individual that carriers a germline mutation in one allele, LOH of the remaining wild-type allele leads to the complete loss of function. Indeed, both for somatic mutations and LOH, tumors become homozygously inactivated [32]. LOH has been identified as a major second inactivation hit in several tumors types, being an example patients with Hereditary Retinobastoma [20]. Further, promoters of genes frequently overlap CpG islands, which are regions that have a high content of CpG sites, dinucleotides of cytosine (C) and guanine (G). Methylation of cytosines at CpG sites (5-methylcytosine) inactivates the expression of genes [33], by preventing the binding of transcription factors [34]. Methylation of *CDH1* has been reported as a somatic second hit, for example in approximately 50% of the primary tumors from Hereditary Diffuse Gastric Cancer patients [35].

2. Gastric Cancer

2.1. Clinical and Histological Features

Gastric cancer (GC) is still the fifth most incident type of cancer, accounting for 7% (roughly 1 million) of all cancer cases [6], despite the recent improvements in diagnosis [36,37]. According to Globocan, in 2012, the incidence of GC was higher in particular areas of the globe, such as Eastern Asia and Central and Eastern Europe, and higher among men than women [6,38]. GC is also the third cause of cancer-related deaths, with approximately 0.7 million of deaths worldwide, in 2012 [6]. Clinically, the absence of symptoms in an early stage of the disease restricts the initial diagnosis, reduces the chance of cure by surgery, and may explain the high mortality rate [37]. In fact, the majority of cases are diagnosed at advance stages of the disease (III and IV), when symptoms and metastasis are already present, and when few therapeutic options are available [31,39-41]. These facts explain the GC patients' poor 5-year survival rate of less than 25% [41,42].

GC is a multifactorial disease with a variety of genetic and environmental factors that increase the susceptibility to develop cancer in the stomach [43]. *Helicobacter pylori* infection, smoking and diet (*e.g.* salted food and smoked meat) play an important cumulative effect as key environmental factors [36,43,44].

GC is an heterogenous disease commonly displayed as adenocarcinomas (*i.e.* a malignant neoplasia originated from epithelial tissue). According to Lauren, gastric adenocarcinomas can be histologically classified as diffuse or intestinal [45]. **Diffuse carcinomas** are more prevalent in female younger individuals and are more aggressive than intestinal carcinomas [46]. These tumors are characterized by isolated infiltrating neoplastic cells with absent or impaired expression of E-cadherin, an epithelial cell adhesion molecule [41]. In addition, these cells do not form glandular structures and may have a signet ring morphology with high mucin content [47]. **Intestinal carcinomas** tend to occur in male elderly individuals and in high risk areas [43], such as Eastern Asia and Central and Eastern Europe [48]. Intestinal-type tumours are highly associated with environmental factors, such as *Helicobacter pylori* infection [43]. According to the Correa model [49], upon *Helicobacter Pylori* infection, a multistep process occurs in which normal gastric mucosa is transformed into an invasive carcinoma. In fact, long term infection leads to an inflammation process, termed gastritis (Figure 1). Subsequently, this gastritis can persist causing atrophy of the gastric mucosa, and a gradual replacement of gastric cells by intestinal cells (Goblet cells), usually absent in the stomach, in a process termed intestinal metaplasia (Figure 1). Intestinal metaplasia may progress to dysplasia and, ultimatly to GC (Figure 1). Thus, these tumors are characterized by solid masses of well-

differentiated glandular structures with atrophic gastritis and peripheral intestinal metaplasia, in which intercellular adhesion is preserved [41].

Diffuse and Intestinal tumours can have a sporadic nature or appear in a familial context.

2.2. Sporadic Gastric Cancer

The majority of GC cases (approximatelly 90%) are sporadic, and result from the complex interplay of genetic and environmental factors (Figure 1) [50]. In fact, these factors affect different stages of the disease, from precancerous lesions to tumor initiation and progression. For instance, genetic polymorphisms in interleukin 1 beta (*IL-1B*) gene have been associated with GC [51]. In particular, individuals carrying the polymorphism IL-1B-511*T, have an increased susceptibility to the development of gastritis upon *H pylori* infection, and intestinal GC [52,53].

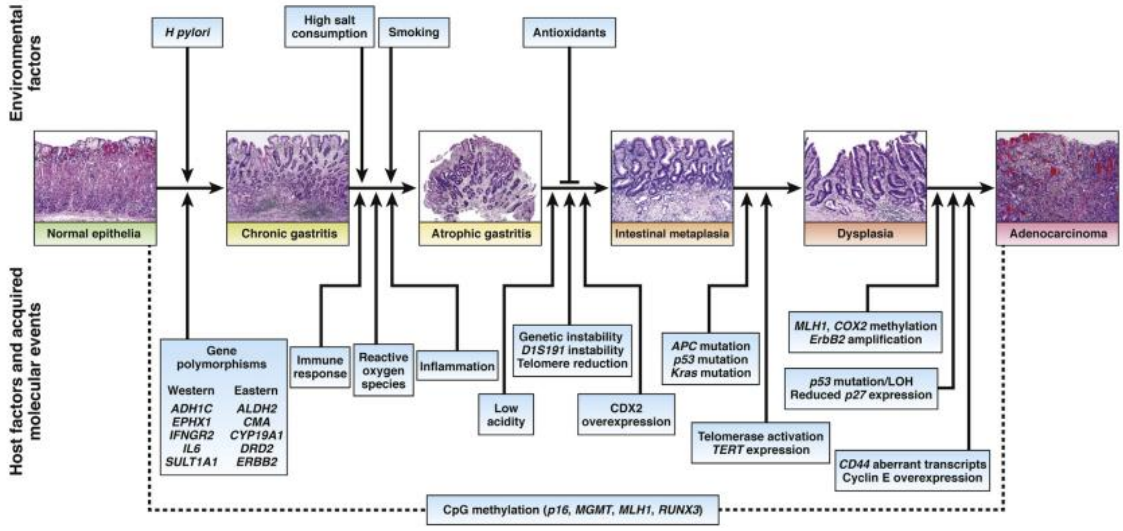


Figure 1. *Major factors that lead to intestinal GC, both environmental and molecular events, adapted from 50*

Beyond genetically determined factors, other somatic genetic alterations have been recurrently found in GC [50]. In fact, during the last decade, several studies (mostly based on next generation sequencing technologies) have been dissecting the molecular lansdcape of GC [54-56]. For instance, The Cancer Genome Atlas (TCGA) project have proposed a molecular classification of GC into four subtypes: a) tumours positive for Epstein–Barr virus, which harbor recurrent *PIK3CA* mutations, extreme DNA hypermethylation (*e.g. CDKN2A* promoter) and amplification of *JAK2*, *PD-L1*, and *PD-*

L2; b) microsatellite unstable tumours, which show an increased mutation rate and hypermethylation (including hypermethylation at the *MLH1* promoter); c) genomically stable tumours, which are mainly of the diffuse histology and highly enriched for *RHOA* mutations; and d) tumors with chromosomal instability, which show an enrichment of *TP53* mutations and focal amplification of oncogenes like *ERBB2* [(54)].

Among the recurrently found mutated genes in GC are: *TP53* (50% of intestinal-type GC cases, *ARID1A*, (10%-15% of cases), *KRAS* (approximately 20%) and *APC* (approximately 20%) [(56-58)].

MSI phenotype is observed in 15%-30% of GC cases, mainly of the intestinal-type and in older patients [(43)]. MSI tumours are characterized by alterations in genes capable of reparing deletions or insertions in nucleotide repeats (microssatelite regions), known as mismatch repair genes (MMR) – *MSH2, MSH3, MSH6, MLH1, PMS2* and others [(59)]. Mutations and promoter methylation of *MSH2* and *MLH1*, respectively, have been described in sporadic GC [(50)](Figure 1). As a consequence, tumours harbour numerous mutations in microsatellite regions, specifically in non-coding regions, both in oncogenes and tumor suppressor genes [(60)]. Genes that have microsatellite regions in the coding region, such as *TGFβRII* and *MSH3*, could likewise be mutated [(39,61)] , but hotspot coding mutations may also occur (*KRAS* and *PI3KCA*).

CIN phenotype is observed in approximately 50% of GC cases of both histological types [(62,63)]. CIN tumours are characterized by changes in chromossomal copy number, such as deletions, amplifications, LOH, and structural abnormalities [(63,64)]. Targets of CIN include oncogenic pathways, such as amplification of *HER2* [(65,66)].

Promoter hypermethylation of tumor suppressor genes, such as *CDH1* [(41)], *RUNX3*, *p16* and *MLH1* can decrease the expression of the corresponding proteins (Figure 1), being *MLH1* methylation a characteristic of MSI tumors [(50)]. In fact, methylation is an early step in GC carcinogenesis and tendes to accumulate in a multistep pattern (Figure 1). Furthermore, *H pylori* infection can induce DNA methylation in the genome of stomach cells, due to infiltration of inflamatory cells [(39,50,67)].

2.3. Familial Gastric Cancer

Familial GC is characterized by an increased risk of GC development within families and occurs in approximately 10% of all GC cases [(68)]. From these, at least three main syndromes, with primary predisposition to stomach cancer, have been described:

Hereditary Diffuse Gastric Cancer (HDGC), Gastric and Proximal Polyposis of the Stomach (GAPPS) and Familial Intestinal Gastric Cancer (FIGC) [46].

Hereditary Diffuse Gastric Cancer (HDGC) was the first hereditary gastric cancer syndrome to be recognized, and follows an autosomal dominant inheritance pattern, in which a mutation occurs in only one copy of the gene [68]. Inherited causative mutations and deletions in the E-cadherin gene (*CDH1*) explain roughly 45% of the cases, whilst 5% are explained by mutations in *CTNNA1, BRCA2, SDHB, PRSS1, PALB2, STK11, ATM* and *MSR1* [69]. The remaining 50% of HDGC cases do not have so far an identified germline cause. Nevertheless, the importance of *CDH1* gene in these cases is highlighted by the fact that 70% of HDGC individuals that do not harbor mutation in *CDH1*, display germline *CDH1* monoallelic RNA downregulation [70]. *CDH1* gene encodes for E-cadherin protein, responsible for cell to cell adhesion and important for the maintenance of epithelial architecture [71]. Somatic CDH1 promoter hypermethylation of the second allele is an early event in tumor development and leads to the inactivation of the gene in around 50% of HDGC tumors [35]. Consequently, E-cadherin expression is reduced or absent, leading to loss of cell adhesion and, consequently, increased proliferation, invasion and metastasis [71]. Lobular breast cancer is part of the tumor spectrum of HDGC families, as well as colon cancer [72]. A similarity has been found between diffuse gastric cancer and lobular breast cancer, since high mucin content with associated signet ring features and the loss of E-cadherin expression are important characteristics of both cancers [73]. In addition, *CDH1* mutation carriers have a high lifetime risk (80% in both men and women) by the age of 80 years of developing diffuse GC [68,69], and women have 60% of probability of developing lobular breast cancer [68].

To help diagnosing this syndrome, the International Gastric Cancer Linkage Consortium (IGCLC) established the following clinical criteria:

1. Two or more cases of gastric cancer, one confirmed case of diffuse gastric cancer in an individual younger than 50 years;
2. Three or more confirmed diffuse gastric cancer cases in first-degree or second-degree relatives, independent of age of onset;
3. Diffuse gastric cancer before age 40 years without a family history;
4. Personal or family history of diffuse gastric cancer and lobular breast cancer, one of which must be diagnosed before age 50 years. [44,46,68,72,74,75]

Genetic counseling is of capital importance in these families, so the individuals can make an informative decision in undertaking a genetic screening and considering the preventive measures available. The currently available genetic screening consists in

sequencing of *CDH1* and *CTNNA1* coding regions and perform multiplex ligation-dependent probe amplification to detect large CDH1 deletions [68]. When positive for *CDH1* or *CTNNA1* mutations, individuals may undergo preventive measures, as prophylactic gastrectomy, and intensive surveillance with endoscopy [68]. Endoscopy screening has, however, a poor diagnostic yield in these cases, since diffuse GC is commonly infiltrative and does not grow to the lumen of the stomach [73].

Gastric and Proximal Polyposis of the Stomach (GAPPS) is a recently identified autosomal dominant inheritance syndrome, characterized by fundic gland polyposis, with areas of dysplasia and/or intestinal-type gastric adenocarcinoma, restricted to the proximal stomach [76]. Before considering a diagnosis of GAPPS, other hereditary polyposis syndromes should be excluded, such as FAP, AFAP, MUTYH-associated polyposis and Peutz-Jeghers syndrome. GAPPS differs from FAP and AFAP, due to the tropism to the proximal stomach. Further, GAPPS differs from MUTYH-associated polyposis, an autosomal recessive inheritance syndrome, which is characterized by colorectal polyps. Additionally, whilst Peutz-Jeghers syndrome displays pigmentated gastric polyps, GAPPS is characterized by non-pigmentated polyps in the fundic region of the stomach [68].

Recently, point mutations in the promoter of *APC* were found to be the cause of GAPPS in various families [77]. However, genetic screening is not yet available for these families. Management of this disease includes endoscopic surveillance and prophylactic gastrectomy [68]. Moreover, oesophagogastroduodenoscopy and a colonoscopy should be offered to first-degree relatives of families diagnosed with GAPPS carrying the disease-associated genotype [68].

Familial Intestinal Gastric Cancer (FIGC) is generally characterized by an autosomal dominant inheritance pattern in many families with intestinal type adenocarcinomas without gastric polyposis [78]. FIGC tumors display common macroscopic features with sporadic intestinal-type GC, such as well-differentiated glandular structures with atrophic gastritis and peripheral intestinal metaplasia, with preservation of intercellular adhesion (68).

Epigenetic and/or genetic alterations in *CDH1* have been reported in FIGC tumors [41]. Particularly, 17% of FIGC tumors displayed *CDH1* promoter hypermethylation, 9.4% showed loss of heterozygosity (LOH) and 3.8% harbored in concomitance *CDH1*

promoter hypermethylation and LOH [41]. Importantly, somatic *CDH1* LOH has been claimed as a poor prognosis factor in GC, and is particularly prevalent among FIGC cases [41]. In addition, it has been demonstrated that intestinal tumors arising within FIGC families display similar frequencies of MSI phenotype comparing with sporadic intestinal GC [79,80]. Nevertheless, the molecular characterization of FIGC remains scarce.

In order to diagnose FIGC, three criteria have been described by the IGCLC, depending on the incidence of GC in each country:

a) In countries with high incidence, such as Japan and Portugal:

1. At least three relatives with intestinal GC and one should be a first degree relative of the other two;
2. At least two successive generations should be affected;
3. Gastric cancer should be diagnosed before the age of fifty in one relative.

b) In countries with low incidence, such as UK:

1. At least two first or second-degree relatives affected by intestinal GC, being one diagnosed before the age of fifty, or,
2. Three or more relatives with intestinal GC at any age [68,81-84].

Management of this syndrome is still very controversial. According to Sereno, M., et al. (2011), periodic endoscopic surveillance in first degree relatives should begin 10 years before the youngest case in the family diagnosed with intestinal GC [83]. Giovanni Corso, et al. (2013) proposed yearly endoscopic surveillance by gastroduodenoscopy beginning at the age of 40 years old, or 5 years before the youngest case diagnosed in the family. Moreover, FIGC families should be tested and eradicated for *Helicobacter pylori* and dietary habits should be taken in consideration [78].

Contrary to other hereditary gastric cancer predisposing syndromes, no major genetic risk factor has been identified for this syndrome and, therefore, no genetic screening is available for FIGC families, thus reflecting their poor management.

Understanding FIGC and its major risk factors is important to recognize families and individuals at high-risk of developing GC and refine clinical criteria, allowing an improved genetic counseling and management of patients and families.

2.4. Other Hereditary Cancer Susceptibility Syndromes Associated with Gastric Cancer Risk

Over the past few decades, several other hereditary cancer-predisposing syndromes have been described in which GC is part of their tumor spectrum, such as Peutz-Jeghers Syndrome (PJS), Lynch Syndrome, Familial Adenomatous Polyposis (FAP), Juvenile Polyposis, and others (Table 1) [81,85]. In addition, several genes have been implicated in those syndromes, like *STK11* in Peutz-Jeghers Syndrome; *APC* in Familial Adenomatous Polyposis; *MYH* in MUTYH-associated polyposis.

Peutz-Jeghers Syndrome (PJS) is an autosomal dominant inheritance syndrome, characterized by multiple polyps in the gastrointestinal tract with mucocutaneous pigmentation (95% of the cases), caused by mutations in *STK11*. Gastric cancer has been described to develop at the mean age of 30 years old in 29% of the cases [86].

Lynch Syndrome is caused by mutations in mismatch DNA repair genes (*MSH2, MSH6, MSH3, MLH1, MLH3, PMS1, PMS2*) [87]. Lynch Syndrome is a disease with an autosomal dominant inheritance pattern, characterized by the appearance of polyps (adenomas) in the colorectal region. The estimated time of transformation of adenomas to carcinomas is 1-3 years with a ration 1:1, in comparison with sporadic individuals, which is 7-13 years with a ratio 30:1 [87]. The risk of developing GC with intestinal histology was estimated to be 1-13% [81].

Familial Adenomatous Polyposis (FAP) is an autosomal dominant syndrome caused by *APC* germline mutations and is defined by the development of more than 100 polyps in the colorectum. The lifetime risk of developing colorectal cancer is approximately 100% by the age of 80 years old [88], and the risk of developing intestinal-type GC is estimated to be 2 to 4% [81].

Juvenile Polyposis Syndrome (JPS) is characterized by multiple juvenile polyps in the gastrointestinal tract with an increased risk to develop colorectal cancer [89]. JPS follows an autosomal inheritance transmission of *SMAD4* or *BMPR1A* mutations [89]. In addition, patients with JPS affected by gastric polyps have a 21% risk of developing intestinal or diffuse GC [81].

Hereditary Breast and Ovarian Cancer Syndrome follows an autosomal dominant inheritance transmission, characterized by germline mutations in *BRCA1* and *BRCA2*. Carriers of *BRCA1* or *BRCA2* mutations display a 57–65% or 45–55% risk of developing breast cancer by age 70 years, and a lifetime risk of developing ovarian cancer of 39–44% or 11–18%, respectively [90]. The lifetime risk of developing GC is 5.5% and 2.6%

for *BRCA1* and *BRCA2*, respectively. However, in families with family history of GC, the risk increases to 24% for *BRCA1* carriers and 12% for *BRCA2* carriers [81].

Table 1. *Hereditary cancer-predisposing syndromes associated with gastric cancer, the genetic cause identified, the inheritance pattern, the risk for developing GC and its histology, average age of appearance of polyps and respectively histology and initiation and intervals of endoscopic surveillance*

Syndrome	Genes	Inheritance pattern	GC Risk	Histologic type	Polyposis		Endoscopic Surveillance		Ref
					Age	Histology	Initiation	Interval	
Peutz-Jeghers Syndrome	STK11	Autosomal dominant	29%	Intestinal	16	Hamartomatous polyps	8 years	Based on the findings	81, 85, 86
Lynch Syndrome	MLH1, MSH2, MSH6, PMS2, MSH3	Autosomal dominant	13%	Intestinal	-	-	30-35 years	2-3 years	81, 85, 87
Familial Adenomatous Polyposis	APC	Autosomal dominant	2.1 – 4.2%	Intestinal	8	Hamartomatous polyps	21-30 years	3-5 years	81, 85, 88
Juvenile Polyposis	SMAD4, BMPR1A	Autosomal dominant	21%	Intestinal and Diffuse	41	Hamartomatous polyps	Midteens or when symptoms begin	3 years without polyps, 1 year with	81, 85, 89
Hereditary Breast and Ovarian Cancer Syndrome	BRCA1, BRCA2	Autosomal dominant	5.5%/24% 2.6/12%	Diffuse	-	-	-	-	81, 85, 90
MUTYH-Associated Polyposis	MYH	Autosomal recessive	Very low	-	14	Familial gastric polyposis and gastric adenoma	30-35 years	3-5 years	81, 85, 92
Carney-Stratakis Syndrome	SDHB, SDHC, SDHD	Autosomal dominant	77% by the age of 50 years	-	-	-	Age of onset: <25 years	3 years	81, 85
Cowden Syndrome	PTEN	Autosomal dominant	Not well characterized	-	-	Hamartomatous polyps with low malignancy potencial	-	-	81, 85

The inherited pattern of hereditary cancer-predisposing syndromes could be classified as Mendelian or non-Mendelian, if the disease follows the laws described by Gregor Mendel or not, respectively. Mendelian disorders are commonly monogenic and could follow a dominant or a recessive inheritance pattern [91].

Individuals with an autosomal dominant syndrome, as FAP have a mutant allele in *APC* at the germline level that increases their risk to develop the disease [88]. Later in life, these individuals may acquire a second mutant allele at the somatic level and develop cancer. The pedigrees of these families exhibit multiple successive generations affected

[91]. On the contrary, individuals with an autosomal recessive syndrome inherit mutant alleles from both parents, and both are needed to increase the susceptibility to develop cancer [91]. This way, autosomal recessive syndromes, such as MUTYH-Associated Polyposis, with germline defects in *MYH* gene, are rare [92]. In these syndromes, pedigrees tend to display skips in generations affected [91].

The inheritance pattern of hereditary cancer-predisposing syndromes could be due to alterations in high or low penetrance genes [93]. Penetrance is the percentage of individuals harboring a given mutation or genotype who exhibit an associated phenotype [93]. Meaning that in a disease with 90% of penetrance, 90% of the individuals with the mutation will develop the disease and 10% will not. In syndromes with mutations in high penetrance genes, the majority of individuals with a mutation in the associated gene will display the phenotype and develop the disease during their lifetime [93]. In fact, *MSH2* germline variants have been reported to be highly penetrant for Lynch Syndrome [87]. On the contrary, in syndromes with mutations in moderate or low penetrance genes, the number of individuals with the associated genotype that do not have the disease is higher and the penetrance is lower [94]. For example, variants in the *ATM* gene have been identified as moderate penetrance mutations in hereditary colorectal cancer [95]. Additionally, moderate or low penetrance variants can be more prevalent in the general population, and in this setting represent high-risk alleles. For example, SNP rs2981582 in *FGFR2* gene is a high risk allele for breast cancer with a population frequency of 38% [96]. In addition, the penetrance can also be influenced by more than one affected gene, that together with environmental factors may be required to install the phenotype. Without the second factor (genetic or environmental), the phenotype may be either absent or less severe [93]. In fact, hypogonadotrophic hypogonadism is derived from two different genes affected: *FGFR1* and *NELF*, that in combination lead to a more severe phenotype [97]. However, it is not clear yet for geneticists if a disease with an oligogenic inheritance pattern need both mutations to trigger the phenotype, or if the co-inheritance of two mutations is irrelevant, because only one of them is able to trigger the phenotype [98]. A quite attractive possibility is a disorder, where the gene responsible for the phenotype is co-inherited with a variant in a modifier gene that also influence the phenotype [93]. Therefore, with multiple possibilities, it is far more challenging for the geneticists to diagnose and manage multigenic diseases [99].

Next-Generation Sequencing (NGS) technologies, such as Whole-Genome Sequencing (WGS), Whole-Exome Sequencing (WES) and Multiplex Custom-panel Sequencing, have been used to dissect the genetic landscape of several diseases. In the particular case of HDGC, germline variants in *PALB2, BRCA1, RAD51C, CTNNA1, BRCA2, STK11, SDHB, PRSS1, ATM* and *MSR1* have been described in *CDH1* mutation negative carriers [69,100].

In order to determine the relative relevance of each variant for a given disease, the data obtained from NGS technologies are submitted to bioinformatics analyses and variant classification. The American College of Medical Genetics and Genomics (ACMG) have described guidelines to help the classification of variants [101]. To support the classification of variants it is fundamental to consult data deposited in databases such as UniProt, OMIM, ClinVar and COSMIC, as well as specific disease-associated databases, such as InSiGHT.

UniProt is a repository of natural variants occurring in the protein sequence, including polymorphisms, variations between strands and others, and their association with the disease in human proteins. However, mutations such as frameshifts and other premature truncating mutations, that induce major protein changes, are not annotated [102,103].

OMIM stands for Online Mendelian Inheritance in Man and displays information regarding the genotype and its association with each disease [104,105].

ClinVar exhibits submissions concerning variants found in patients, both germline and somatic, and their clinical significance [106,107].

The **Catalogue of Somatic Mutations in Cancer**, or **COSMIC**, is a repository of somatic mutations found in various types of cancer specimens, which were manually curated using literature data. Cosmic also displays a Functional Analysis through Hidden Markov Models (FATHMM) Prediction, predicting the functional consequences of coding Single Nucleotide Variants (SNVs) [108].

As one important factor to take into account in the evaluation of the pathogenicity of a variant is the frequency. **dbSNP**, **Exome Aggregation Consortium** (ExAC) and **1000Genome Project** are databases that provide the frequency of millions of variants in various populations [109-111].

The information collected by all these databases, may help the geneticists to retrieve the clinical significance of each variant found by Next-Generation Sequencing technologies.

II. Rational and Objectives

The majority of Gastric Cancer cases have a sporadic nature, however 10% display familial aggregation. These may account for at least three syndromes: Hereditary Diffuse Gastric Cancer (HDGC), Gastric Adenocarcinoma and Proximal Polyposis of the Stomach (GAPPS) and Familial Intestinal Gastric Cancer (FIGC).

Whilst germline defects at the *CDH1* and *APC* genes have been found for HDGC and GAPPS families, respectively, FIGC families remain genetically unexplained. Consequently, there is no genetic screening available for these families and their clinical management is poor. Additionally, the pedigrees of FIGC families often display generations without affected individuals, as well as, late onset intestinal gastric cancer. These characteristics may indicate that increase cancer susceptibility in FIGC is determined by the presence of low or moderate risk alleles occurring in combination.

Therefore, the hypothesis of this dissertation is that co-occurrence, in the germline, of low or moderate risk alleles in one or more cancer-related genes may be the underlying genetic defect of Familial Intestinal Gastric Cancer.

Moreover, in other hereditary cancer susceptibility syndromes, such as Lynch syndrome, Hereditary Breast and Ovarian Cancer, or Ataxia-telangiectasia, among others, there is an unneglectable risk of developing gastric cancer. Therefore, genes causing hereditary cancer susceptibility syndromes, even if only remotely associated with gastric cancer susceptibility, would be good candidates to test as potential causal genes in FIGC families.

The general objective of this dissertation was therefore to identify genetic defects that could increase susceptibility to develop Familial Intestinal Gastric Cancer. The specific aims of this dissertation were to:

1. Identify the germline landscape of 52 FIGC families, by performing Multiplex Custom-panel sequencing;
2. Classify the germline variants, using Illumina's Variant Interpreter and Annotator platforms;
3. Characterize somatic second-hits which may lead to the inactivation of genes with a pathogenic or co-occurring germline variant in FIGC tumors, using Multiplex Custom-panel Sequencing and methylation analysis;
4. Identify and classify other somatic events in FIGC tumors, using Multiplex Custom-panel sequencing and Illumina's Variant Interpreter and Annotator platforms.

III. Materials and Methods

1. Patients and samples

All 52 families diagnosed with FIGC that met at least one criteria from IGCLC [74] were admitted at the Division of General Surgery and Surgical Oncology, University of Siena, Italy. Informed consent was obtained from the 52 patients with hospital's ethics committee approval. Tissue specimens of gastric cancer and matched normal samples were analyzed in this study. DNA and RNA were extracted from tumor (histologically verified to contain a minimum of 70% to 80% of neoplastic cells) and normal frozen tissues.

2. Analysis of germline and somatic variants

2.1. Multiplex Custom-panel Sequencing

Multiplex custom-panel sequencing of 67 genes (table 2) was performed on normal and tumor DNA from 52 FIGC probands using Illumina's TruSeq Custom Amplicon assay on the MiSeq platform (Illumina). All germline and somatic variants were validated by Sanger sequencing.

Table 2. *Panel of genes used in Multiplex Custom-panel Sequencing.*

Condition	Gene	Function
Carney-Stratakis Syndrome / Paraganglioma	SDHB SDHC SDHD	Kreb's cycle, respiratory eletron transport
Cowden Syndrome	PTEN	AKT/PKB regulator
	SDHB SDHD	Kreb's cycle, respiratory eletron transport
Esophageal adenocarcinoma / Barret's esophagous	AKAP12	Activation of cAMP-Dependent PKA
	CTHRC1	Noncanonical Wnt signaling pathway, PCP
	FOXF1	Transcription factor
	MSR1	Inflammation, LDL transport
Familial Adenomatous Polyposis	APC	Wnt signaling pathway
Familial Gastric Cancer	MAP3K6	MAPK signalling pathway, regulation VEGF, Apoptosis

Table 2 (cont). *Panel of genes used in Multiplex Custom-panel Sequencing.*

Condition	Gene	Function
Gastric Cancer	ARID1A	Chromatin remodeling
	ATM	DNA double strand repair
	BCL2L10	Apoptosis
	CASP10	
	CHEK2	DNA double strand repair
	FAT4	Planar Cell Polarity
	FHIT	Purine metabolism
	HSPA5	Regulation of proteins folding and degradation of misfolded proteins
	IDH1	Kreb's cycle
	IDH2	
	PALB2	DNA double strand repair
	PSCA	Proteins metabolism
	RUNX3	Transcription factor
	TP53	Apoptosis
	ITIH2	Cell adhesion to the matrix
	MET	Cell survival, migration and invasion
	MTUS1	AT2 signalling pathway, p36 inhibition, apoptosis
Hereditary Breast and Ovarian Cancer	BRCA1	DNA double strand repair
	BRCA2	
Hereditary Diffuse Gastric Cancer	ATM	DNA double strand repair
	BRCA2	
	CDH1	Cell-cell adhesion
	CTNNA1	
	MSR1	Inflammation, LDL transport
	PALB2	DNA double strand repair
	PRSS1	Degradation of the extracellular matrix,
	SDHB	Kreb's cycle, respiratory eletron transport
	STK11	Cell metabolism, cell polarity, apoptosis and DNA damage response.
Hereditary Mixed Polyposis Syndrome	GREM1	Angiogenesis and Wnt, Hedgehog and Notch signalling pathways
	SCG5	Chaperone
	TGFBR2	Regulation of transcription of proliferating genes
Juvenile Polyposis, Pancreatic Cancer	BMPR1A	Regulation of transcription of proliferating genes
	SMAD4	
Lynch Syndrome	EPCAM	Mismatch Repair
	MLH1	
	MSH2	
	MSH3	
	MSH6	
	PMS1	
	PMS2	
MUTYH Associated Polyposis	MUTYH	Oxidative DNA damage repair
Pancreas Adenocarcinoma	AKR7A3	Detoxification of aldehydes and ketones
	CDKN2A	Cell cycle regulation
	SPINK1	Anti-trypsin activity
Peutz-Jeghers Syndrome	STK11	Regulates cell polarity, G1 arrest

Table 2 (cont). *Panel of genes used in Multiplex Custom-panel Sequencing.*

Condition	Gene	Function
Others Colaborating Projects	BAX	Apoptosis
	C12orf32	DNA double strand repair
	CFTR	Inflammation
	FBXL4	Cell-cycle regulation
	GAB2	Amplification of signal transduction growth factors, cytokines and antigen receptors
	HIC1	Growth regulator
	MCCC1	Leucine catabolism
	NAT2	Activate and deactivate arylamine and hydrazine drugs and carcinogens
	NEK1	Cel cicle regulation
	PLAU	Converts plasminogen into plasmin
	PRR5	Regulates platelet-derived growth factor (PDGF) receptor beta expression
	PXN	Actin-membrane attachment at sites of cell adhesion to the extracellular matrix
	RNF43	Negatively regulate Wnt signaling
	SCARF2	Degradation of acetylated low density lipoprotein
	SCTR	Secretin receptor
	SLC22A4	Polyspecific organic cation transporter
	TMEFF2	Tomoregulin family of transmembrane proteins
	TNFRSF10B	Apoptosis

2.2. Clinical Classification of variants

Germline and somatic variants were classified using Illumina's Variant Interpreter platform and re-classified using Annotator57 platform, which is a bioinformatics tool implemented by our group. Annotator57's classification is based on three public databases: UniProt, OMIM and ClinVar. Additionally, all variants were manually curated using COSMIC data (not used for the integrated classification), as well as the frequency data for the population "Tuscany" available within the 1000Genome project. An integrated classification was defined using the following criteria:

a) If UniProt, OMIM and ClinVar have the same information for a given variant, the integrated classification will correspond to such information, *i.e.* UniProt, OMIM and ClinVar all classify variant X as 'Likely Pathogenic', the classification of variant X will be 'Likely Pathogenic';
b) If at least one of the three databases (UniProt, OMIM or ClinVar) classify a given variant as 'Pathogenic' or 'Likely Pathogenic' or 'Likely Benign', and information is missing for other databases, the integrated classification will be 'Pathogenic', 'Likely Pathogenic' or 'Likely Benign', regardless of the frequency obtained from dbSNP or 1000Genome project;
c) If one of the three databases (UniProt, OMIM or ClinVar) classify a given variant as 'Pathogenic' or 'Likely Pathogenic' and other database classify the variant as 'Uncertain Significance' or 'Likely Benign', the integrated classification will be 'Conflicting', regardless of the frequency obtained from dbSNP or 1000Genome project;

d) If there is no information in UniProt, OMIM and ClinVar, the variant will be classified as ´Variant of Unknown Significance' (VUS).

In addition, a potential risk allele was considered whenever a specific variant was more frequent in the FIGC cohort than in Global, or Europe or Tuscany populations. For example, variant c.1256-2A>G in the *MAP3K6* gene was considered a potential risk allele because it was more frequent in the FIGC cohort than in Tuscany (4% *vs* 0.47%, respectively), despite no information was available at UniProt, OMIM, and ClinVar databases.

2.3. Splicing prediction

The impact of germline variants, located in splice regions, on the function of the protein was evaluated using splicing prediction softwares, such as: Human Splicing Finder [(112)] and NetGen2 [(113)]. These softwares displayed a score between 0 and 1 to demonstrate the probability of the splice donor or acceptor to occur in the region analyzed, revealing all splice donors and acceptors for the input sequence. A score of zero means a probability of 0% of the splicing occur in that region and a score of one displays a probability of 100% of the splicing occur.

3. Analysis of Promoter Methylation

Prediction of CpG islands in selected genes was performed using the bioinformatics tool, CpG Island Searcher, DBCAT [(114)] and sequence data retrieved from the Ensembl database v90 [(115)]. For each gene, a CpG island was defined according to the following criteria:

a) Genomic sequence length of CpG island with ≥200 bp;
b) A percentage of GC content ≥55%;
c) CpG dinucleotides observed/expected ratio ≥0.65.

In order to evaluate the methylation status of each promoter, bisulfite modification followed by PCR and Sanger sequencing were performed. Bisulfite Modification consists in the deamination of unmethylated cytosines (C) into uracils (U), whereas methylated cytosines (5mC) are protected from deamination and are not converted, thus remaining a C upon treatment. Bisulfite treated DNA is further amplified by PCR using primers specifically designed for sequences without CpG sites. After Sanger Sequencing, the U

derived from deamination of unmethylated C is read as thymine (T) and the C derived 5mC remained as a C, Figure 2 [116].

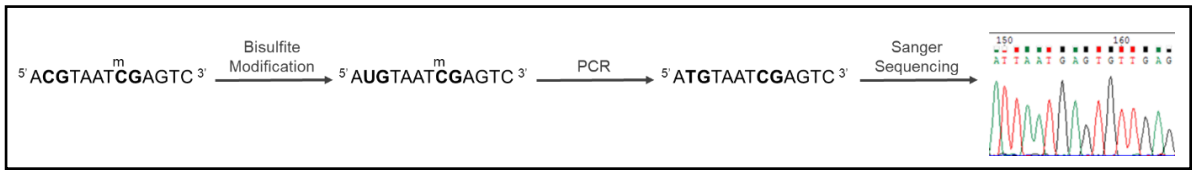


Figure 2. *Bisulfite modification principle and Sanger sequencing*

Approximately 200ng of DNA extracted from cell lines, tumor and normal tissues were treated with bisulfite, using EpiTeck Bisulfite kit (Qiagen, Hilden, Germany). Bisulfite conversion was performed using the reagents described in table 3 and the thermal cycle protocol described in table 4. The desulfonation and DNA purification was completed in an EpiTeck spin column membrane and the DNA was eluted in 20µl of UltraPure DNAse/RNAse-Free Water (Gibco, Invitrogen, Oregon, USA).

Table 3. *Bisulfite reaction components.*

Component	Volume per reaction
DNA solution	200ng
Rnase-free water	40µl - $V_{DNA\ solution}$ µl
Bisulfite Mix	85 µl
DNA Protect Buffer	15 µl
Total Volume	**140 µl**

Table 4. *Thermal cycle protocol.*

Step	Time	Temperature
Denaturation	5 min	95ºC
Incubation	25 min	60ºC
Denaturation	5 min	95ºC
Incubation	85 min	60ºC
Denaturation	5 min	95ºC
Incubation	175 min	60ºC
Hold	Indefinitive	20ºC

Promoter methylation analysis was carried out in 13 genes, by amplifying bisulfite treated DNA using primers specifically designed for sequences without CpG sites (Table 5) and the PCR program depicted in Table 6 (the annealing temperatures of each set of primers can be found in Table 5). For the genes that a CpG island was not predicted based on the abovementioned criteria, primers were designed according to studies that had previously reported methylation analysis for those genes. Primers described in table 5 were used for amplification and sequencing, with the exception of primers R4 and F5 in *FAT4*, which were only used for sequencing analysis. Positive and negative controls are described in table 5.

Table 5. *CpG island analyzed for each gene and respective primers.*

Gene	CpG island	Region Analyzed hg38	Primer		Anealing Temperatures	Size	+ Controls	- Controls
SDHB	Not found[142]	204bp upstream ATG (chr1: 17054019); 23 CpG sites	**F1**: 5'- GGGGAAGTTAAATGGGTAT -3'	**R1:** 5'- TCAACCCCACCCCTTAACC -3'	58ºC, 56ºC, 54ºC	176bp	-	Hela
BRCA2	Not found[137]	1093bp upstream ATG (chr13:32316461); 27 CpG sites	**F1:** 5'- GTTGGGATGTTTGATAAGGAAT -3'	**R1:** 5'- ATCACAAATCTATCCCCTCA -3'	56ºC, 54ºC, 52ºC	224bp	-	HCT-116
MAP3K6	Promoter	165bp upstream ATG (chr1:27366597); 74 CpG sites	**F1:** 5'- GGATTTTTTTTTGGATTTTAG -3'	**R1:** 5'- AATCCTACCAACAACTACC -3'	55ºC, 53ºC, 51ºC	205bp	-	MCF7
			F2: 5'- GTAGTTGTTGGTAGGATT -3'	**R2:** 5'- CCTCCCAAAACCCTAAATC -3'	57ºC, 55ºC, 53ºC	317bp		
	Gene body[31]	474bp downstream of the ATG (chr1:27366597); 30 CpG site	**F3:** 5'- GGTAGTTTGATTATGAGTATA -3'	**R3:** 5'- AACCCAATCCACAAAACTC -3'	56ºC, 54ºC, 52ºC	239bp	MKN74	-
MSH6	Promoter[126]	576bp upstream ATG (chr2:47783234); 59 CpG sites	**F1:** 5'- GAAGGTGAATTGTTGATTAAAG -3'	**R1:** 5'- CTAAACTCCCCTTCCCTCA -3'	60ºC, 58ºC, 56ºC	252bp	-	HCT-116
			F2: 5'- TGAGGGAAGGGGAGTTTAG -3'	**R2:** 5'- CCCAATAACCAATCAACAAAC -3'	60ºC, 58ºC, 56ºC	162bp		
MSH2	Promoter	263 base pairs upstream ATG (chr2:47403192); 76 CpG sites	**F1:** 5'- GAAAGGAGTTTTATTAAGGATG -3'	**R1:** 5'- CACACCCACTAAACTATTTCC -3'	60ºC, 58ºC, 56ºC	200bp	-	HCT-116
			F2: 5'- GGAAATAGTTTAGTGGGTGTG -3'	**R2:** 5'- CCTAAATCTTAAACACCTCC -3'	58ºC, 56ºC, 54ºC	207bp		
			F3: 5'- GGAGGTGTTTAAGATTTAGG -3'	**R3:** 5'- CAATACATTAAAATACCTAACAC -3'	58ºC, 56ºC, 54ºC	276bp		
CASP10	Promoter	4161bp upstream ATG (chr2:201185778); 18 CpG sites	**F1:** 5'- GTAGTTTAGGTTGGAGTGTG -3'	**R1:** 5'- CCATATCCTAAACATCATTC -3'	56ºC, 54ºC, 52ºC	264bp	-	HCT-116
			F2: 5'- GGTTAAGGAGGGTGGATTAT -3'	**R2:** 5'- CCCTTTCTTATATCCACATAC -3'	60ºC, 58ºC, 56ºC	188bp		
ATM	Promoter	5204bp upstream ATG (chr11:108227625); 68 CpG sites	**F1:** 5'- GGGAAAATTTTTGGTTTTAAAGG -3'	**R1:** 5'- CCTTATTATAATTCCTACTATA -3'	56ºC, 54ºC, 52ºC	102bp	-	HCT-116
			F2: 5'- TATAGTAGGAATTATAATAAGG -3'	**R2:** 5'- CTCTCACCCACCCTCTTC -3'	56ºC, 54ºC, 52ºC	244bp		
			F3: 5'- GAAGAGGGTGGGTGAGAG -3'	**R3:** 5'- CCCCTACCACTACACTC -3'	57ºC, 55ºC, 53ºC	347bp		
CTHRC1	Promoter	389bp upstream ATG (chr8:103371657); 73 CpG sites	**F1:** 5'- GTGTTTTTAAAATGTTTATTTAGG -3'	**R1:** 5'- CTCCCTCCCTTAATTCCTC -3'	60ºC, 58ºC, 56ºC	237bp	MKN74	-
			F2: 5'- GAGGAATTAAGGGAGGGAG -3'	**R2:** 5'- CCTTTTACTTCCCCTTAAA -3'	55ºC, 53ºC, 51ºC	232bp		
			F3: 5'- TTTAAGGGGAAGTAAAAGG -3'	**R3:** 5'- CTAAATTCCAAAACTCACTAC -3'	55ºC, 53ºC, 51ºC	216bp		
MTUS1	Promoter	45952bp upstream ATG (chr8:17755807); 108 CpG sites	**F1:** 5'- GTTAGAGGTAAGTAGGAG -3'	**R1:** 5'- AAAACTCCACTCCAAC -3'	49ºC, 47ºC, 45ºC	379bp	-	HCT-116
			F2: 5'- GTTGGAGTGGAGTTTT -3'	**R2:** 5'- CCCTTTACACCACTAAC -3'	49ºC, 47ºC, 45ºC	231bp		
			F3: 5'- GTTAGTGGTGTAAAGGG -3'	**R3:** 5'- CCTTTAACCCTCAACTTC -3'	52ºC, 50ºC, 48ºC	186bp		
TGFβRII	Promoter	483bp upstream ATG (chr3:30606884); 54 CpG sites	**F1:** 5'- GGAATTTTTGAGTGGTGTGG -3'	**R1:** 5'- CTCACTCAACTTCAACTCAAC -3'	60ºC, 58ºC, 56ºC	177bp	-	MCF7
			F2: 5'- GTTGAGTTGAAGTTGAGTGAG -3'	**R2:** 5'- CAAACCCCTAAACAACCCC -3'	60ºC, 58ºC, 56ºC	293bp		
FAT4	Promoter[40,150]	1605bp upstream ATG (chr4:125316412); 48 CpG sites	**F1:** 5'- GGGAGGTTGAAGTGATTAG -3'	**R1:** 5'- CAACACTCAAAAACTTTACTC -3'	58ºC, 56ºC, 54ºC	177bp	-	HCT-116
				R4: 5'- CCCCTAACTCCCAATCC -3' (seq)		121bp		
			F2: 5'- GGAGTTTTGGTTGTTGTTTG -3'	**R2:** 5'- CCTAAAATCCCCACTCCTC -3'	58ºC, 56ºC, 54ºC	216bp		
			F5: 5'- GTTGTAGGAGGGGAAGG -3' (seq)			152bp	NCI	-
			F6: 5'- GTTTAGATTATTTGGATTTAAA -3'	**R6:** 5'- CCAACAAAAATTCAACTAAC -3'	54ºC, 52ºC, 50ºC	211bp		
MSR1	Promoter	80392bp upstream ATG (chr8:16177988); 9 CpG sites	**F1:** 5'- GAGATGGAGATTTATTTTG -3'	**R1:** 5'- TAAAATAAAACTTCCAAACC -3'	52ºC, 50ºC, 48ºC	230bp	-	-
SDHD	Promoter	285bp upstream ATG (chr11:112086908); 23 CpG sites	**F1:** 5'- GTATTTGTGTAGTAAATTG -3'	**R1:** 5'- CTCAAAATCATCCACCAACCC -3'	52ºC, 50ºC, 48ºC	218bp	-	HCT-116

CpG islands analyzed in the promoter region or in the gene body; the regions analyzed were described according to ATG position, indicated in parenthesis; F: primer forward, R: primer reverse. Annealing temperatures used in the touchdown PCR. Size: amplicon product size, bp: base pairs, + controls: positive controls, - controls: negative controls.

Table 6. *PCR amplification program.*

Step	Time	Temperature	Nº of Cycles
Initial denaturation and polymerate activation	15 min	95ºC	1
Denaturation	30 seg	94ºC	
Annealing	90 seg	60ºC	3
Extension	90 seg	72ºC	
Denaturation	30 seg	94ºC	
Annealing	90 seg	58ºC	3
Extension	90 seg	72ºC	
Denaturation	30 seg	94ºC	
Annealing	90 seg	54ºC	35
Extension	90 seg	72ºC	
Final Extension	10 min	72ºC	1

After amplification, PCR products were loaded in a 2% agarose gel with 1x loading buffer and purified using two distinct methods: ExoSap it Express (Affymetrix USB, USA), with an incubation of 4 minutes at 37ºC and 1 minute at 80ºC; or using a Gel Band Purification Kit (GE Healthcare, UK). Then, the purified DNA was submitted to Sanger Sequencing using primers listed in table 5, the reagents depicted in table 7, and the thermal cycle protocol described in table 8. The presence of a T peak revealed that cytosines were unmethylated, whereas the C peak revealed the presence of methylated cytosines. If a double T and C peaks were observed, the CpG site was considered hemi-methylated. Tumors displaying more than 25% of methylated or hemi-methylated CpG sites were considered methylated and hemi-methylated, respectively.

Table 7. *Sanger Sequencing components.*

Component	Volume per reaction/µl
Big Dye Buffer	1
Big Dye	0,5
Primer F/R 10µM	0,4
DNAse/RNAse-Free Water	2,1
Purified Sample	1

Table 8*. Sanger Sequencing thermal cycle program.*

Step	Time	Temperature	Nº of Cycles
Incubate	2 min	96ºC	1
Denature	30 seg	96ºC	
Annealing	15 seg	54ºC	35
Extention	3 min	60ºC	
Final Extension	10 min	60ºC	1

4. RT-PCR and quantification of APC, CTHRC1 and β-Catenin mRNA expression

A first-strand complementary DNA (cDNA) was obtained from 1µg of total RNA from HCT-116 and MKN74 cell lines, commercial total RNAs from normal stomach and normal breast (Stratagene, Agilent Technologies, TX, USA), tumor and normal samples from the proband of family 12 using random hexamer primers (Invitrogen, Oregon, USA) and DNAse/RNAse-

free water in a final volume of 12µl. This mix was incubated at 70ºC for 10 minutes and then placed at 4ºC for 2 minutes. Then, the first-strand reaction was catalyzed by Superscript II Reverse Transcriptase (Invitrogen, Oregon, USA) at 37ºC for 60 minutes using the reagents described in table 9.

Table 9. *Components of Reverse Transcriptase Mix.*

Buffer 5x	4
DTT 0,1M	2
dNTPs	1
Rnasin	0,2
RT	0,75
H_2O	0,75
Total volume	8,7

Quantitative real time PCR was performed in triplicates for the target genes *APC, CTHRC1, β-Catenin* and *18s* RNA (endogenous control) using as probes sets: Hs, PT.56a.3539689, Hs.PT.58.39259295, Hs00355045_m1 and Hs99999901_s1 (IDT and TaqMan), respectively, and an ABI Prism 7000 Sequence Detection System. Specifically, the qRT-PCR reaction was performed using 50-100ng of cDNA and the KAPA probe fast enzyme (table 10) with a PCR program described in table 11. Data was analyzed with the comparative C_t ($2^{-\Delta Ct}$) method [(117)].

Table 10. *Components of RT-PCR.*

Component	Volume/µl
KAPA PROBE FAST qPCR Master mix	5
Rox Low	1
cDNA	1
H2O	3,5
Probe	0,5
Total Volume	10

Table 11. *Thermal cycle.*

Time	Temperature	Nº of Cycles
2m	50ºC	1
10m	95ºC	1
15s	95ºC	45
1m	60ºC	

IV. Results and Discussion

1. Germline Landscape of Familial Intestinal Gastric Cancer Families

FIGC is characterized by the development of intestinal type adenocarcinomas without gastric polyposis, segregating within families. Contrary to HDGC and GAPPS, FIGC remains genetically unexplained. Hence, no genetic risk factor has been identified for this syndrome, and no genetic screening is available for FIGC patients [(68)]. Therefore, the first two aims of this thesis were identifying and classifying germline variants that may increase susceptibility to develop FIGC.

Probands of 52 FIGC families were screened for the presence of germline variants using a Multiplex Custom-panel of 67 gastrointestinal cancer-associated genes with Illumina's MiSeq platform. Overall, 36 germline variants present in 17 genes were found in 24/52 families analyzed (Table 12). Illumina's software classified the 36 variants as: pathogenic (10 variants), likely pathogenic (3 variants), and variants of unknown significance (VUS) (23 variants), Table 12.

To refine the variant classification and improve the knowledge on VUS towards prioritization of families for further studies, the 36 variants were re-classified using a software developed in house. This software, named Annotator57, relies on criteria defined in section III, 2.2 of the Materials and Methods.

The Annotator57 software allowed answering aim 2 of this thesis, and re-classify the 36 variants as: Likely Pathogenic (1 variant), Conflicting (11 variants), Likely Benign (3 variants) and VUS (21 variants). Furthermore, this software also collected the frequency of identified variants. Indeed, most of them were very rare (<1%: 19 variants) or rare (>1% and <5%: 9 variants), and eight variants were absent from the ExAC and the 1000Genome project, and were therefore novel (Table 12). Detailed information on the classification of germline variants is described in Supplementary Table 1.

By comparing the performance of Illumina's Variant Interpreter and Annotator57 for variant classification, 4/23 variants classified as VUS by the Illumina's Variant Interpreter were reclassified into Conflicting (n=1) and Likely Benign (n=3). Annotator57 was not able to reclassify 19/23 variants that remained as VUS, due to lack of information deposited in the databases analyzed. In addition, the 10 Pathogenic variants classified with Illumina's software were re-classified as Conflicting (10 variants). Additionally, one variant was classified as Likely Pathogenic by both softwares. All these findings, suggest that Annotator57 allowed an increased knowledge, particularly on variants of unknown significance (Table 12). Furthermore, 2 variants classified as Likely Pathogenic by Illumina's

Variant Interpreter were re-classified as VUS, using Annotator57. These results may reflect an update in the databases analyzed, and highlighting the need of using constant online real time search. Further, Annotator57 discloses a classification based on clinical databases (*e.g.* ClinVar) and focused on the disorder in question, due to the use of Text Minning approaches. The frequency annotated by Annotator57 is focused on the population of the cohort, in this particular case, Tuscany. In fact, classification using Annotator 57 also relies on information retrieved from papers depicted in Pubmed and annotates using the last version of the genome, with up to date variants. Moreover, allows traceback of the classification, to confirm in which databases the information was found and which information. Therefore, the classification obtained with Annotator57 was used from this point onwards.

Working under the hypothesis that FIGC is a syndrome that does not strictly follows Mendel's laws, but rather presents as a polygenic disease with co-occurrence of low or moderate risk alleles as genetic risk factors. Under the same premise, it was assumed that variants with very rare or rare frequencies could present a cancer-predisposing effect, if co-occurring in the same patient. Thus, if a very rare variant, in a relevant gene, is classified as benign and co-occurs with other variants classified as likely pathogenic, pathogenic or even VUS, the overall impact of those variants may also be considered cancer-predisposing. In summary, for the study of susceptibility related to co-occurring variants, diverse types of variants were included, since association of two or more of these may have a cumulative effect for FIGC susceptibility.

Subsequently, the 24 FIGC families that harbored germline variants were analyzed in detail. The most frequently mutated genes were *MSH6* (17%, 6/36 variants), *SDHD* (11%, 4/36 variants), *MAP3K6* (11%, 4/36 variants), *ATM* (8%, 3/36 variants), and *MTUS1* (8%, 3/36 variants). Furthermore, 50% of the families harbored germline variants in DNA repair genes and 25% of the families carried germline variants in genes associated with metabolism. In addition, not only the same variant *(e.g* p.Val878Ala, *MSH6*) appeared in different families (F9, F10, F11, F18), but also that the same family often carried more than one variant (F1, F4, F7, F8, F11, F12, F13, F14, F16 and F19). In fact, 42% (10/24) of the families displayed co-occurrence of germline variants (Table 12 and 13). This finding supports our hypothesis and might indicate that FIGC is not a monogenic disease, but rather a polygenic disorder caused by a combination of moderate or low risk alleles [(93)]. Moreover, DNA repair genes were frequently altered in co-occurrence with metabolism associated genes (F1, F4 and F11). In fact, the germline variants in metabolism genes (*SDHB* and *SDHD*) are very rare at a global scale, but present in similar frequencies in the Tuscany population and in our FIGC cohort. Despite this fact, we reasoned that these still rare variants when in co-

occurrence with very rare variants in other important predisposing genes may still predispose to the development of FIGC, and therefore be classified as potential risk alleles.

Table 12. *Germline Variants identified in FIGC families by multiplex custom-panel based sequencing.*

Family	Gene	Chr	Position GRCh38	cDNA	Protein	Consequence	Freq cohort	Population Frequency	Illumina's Classification	Annotator57's Classification	Integrated Classification	Potential Risk Allele
1	MSH6	2	47799219	c.1236G>C	p.Lys412Asn	Missense	2%	NA	VUS	NA	VUS	Yes
	MSH6	2	47800241	c.2258C>T	p.Ser753Phe	Missense	2%	NA	VUS	Uncertain Significance - 1	VUS	Yes
	SDHD	11	112086941	c.34G>A	p.Gly12Ser	Missense	4%	Very Rare Global (1%), Rare Eur (2%), Rare Tusc (4,7%)	Pathogenic	Likely Pathogenic - 2; Uncertain Significance - 1	Conflicting	Yes
2	SDHD	11	112086945	c.38C>T	p.Ala13Val	Missense	2%	Very Rare Global (0,006%), Very Rare Eur (0%)	VUS	NA	VUS	Yes
3	MAP3K6	1	27362252	c.1256-2A>G	-	Splice acceptor	4%	Very Rare Global (0,4%), Rare Eur (1,4%), Very Rare Tusc (0,9%)	Likely Pathogenic	NA	VUS	Yes
4	MSH2	2	47429940	c.1275A>G	c.1275A>G(p.=)	Splice region, Synonymous	2%	Very Rare Global (0,02%), Very Rare Eur (0,05%)	VUS	Uncertain Significance - 1	VUS	Yes
	SDHD	11	112087953	c.149A>G	p.His50Arg	Missense	2%	Rare Global (1,3%), Rare Eur (3,4%), Rare Tusc (2,8%)	Pathogenic	Likely Pathogenic - 2; Uncertain Significance - 1	Conflicting	Yes
5	ATM	11	108329200	c.7269A>T	p.Glu2423Asp	Missense	2%	NA	VUS	Uncertain Significance - 1	VUS	Yes
6	ITIH2	10	7731989	c.1640C>T	p.Ala547Val	Missense	2%	Very Rare Global (0,02%), Very Rare Eur (0,004%)	VUS	NA	VUS	Yes
7	MAP3K6	1	27358259	c.2837C>T	p.Pro946Leu	Missense	2%	Very Rare Global (0,4%), Very Rare Eur (1%), Very Rare Tusc (0,9%)	VUS	NA	VUS	Yes
	BRCA1	17	43125268	c.-20+2dupT	-	Splice region, Intron	2%	NA	VUS	NA	VUS	Yes
8	MAP3K6	1	27362252	c.1256-2A>G	-	Splice acceptor	4%	Very Rare Global (0,4%), Rare Eur (1,4%), Very Rare Tusc (0,9%)	Likely Pathogenic	NA	VUS	Yes
	CASP10	2	201193048	c.506G>T	p.Cys169Phe	Missense	2%	Very Rare Global (0,002%), Very Rare Eur (0,02%)	VUS	NA	VUS	Yes
9	MSH6	2	47800616	c.2633T>C	p.Val878Ala	Missense	8%	Very Rare Global (0,8%), Rare Eur (2,6%), Rare Tusc (4,7%)	Pathogenic	Likely Pathogenic - 2; Likely Benign - 1	Conflicting	Yes
10	MSH6	2	47800616	c.2633T>C	p.Val878Ala	Missense	8%	Very Rare Global (0,8%), Rare Eur (2,6%), Rare Tusc (4,7%)	Pathogenic	Likely Pathogenic - 2; Likely Benign - 1	Conflicting	Yes
11	MSH6	2	47800616	c.2633T>C	p.Val878Ala	Missense	8%	Very Rare Global (0,8%), Rare Eur (2,6%), Rare Tusc (4,7%)	Pathogenic	Likely Pathogenic - 2; Likely Benign - 1	Conflicting	Yes
	ATM	11	108289671	c.4306C>T	p.His1436Tyr	Missense	2%	Very Rare Global (0,04%), Very Rare Eur (0,2%), Very Rare Tusc (0,9%)	VUS	Uncertain Significance - 1	VUS	Yes
12	CTHRC1	8	103371707	c.57_62dupCCTGCT	p.Leu20_Leu21dup	Inframe insertion	2%	NA	VUS	NA	VUS	Yes
	BRCA2	13	32337163	c.2812_2815dupGCAA	p.Thr939SerfsTer7	Frameshift Indels	2%	NA	Likely Pathogenic	NA	Likely Pathogenic	Yes
13	SDHB	1	17027802	c.487T>C	p.Ser163Pro	Missense	4%	Rare Global (1,9%), Rare Eur (3,4%), Rare Tusc (3,7%)	Pathogenic	Likely Pathogenic – 2; Uncertain Significance - 1	Conflicting	Yes
	MTUS1	8	17655975	c.2996A>G	p.Glu999Gly	Missense	2%	Very Rare Global (0,6%), Rare Eur (2,6%), Rare Tusc (3,7%)	VUS	NA	VUS	Yes
14	MTUS1	8	17755177	c.631T>G	p.Ser211Ala	Missense	2%	Very Rare Global (0,8%), Rare Eur (3,4%), Rare Tusc (3,7%)	VUS	NA	VUS	Yes
	ATM	11	108267198	c.2494C>T	p.Arg832Cys	Missense	2%	Very Rare Global (0,04%), Very Rare Eur (0%), Very Rare Tusc (0%)	VUS	Uncertain Significance - 1	VUS	Yes
15	MTUS1	8	17684434	c.2732A>C	p.Lys911Thr	Missense	2%	Very Rare Global (0,9%), Rare Eur (1,8%), Very Rare Tusc (0,9%)	VUS	Likely Benign - 1	Likely Benign	Yes
16	MSH2	2	47475052	c.1787A>G	p.Asn596Ser	Missense	2%	Very Rare Global (0,06%), Very Rare Eur (0,07%)	VUS	Likely Pathogenic – 2; Uncertain Significance - 1	Conflicting	Yes
	TGFBR2	3	30671751	c.643C>T	p.Arg215Cys	Missense	2%	Very Rare Global (0,02%), Very Rare Eur (0%)	VUS	NA	VUS	Yes
	FAT4	4	125316935	c.524G>T	p.Arg175Leu	Missense	2%	Very Rare Global (0,2%), Very Rare Eur (1%)	VUS	Likely Benign - 1	Likely Benign	Yes
17	APC	5	112839078	c.3484T>C	p.Tyr1162His	Missense	2%	NA	VUS	NA	VUS	Yes
18	MSH6	2	47800616	c.2633T>C	p.Val878Ala	Missense	8%	Very Rare Global (0,8%), Rare Eur (2,6%), Rare Tusc (4,7%)	Pathogenic	Likely Pathogenic - 2; Likely Benign - 1	Conflicting	Yes
19	MAP3K6	1	27356633	c.3481C>G	p.Pro1161Ala	Missense	2%	Very Rare Global (0,005%), Very Rare Eur (0,009%)	VUS	NA	VUS	Yes
	MSR1	8	16155085	c.877C>T	p.Arg293Ter	Stop gained	2%	Very Rare Global (1,1%), Rare Eur (2,4%), Rare Tusc (1,7%)	Pathogenic	Likely Pathogenic - 1; Uncertain Significance - 1	Conflicting	Yes
20	FAT4	4	125451863	c.10847C>T	p.Thr3616Met	Missense	2%	Very Rare Global (0,3%), Very Rare Eur (0,6%), Very Rare Tusc (0,9%)	VUS	Likely Benign - 1	Likely Benign	Yes
21	SDHD	11	112086941	c.34G>A	p.Gly12Ser	Missense	4%	Very Rare Global (1%), Rare Eur (2%), Rare Tusc (4,7%)	Pathogenic	Likely Pathogenic - 2; Uncertain Significance - 1	Conflicting	Yes
22	SDHB	1	17027802	c.487T>C	p.Ser163Pro	Missense	4%	Rare Global (1,9%), Rare Eur (3,4%), Rare Tusc (3,7%)	Pathogenic	Likely Pathogenic – 2; Uncertain Significance - 1	Conflicting	Yes
23	CDH1	16	68813323	c.1148A>G	p.Gln383Arg	Missense	2%	NA	VUS	NA	VUS	Yes
24	MSR1	8	16168606	c.482C>A	p.Thr161Asn	Missense	2%	Very Rare Global (0,08%), Very Rare Eur (0,2%) Very Rare Tusc (0,9%)	VUS	NA	VUS	Yes
25-52	Germline variants were not found											

Description of germline variants, including both frequency in the cohort (freq cohort: % of mutated families), ExAC and 1000genomes and Clinical Significance classification from Illumina's Variant Interpreter, Annotator57 (in 3 databases) and integrated classification. Eur: Europe; Tusc: Tuscany; VUS: Variant of Unknown Significance; NA: Not available; Very Rare: <1%; Rare: >1, <5; Not Rare: >5.

Table 13. *Characterization of germline variants found in each family.*

Gene	Gene Freq	Alteration	Alt Freq	F1*	F2	F3	F4*	F5	F6	F7*	F8*	F9	F10	F11*	F12*	F13*	F14*	F15	F16*	F17	F18	F19*	F20	F21	F22	F23	F24
MSH6	17%	p.Lys412Asn	3%																								
		p.Val878Ala	11%																								
		p.Ser753Phe	3%																								
SDHD	11%	p.Gly12Ser	6%																								
		p.His50Arg	3%																								
		p.Ala13Val	3%																								
ATM	8%	p.Glu2423Asp	3%																								
		p.His1436Tyr	3%																								
		p.Arg832Cys	3%																								
MSH2	6%	c.1275A>G(p.=)	3%																								
		p.Asn596Ser	3%																								
TGFBR2	3%	p.Arg215Cys	3%																								
FAT4	6%	p.Arg175Leu	3%																								
		p.Thr3616Met	3%																								
MAP3K6	11%	c.1256-2A>G	6%																								
		p.Pro946Leu	3%																								
		p.Pro1161Ala	3%																								
BRCA1	3%	c.-20+2dupT	3%																								
CASP10	3%	p.Cys169Phe	3%																								
MSR1	6%	p.Arg293Ter	3%																								
		p.Thr161Asn	3%																								
MTUS1	8%	p.Lys911Thr	3%																								
		p.Glu999Gly	3%																								
		p.Ser211Ala	3%																								
SDHB	6%	p.Ser163Pro	6%																								
BRCA2	3%	p.Thr939SerfsTer7	3%																								
CTHRC1	3%	p.Leu20_Leu21dup	3%																								
ITIH2	3%	p.Ala547Val	3%																								
APC	3%	p.Tyr1162His	3%																								
CDH1	3%	p.Gln383Arg	3%																								

Gene Freq: Frequency of each gene's variants in a total of 36 variants; Alt Freq: Frequency of each variant in a total of 36; Blue: Likely Pathogenic variant; Red: Conflicting Variant; Green: Likely Benign variant; Grey: Variant of Unknown Significance. * represents families with co-occurrence of germline variants.

2. Integration of Germline and Somatic Events of Familial Intestinal Gastric Cancer Families

Co-occurrence of variants in genes with low or moderate risk might explain some of the characteristics observed in FIGC families, such as late onset of gastric cancer and skipping of affected generations [93]. These findings led us to prioritize the 10 FIGC families displaying co-occurrence of germline variants (F1, F4, F7, F8, F11, F12, F13, F14, F16 and F19), for further studies.

Therefore, the somatic second hits (aim 3) was characterized at potentially causative genes that may lead to their inactivation in FIGC tumors, by integrating the germline and somatic sequencing data of each family and by performing promoter methylation analysis.

Detailed characterization of Families displaying co-occurrence of Germline Variants

2.1. Family 1

Family 1 encloses a female proband (F1) of 78 years old with intestinal-type GC and vaginal cancer. By analyzing the pedigree of this family, it is observed that the brother of the proband had gastric cancer at the age of 71 (Figure 3, Panel A).

2.1.1. Germline landscape

F1 had three germline variants, two in the *MSH6* gene, which belongs to the Mismatch Repair (MMR) pathway, and one in the *SDHD* gene that participates in the respiratory chain and Krebs cycle (Table 12).

The two variants found in *MSH6* gene were classified as VUS and their frequency at ExAC and 1000Genomes project is currently unknown. These variants, which are novel, were unique in this FIGC cohort (Table 12). Variant p.Lys412Asn is located in the MutSdomain I and variant p.Ser753Phe is located in the MutS domain III (Figure 3, Panel B). Since the MutSdomain I is essential for *MSH2* binding and proper repair of mismatch errors in the DNA [118], the variant at this domain may hamper the binding of *MSH2* and inhibit the function of both *MSH6* and *MSH2* proteins. MutS domain III is constituted of two subdomains that, alongside domain IV, suffer a large conformational change for the DNA to bind to domain I [119]. It is possible that a single amino acid change at this position of the protein interferes with this conformational change, but that remains to be proved. The two germline variants found in *MSH6* gene may increase FIGC susceptibility by: 1) be present

in different alleles and both contribute to increased susceptibility; 2) be present in the same allele and only one of them contributes to increased susceptibility; 3) be present in the same allele and both contribute to increased susceptibility; 4) be present in different alleles and still only one of them contributes to increased susceptibility.

The variant found in the *SDHD* gene (p.Gly12Ser) was classified as conflicting, since classifications as likely benign, likely pathogenic and uncertain significance were found in the three databases analyzed (Table 12). In addition, this variant was found to be very rare in European population (0.9%) and slightly less rare in Tuscany population (2.34%). Further, this variant is not unique in this cohort since it was also present in F21, being enriched in this FIGC cohort (Table 12). The variant p.Gly12Ser was located in the transit peptide mitochondrion domain, which is responsible for the protein transport to the mitochondria and might impair the function of the protein due to the absent transport to the mitochondria and respective anchoring (Figure 3, Panel C).

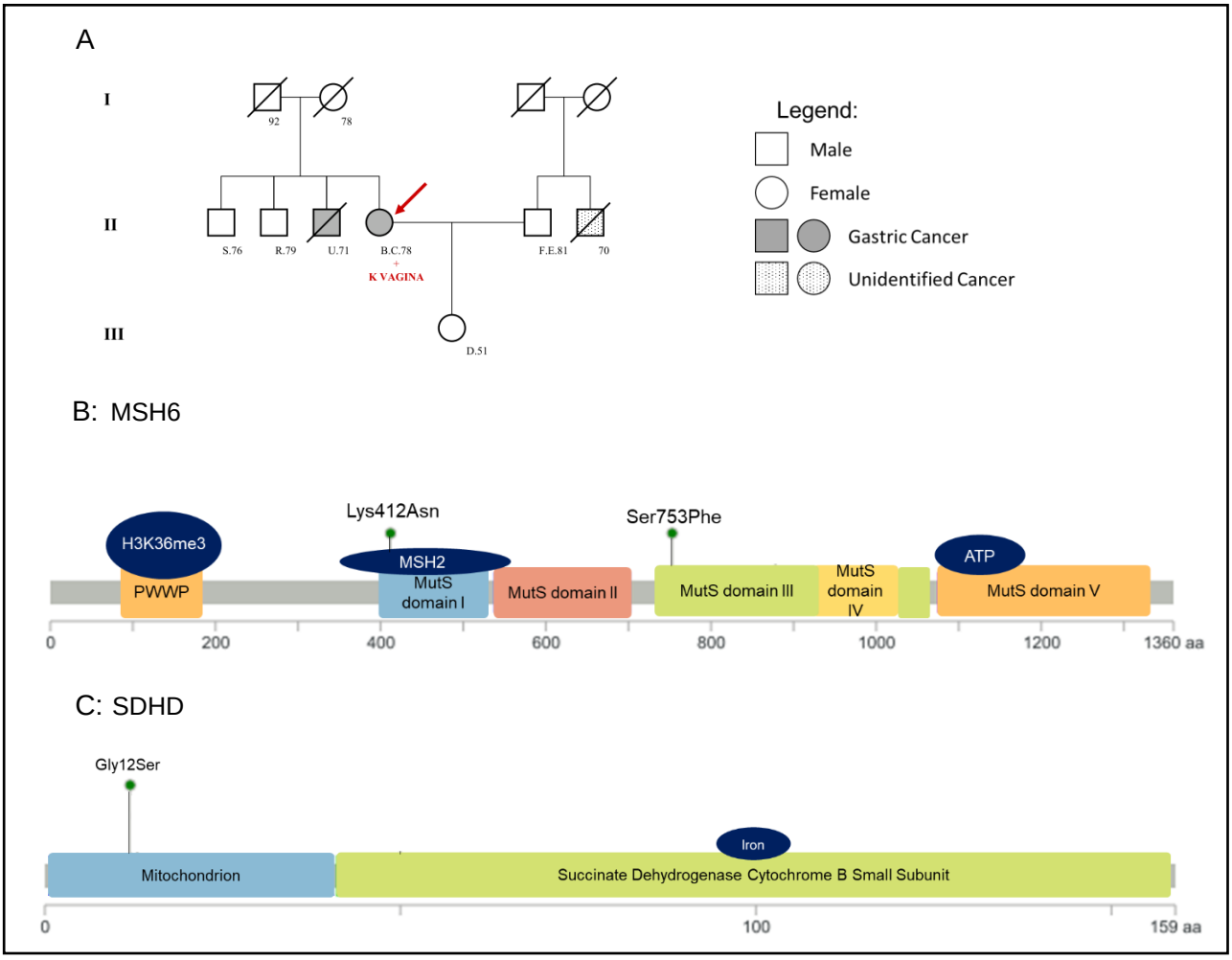


Figure 3. *Pedigree and germline variants of Family 1. A: Pedigree of family 1. The proband is displayed by an arrow. B: Schematic representation of MSH6 protein, its domains, interacting molecules and germline variants found represented by green dots (missense variants). aa: amino acids. [57,58] C: Schematic representation of*

SDHD protein, its domains, interacting molecules and germline variant found represented by a green dot (missense variants) [57, 58]

2.1.2 Somatic Landscape

The sequencing data of F1 tumor revealed eight variants in genes *ARID1A, MSH6, FAT4, AKAP12, ATM, BRCA2, IDH2* and *PALB2* (Figure 4, Panel A). In addition, previous analysis of F1 tumor have also identified a *KRAS* mutation (G13D), hypermethylation of *MLH1* promoter and microsatellite instability (MSI) phenotype.

The *MSH6* germline variant located in the domain of interaction with *MSH2* could explain the MSI phenotype, which is usually driven by *MSH2* inactivation [(120)]. Interestingly, the somatic variants presented at the *AKAP12, ATM, BRCA2, IDH2* and *PALB2*, might be a consequence of the MSI phenotype, due to their location in microsatellite regions. Of notice, the *MLH1* promoter was found to be methylated and *TP53* variants were not found at the somatic level. These findings are also characteristic of the MSI phenotype [(59)].

The Krebs Cycle pathway might be also impaired, since F1 harbored germline and somatic variants in *SDHD* and *IDH2* genes, respectively [(121)]. These findings may indicate that the occurrence of variants is important for the formation of the tumor, since *SDHD* and *IDH2* belong to the same pathway.

Similarly, *ATM, BRCA2* and *PALB2* genes participate in the same pathway - Homologous Recombination (HR) [(122)]. Therefore, somatic variants found in these genes may create an imbalance of repair of double strand breaks in the DNA molecule. Recognition of these errors might not occur, due to the impairment of *ATM*, as well as, the absent formation of the complex *BRCA2-PALB2*, important for the binding of DNA [(122)].

2.1.3 Characterization of genetic and epigenetic second hits

Family F1 harbored three germline variants, two in *MSH6* and one in *SDHD* (Figure 3). To understand whether these genes could be further inactivated at the somatic level, supporting their role as potential causative genes in this family, analysis of their coding sequence for somatic variants was performed, as well as, their CpG islands for traces of cancer-associated promoter methylation.

A missense somatic variant in *MSH6* (p.Ala1055Thr) was found in the tumor of F1 proband (Figure 4, Panel A; Table 15). This variant was located in MutS domain III, downstream of the MutS domain IV (Figure 4, Panel B) and was very rare globally (0,0008%). According

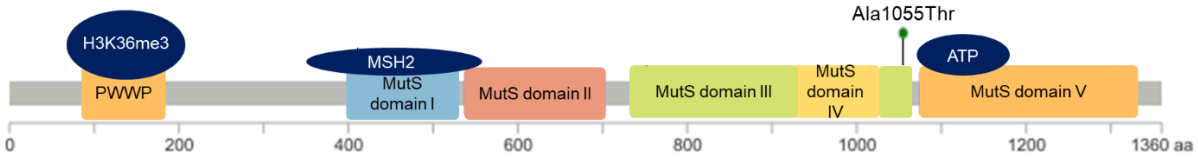
to the InSiGHT database [123,124], this variant has been found in a Lynch Syndrome patient, at the germline level, and was classified as a Class III variant (uncertain), due to insufficient evidence. Additionally, this variant has not been annotated nor curated in COSMIC database. Further, variants in the same region have been classified as likely pathogenic, in the particular case of codon 1055 [125]. Thus, p.Ala1055Thr in *MSH6* may lead to the somatic inactivation of the wild-type allele of *MSH6*. *SDHD* did not show the presence of somatic mutations in the tumor from F1.

In parallel, the methylation status of the *MSH6* and *SDHD* promoters was analyzed. Two sets of primers were designed to assess the CpG island of *MSH6*, according to Goodfellow, P. J. et al. [126]. As shown in figure 4, Panel C, the CpG sites analyzed (1 to 59) of *MSH6* promoter were neither methylated in F1 tumor nor in HCT-116 cell line, used as negative control. A set of primers was used to screen the CpG island of *SDHD* promoter. Similarly to *MSH6*, the CpG sites of *SDHD* that were possible to analyze in F1 tumor (5 to 17) were not methylated, mimicking the result obtained for the negative control, HCT-116 (Figure 4, Panel D).

A

Gene	Chr	Position GRCh38	cDNA	Protein	Consequence	Freq cohort
ARID1A	1	26771139	c.3219G>A	p.Trp1073Ter	Stop gained	1
MSH6	2	47801146	c.3163G>A	p.Ala1055Thr	Missense	1
FAT4	4	125406935	c.5363G>A	p.Arg1788His	Missense	1
AKAP12	6	151351864	c.3479dupC	p.Asp1161Ter	Frameshift Indels	1
ATM	11	108229266	c.278delA	p.Lys93ArgfsTer23	Frameshift Indels	1
BRCA2	13	32339422	c.5073delA	p.Lys1691AsnfsTer15	Frameshift Indels	1
IDH2	15	90088686	c.435delG	p.Thr146LeufsTer15	Frameshift Indels	1
PALB2	16	23635707	c.839delA	p.Asn280ThrfsTer8	Frameshift Indels	1

B: MSH6



C: MSH6

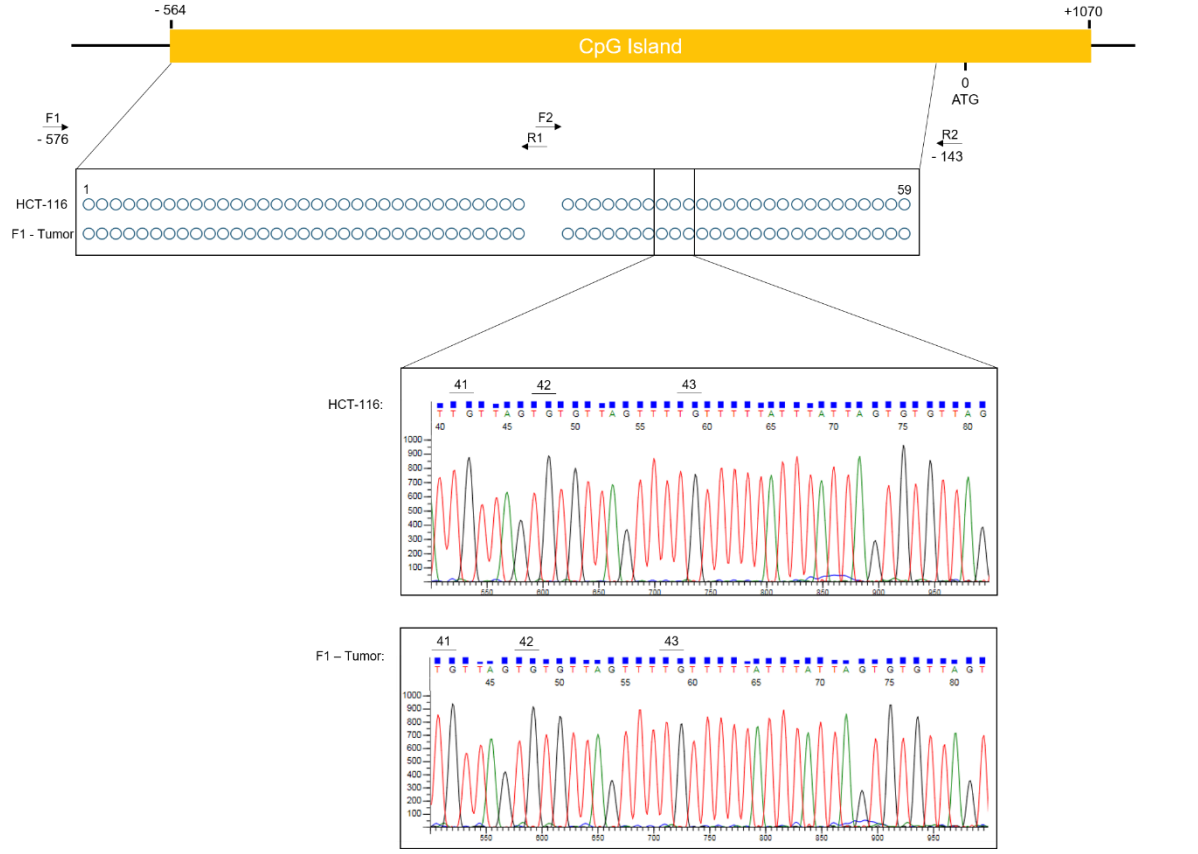


Figure 4. *Somatic Events in Family 1. A: Somatic variants found in family 1 by multiplex custom-panel based sequencing. B: Schematic representation of MSH6 protein, its domains, interacting molecules and somatic variants found represented by green dots (missense variants). aa: amino acids [57,58]. C: MSH6 Promoter Methylation Analysis. Schematic representation of the CpG island located at the MSH6 promoter. Open circles represent a non-methylated CpG; Representative examples of sequences of each sample analyzed: HCT-116 cell line (negative control) and F1 tumor. D: SDHD Promoter Methylation Analysis. Schematic representation of the CpG island located at the SDHD promoter. Open circles represent a non-methylated CpG and yellow circles represent not accessed CpG sites; Representative examples of sequences of each sample analyzed: HCT-116 cell line (negative control) and F1 tumor.*

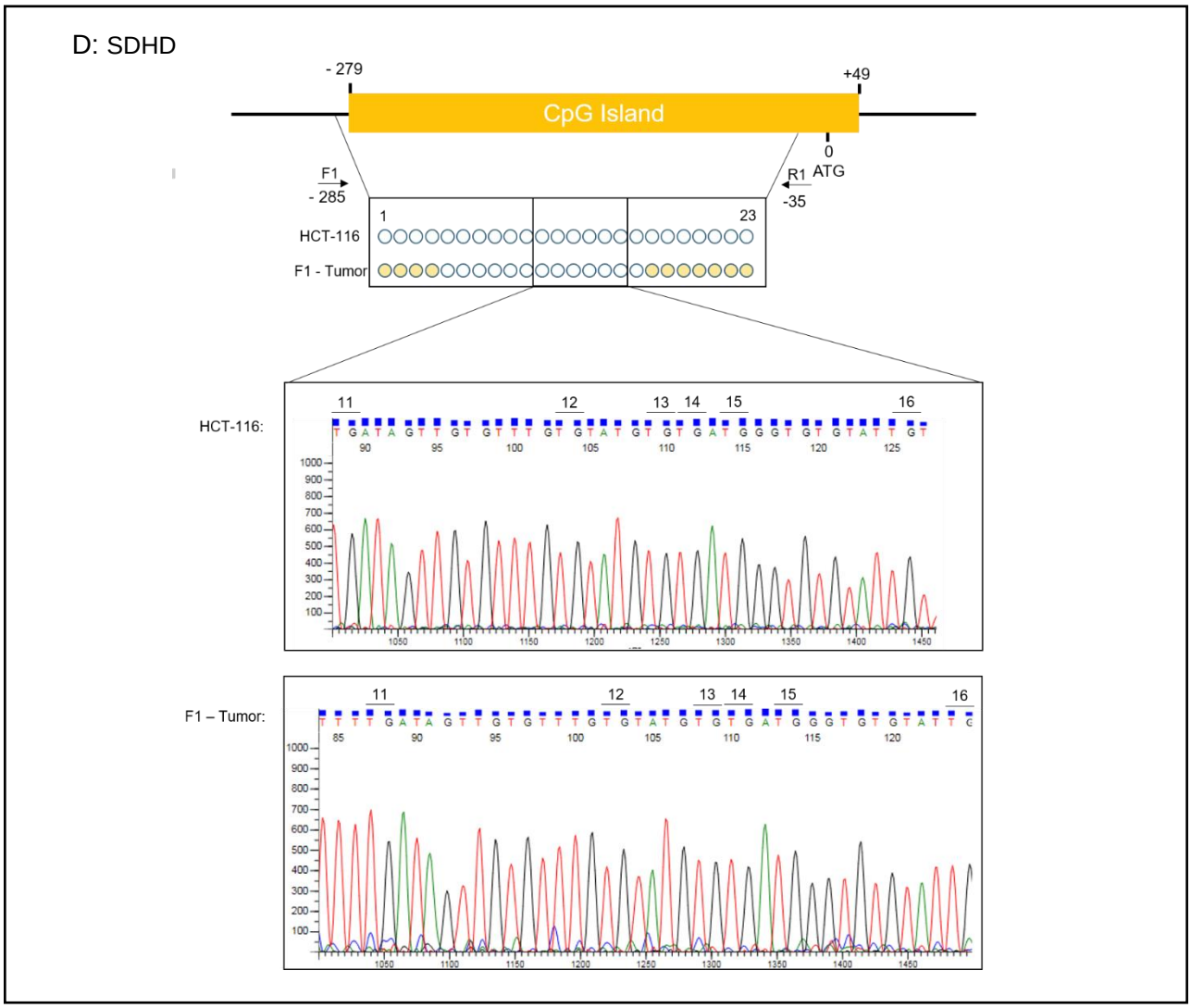


Figure 5 (cont.). *Somatic Events in Family 1. D: SDHD Promoter Methylation Analysis. Schematic representation of the CpG island located at the SDHD promoter. Open circles represent a non-methylated CpG and yellow circles represent not accessed CpG sites; Representative examples of sequences of each sample analyzed: HCT-116 cell line (negative control) and F1 tumor.*

In summary, F1 harbored two novel germline variants in *MSH6* classified as VUS, and a third *MSH6* somatic variant also very rare, classified as VUS. These mutations may or not occur in the same allele, however their co-existence supports that *MSH6* may be inactivated in this family, likely contributing to the disease phenotype. According to the International Society for Gastrointestinal Hereditary Tumors (InSiGHT) database [123,124] and TCGA dataset of sporadic gastric adenocarcinomas [54,57,58], mutations in HNPCC are found throughout the *MSH6* gene. In fact, both germline variants in *MSH6* were not found in Lynch Syndrome patients (InSiGHT database [123,124]). Thus, these variants may be low or moderate risk alleles for FIGC, and likely not be sufficiently deleterious to cause Lynch Syndrome. Moreover, these alleles may contribute to increase the risk to develop FIGC when occurring in concomitance with other variants (in this particular case, variants at

SDHD, which pathway seems also to be impaired in the tumor due to presence of a somatic mutation in IDH2).

2.2. Family 4

Family 4 was represented by a male proband (F4) of 78 years old with intestinal-type GC. The pedigree showed that his sister had lung cancer at 84 years old and his son had kidney cancer at the age of 58. In addition, it is observed that the proband's cousin had GC at 84 years old and the grand daughter of his cousin had GC at the age of 60 (Figure 5, Panel A).

2.2.1. Germline Events

F4 had two germline variants in *SDHD* and *MSH2* genes (Table 12).

The variant found in *SDHD* gene (p.His50Arg) was rare both in Europe and in Tuscany populations (1.7% and 1.4%, respectively) and unique in the cohort (Table 12). This variant was enriched in this cohort and classified as conflicting, due to contradictory classifications (likely benign, likely pathogenic and uncertain significance) depicted in the three databases analyzed (Table 12). Further, it was located in the Succinate Dehydrogenase Cytochrome B Small Subunit domain (Figure 5, Panel B), membrane-anchoring subunit which is responsible for transferring electrons from succinate to ubiquinone and involved in complex II of the electron transport chain [127,128].

The variant found in *MSH2* gene, another member of MMR pathway, was very rare globally (0.009%) and unique in the cohort (2%), being enriched. Synonymous variant p.Glu424Glu was classified as VUS (Table 12) by Annotatorr57 and as a class III uncertain significance variant, due to insufficient evidence in the InSiGHT database [123,124]. Furthermore, this variant was located in the splice region of exon 7 and 8 (Table 12). Therefore, *in-silico* tools (Human Splicing Finder and NetGene2) were used to better understand the effect of this variant on splicing defects. The wild-type codon AAG was predicted to have a score of 0.83 (*i.e* a 83% probability of the splice donor to occur at this codon) as a splice donor (Figure 5, Panel D). In fact, in the mutant sequence exon 7 displays the same amino acid (Glu) as the Wild-type sequence, however the splice donor in the mutant allele occurs 47bp upstream in comparison with the wild-type allele, with a score of 0.5. Thus, the mutant allele has a decrease in 47bp in exon 7, which have an impact in the frame and could be deleterious, due to a premature stop codon, for example.

This variant was located in MutS domain III, which allows the heterodimer complex formation MSH2/MSH6 and MSH2/MSH3 (Figure 5, Panel C) [118] and may impact the heterodimer formation and, consequently, the repair in microsatellite regions.

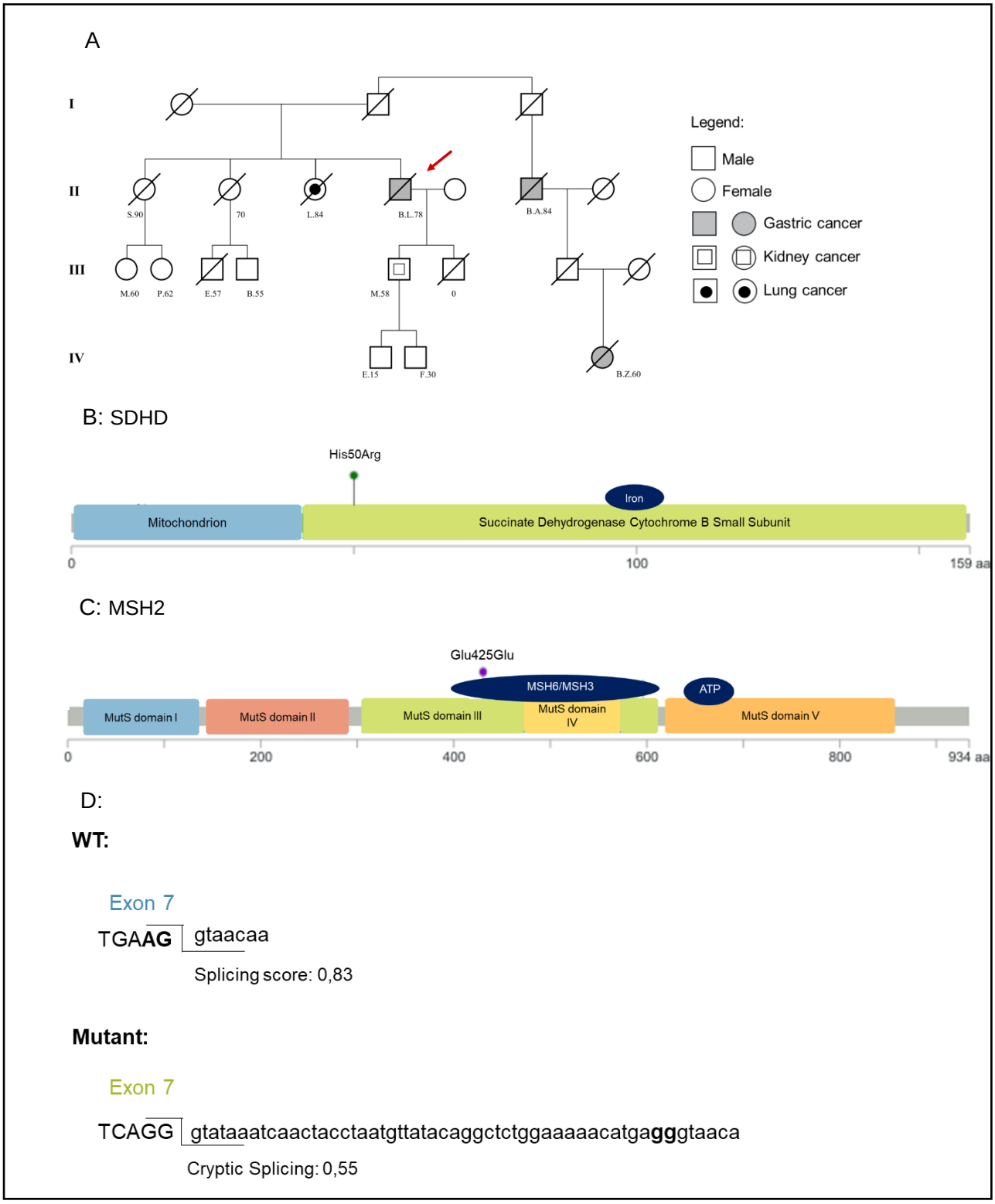


Figure 6. *Pedigree and germline variants of Family 4. A: Pedigree of family 4. The proband is displayed by an arrow. B: Schematic representation of SDHD protein, its domains, interacting molecules and germline variant found represented by a green dot (missense variants). aa: amino acids [57,58]. C: Schematic representation of MSH2 protein, its domains, interacting molecules and germline variant found represented by a purple dot*

(splicing variants). D: Splicing prediction of variant Glu425Glu in MSH2. In bold is represented the variant alteration [57,58].

2.2.2 Somatic Landscape

Somatic variants were not found in the panel of genes analyzed. However, previous analysis of this family, revealed that the tumor is microsatellite stable (MSS) and has loss of heterozygosity (LOH) at the *CDH1* gene, which may lead to poor survival [(41)].

2.2.3. Characterization of genetic and epigenetic second hits

There were no somatic mutations affecting *SDHD* and *MSH2*. Therefore, methylation status of their promoters was assessed to check if promoter methylation could be the somatic second inactivation hit.

Using the same strategy as in F1 family, it was observed that CpG island of *SDHD* was not methylated in F4 tumor (CpG sites 5 to 18) (Figure 6, Panel A).

Three sets of primers were designed to assess the methylation status of the *MSH2* CpG island. Figure 6, Panel B showed that the *MSH2* CpG island was not methylated in the negative control (HCT-116, CpG sites 1 to 73), as well as, in the F4 tumor, CpG sites 1 to 44 and 52 to 73.

By combining the germline and somatic data, it can be inferred that the co-occurrence of germline variants at the *MSH2* and *SDHD* genes was necessary to display the disease. In fact, the observed MSS phenotype of F4 tumor was also contradictory to a *MSH2* driven tumor [(120)], reinforcing that this variant might not be sufficient to cause Lynch Syndrome, but may be important if co-occurring with a *SDHD* germline variant in the context of FIGC.

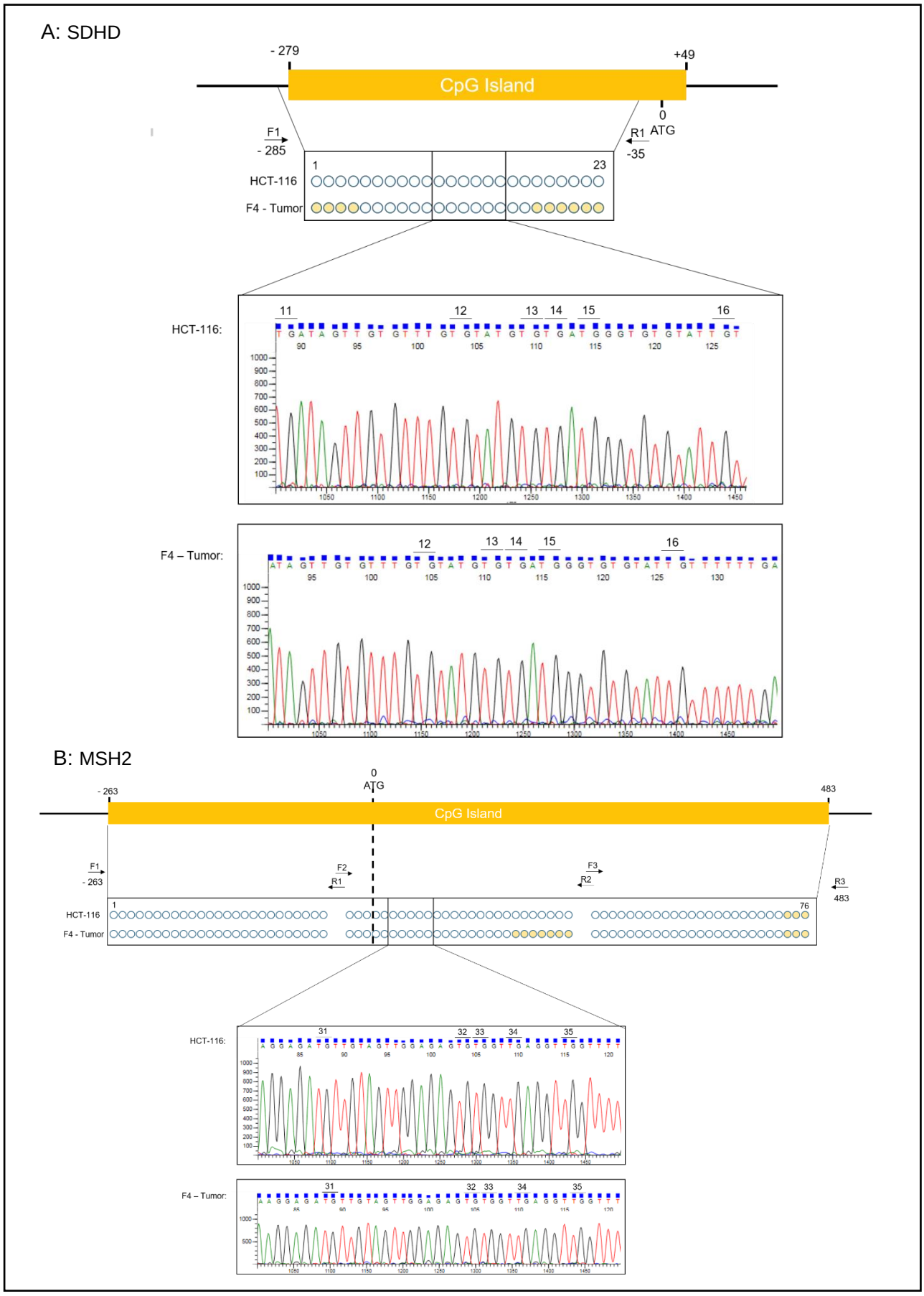


Figure 7. *Somatic Second Hit in Family 4. A: SDHD Promoter Methylation Analysis. Schematic representation of the CpG island located at the SDHD promoter. Open circles represent a non-methylated CpG and yellow*

circles represent not accessed CpG sites; Representative examples of sequences of each sample analyzed: HCT-116 cell line (negative control) and F4 tumor. B: MSH2 Promoter Methylation Analysis. Schematic representation of the CpG island located at the MSH2 promoter. Open circles represent a non-methylated CpG and yellow circles represent not accessed CpG sites; Representative examples of sequences of each sample analyzed: HCT-116 cell line (negative control) and F4 tumor.

2.3. Family 7

Family 7 was represented by a 66 years old male proband (F7) with intestinal-type GC. By analyzing the pedigree of family 7, it was observed that both grandparents of F7 were diagnosed with GC at 80 years old. From the 6 children that these individuals had, 5 had GC, including the mother of the proband, at 88 years old, two uncles at 70 and 60 years old and two aunts at 44 and 80 years old. Four cousins of the proband were also affected: one had leukemia at 22 years old, other had breast cancer at 24 years old, other had prostate cancer at 60 and a cousin with bladder cancer. A second degree cousin of the proband had brain cancer at 40 years old. The father of the proband had GC at 45 years old and his brother had an unidentified cancer at 80 years old, as well as, a proband's cousin at 40 years old. Two second degree uncles were diagnosed with GC at 79 years old and prostate cancer at 73 years old, respectively, and a second degree aunt was diagnosed with Gynaecological cancer. Two third degree cousins were affected, one was diagnosed with breast cancer and the other was diagnosed with osteosarcoma at 65 years old (Figure 7, Panel A).

2.3.1. Germline Landscape

F7 had two germline variants, one in *MAP3K6* and one in *BRCA1* genes (Table 12). *MAP3K6* encodes for a serine/threonine kinase that participates in the regulation of vascular endothelial growth factor (VEGF) expression [(129)]. The variant found in *MAP3K6* (p.Pro946Leu) was very rare both in European and Tuscany populations (0.5% and 0.47%, respectively) and was classified as VUS (Table 12). Further, this variant was unique in the cohort analyzed (2%) and enriched in comparison with European and Tuscany populations (Table 12) Moreover, p.Pro946Leu was not located in any known domain of the protein (Figure 7, Panel B). Interestingly, this variant was previously found in a familial gastric cancer cohort [(31)], increasing the likelihood of this variant being a susceptibility factor in the context of gastric cancer.

The *BRCA1* gene is involved in the HR pathway by repairing double strand errors in the DNA molecule [(122)]. The variant c.-20+2dupT in *BRCA1* was not yet described in the ExAC

and 1000Genome project database and was unique in the cohort. Further, was located in intron 1, specifically in a splice region, and was classified as VUS (Table 12).

Splicing prediction revealed that the duplication of T did not affect this splice acceptor, being scored in 0.79 (a 79% probability of occurrence of splice acceptor) in both the wild-type and mutant sequence. The T nucleotide duplication leads to a frameshift mutation, however, since it occurred before the ATG, there is no alteration of the sequence frame (Figure 7, Panel C). Nevertheless, this region is associated with highest promoter activity, as well as, E2F transcription factor binding, important to induce BRCA1 activity [(130)].

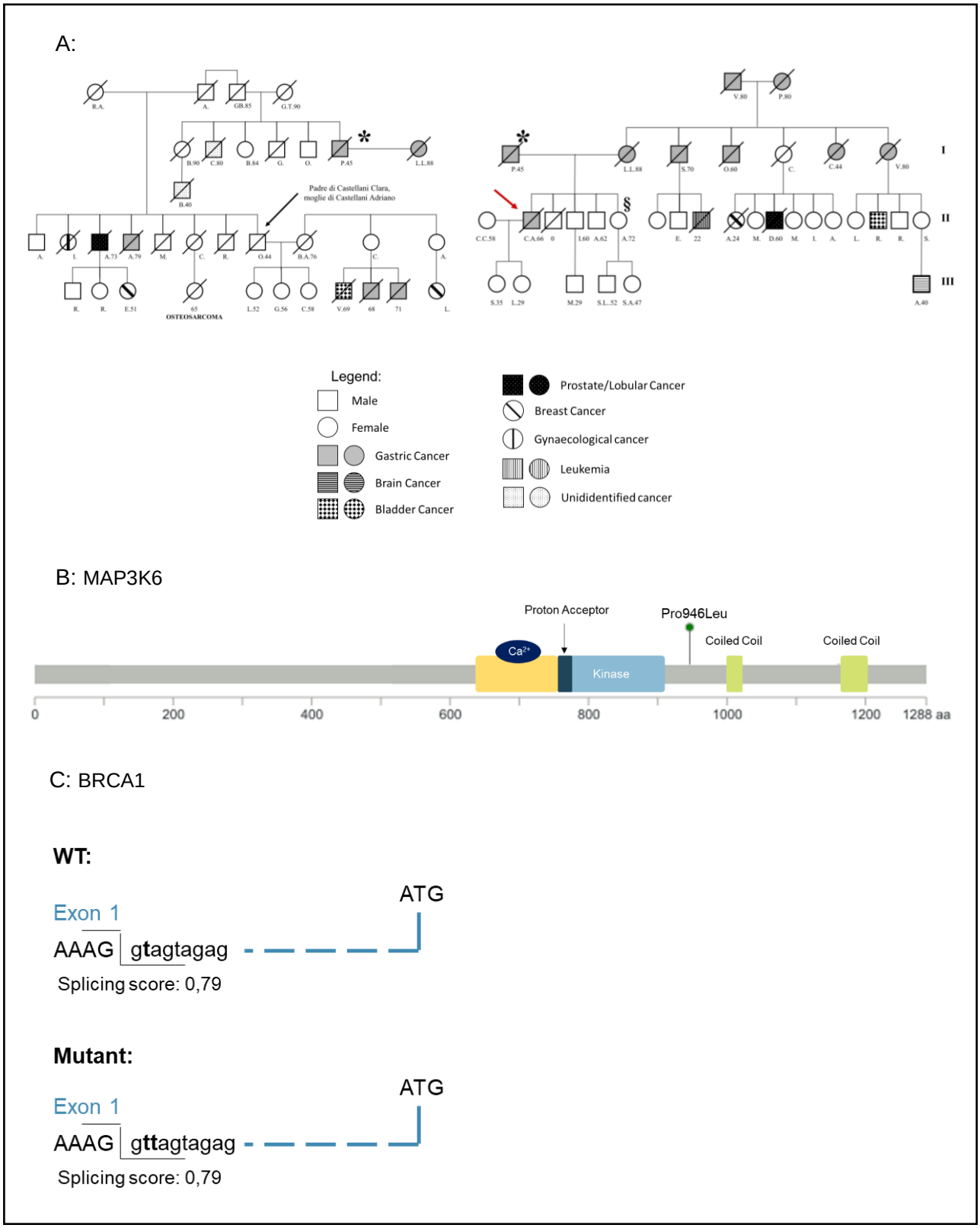


Figure 8. *Germline variants of Family 7. A: Pedigree of family 7. The proband is displayed by an arrow. B: Schematic representation of MAP3K6 protein, its domains, interacting molecules and germline variant found represented by a green dot (missense variants). aa: amino acids [57,58]. C: Splicing prediction of variant c.-20+2dupT in BRCA1. In bold is represented the variant alteration.*

2.3.2 Somatic Landscape

The F7 tumor is MSS (previous analysis) and had one somatic variant in the *TP53* gene, p.His168Leu (Figure 8, Panel A).

2.3.3 Characterization of genetic and epigenetic second hits

MAP3K6 and *BRCA1* were not found to be mutated at the somatic level. Therefore, promoter hypermethylation was tested to assessed the second inactivation hit of *MAP3K6.*

Methylation status of the *MAP3K6* promoter was evaluated using two sets of primers. As shown in Figure 8, Panel B, the *MAP3K6* promoter (CpG sites 4 to 26) was not methylated in F7 tumor. Given that it was previously reported in a Familial Gastric Cancer (FGC) tumor, complete methylation of a CpG island within the *MAP3K6* gene, in opposition to the normal counterpart, this CpG island (located in exon 15 of *MAP3K6*) was analyzed and predicted to harbor promoter associated features, by using the same strategy as in reference [31].

As shown in Figure 8, Panel C, the CpG island located in gene body of *MAP3K6* (CpG sites 5 to 19) was hemi-methylated both in tumor and normal tissue of F7 proband. Due to the fact that the normal tissue was gastric mucosa adjacent to the tumor, it can be inferred that inflammation or other tumor-related phenomena could lead to methylation of the wild-type allele. Therefore, an unrelated normal gastric mucosa from a bariatric surgery was analyzed and found to be hemi-methylation.

Additionally, to confirm these results, UCSC database [131] was analyzed to assess the methylation status of a normal stomach. The currently available information shows that the gene body CpG island of *MAP3K6* is also hemi-methylated in a normal stomach (Figure 8, Panel D), which may suggest that neither second somatic mutation nor CpG island hypermethylation are the second inactivation hits in F7.

In summary, F7 had two germline variants, in *MAP3K6* and in *BRCA1.* The *MAP3K6* germline variant found in F7 has been previously identified in 2 unrelated families with aggregation of gastric cancer [31], supporting a potential role for this protein in gastric cancer. Additionally, the *BRCA1* germline variant is located in a highest promoter activity region, to which E2F transcription factors binds. One possibility is that the variant c.-20+2dupT in *BRCA1* may inhibit the binding of E2F, leading to decrease in BRCA1 protein expression. Furthermore, 3 familial members of F7 have been diagnosed with breast cancers, however, c.-20+2dupT has not been found in Breast and Ovarian Hereditary Syndrome. Thus, it can be inferred that this variant display a low or moderate risk allele. Interestingly, in a sporadic

cohort of GC (TCGA) [57,58], 2 individuals (approximately 1%) display co-occurrence of somatic variants in *MAP3K6* and *BRCA1*. Thus, it could be inferred that co-occurrence of low or moderate risk alleles in *MAP3K6* and *BRCA1* may increase predisposition to FIGC.

A

Gene	Chr	Position GRCh38	cDNA	Protein	Consequence	Freq cohort
TP53	17	7675109	c.503A>T	p.His168Leu	Missense	1

B: MAP3K6

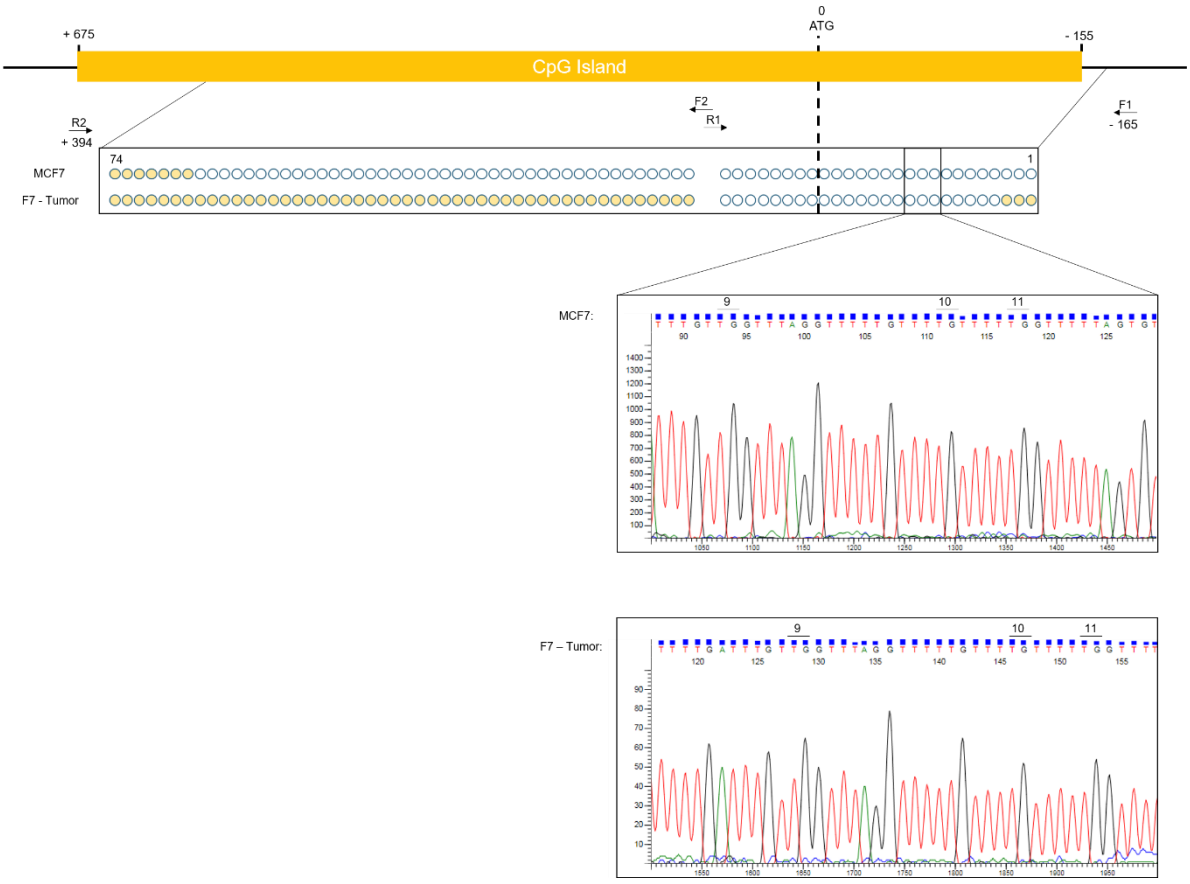
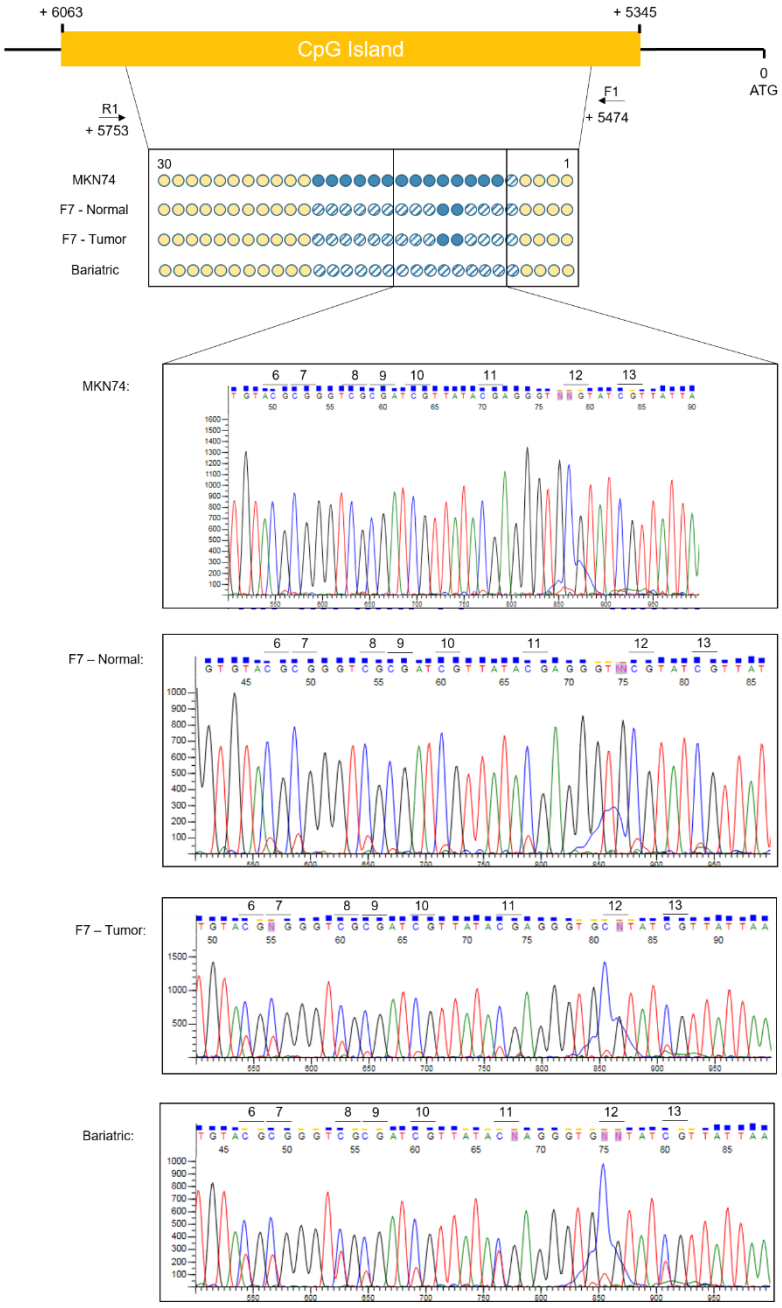


Figure 9. *Somatic events in Family 7. A: Somatic variants found in family 7 by multiplex custom-panel based sequencing. B: MAP3K6 Promoter Methylation Analysis. Schematic representation of the CpG island located at the MAP3K6 promoter. Open circles represent a non-methylated CpG and yellow circles represent not accessed CpG sites; Representative examples of sequences of each sample analyzed: MCF7 cell line (negative control) and F7 tumor.*

C: MAP3K6


+ 6063
+ 5345
CpG Island
0
ATG
R1
+ 5753
F1
+ 5474
30
1
MKN74
F7 - Normal
F7 - Tumor
Bariatric
MKN74:
F7 – Normal:
F7 – Tumor:
Bariatric:

D:

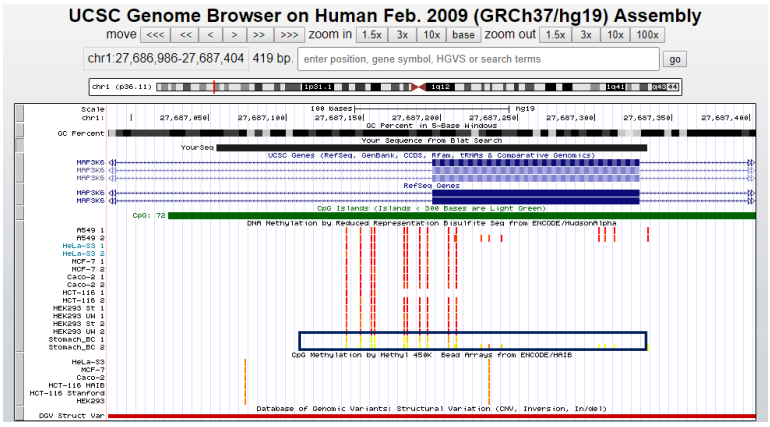

UCSC Genome Browser on Human Feb. 2009 (GRCh37/hg19) Assembly
move
zoom in
zoom out
chr1:27,686,986-27,687,404
419 bp.
enter position, gene symbol, HGVS or search terms
go
MAP3K6
RefSeq Genes
CpG Islands (Islands < 300 Bases are Light Green)
DNA Methylation by Reduced Representation Bisulfite Seq from ENCODE/HudsonAlpha
CpG Methylation by Methyl 450K Bead Arrays from ENCODE/HAIB
Database of Genomic Variants: Structural Variation (CNV, Inversion, In/del)

Figure 8 (cont.). *Somatic events in Family 7. C: MAP3K6 gene body Methylation Analysis. Schematic representation of the CpG island located at the MAP3K6 gene body. Open circles represent a non-methylated CpG; blue circles represent methylated CpG, white circles with blue risks represent hemi-methylated CpG and yellow circles represent not accessed CpG; Representative examples of sequences of each sample analyzed: MKN74 cell line (negative control), F7 normal, F7 tumor and sample of bariatric surgery. D: UCSC genome browser, displaying gene body CpG island of MAP3K6. The blue rectangle represents the region amplified in normal stomach and are represented in yellow and orange, code for hemi-methylated region.*

2.4. Family 8

Family 8 was represented by a 73 years old female proband (F8) with intestinal-type GC. By analyzing the pedigree, it was observed that a sister of the proband was diagnosed with gastric ulcer (Figure 9, Panel A). Furthermore, the father of the probands husband was diagnosed with GC at 40 years old. The proband had one daughter diagnosed with GC at 49 years old and a granddaughter diagnosed with gastric ulcer at 21 years old (Figure 9, Panel A).

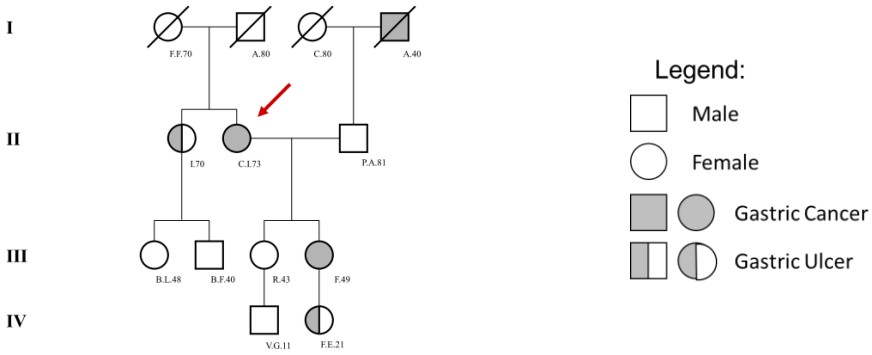
2.4.1. Germline Landscape

F8 had two germline variants in *CASP10* and in *MAP3K6* genes (Table 12).

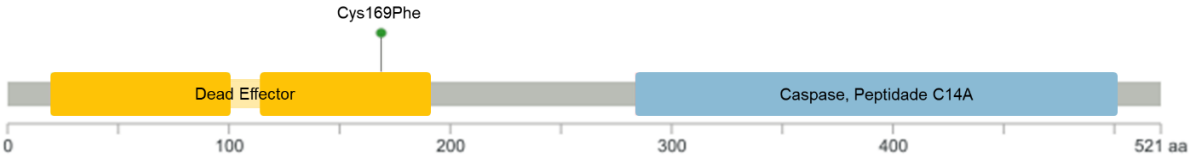
CASP10 belongs to the caspase family, playing a role similar to *CASP8* in apoptosis [(90)]. The missense variant p.Cys169Phe in *CASP10* was very rare in the global population and was classified as VUS (Table 12). This variant was unique in the cohort analyzed (2%), being enriched in comparison with global population (0.0008%) and was located in the Dead Effector domain (Table 12, Figure 9, Panel B), which is described as a protein-protein interaction domain involved in apoptosis [(132)]. It may be speculated that this very rare variant could hamper the protein-protein interaction and inhibit apoptosis.

The *MAP3K6* variant (c.1256-2A>G) was very rare in Europe and Tuscany populations (0.7% and 0.47%, respectively) and recurrently appeared in F3 and F8 from this cohort (4%), being enriched in comparison with European (0.7%) and Tuscany (0.47%) populations (Table 12). This variant was classified as VUS and was located in intron 8, between exon 8 and exon 9 and can give rise to splice defects (Table 12). In fact, splicing prediction of the missense alteration A>G revealed a decreased probability of occurring a splice acceptor (from 0.94-wild-type to 0.65-mutant) and a creation of a new splice acceptor. In this last scenario, a frameshift occurs, increasing two nucleotides in the beginning of exon 5 (Figure 9, Panel C), which is predicted to lead to an alteration of the frame in exon 5 that could be deleterious, due to a premature stop codon, for example.

A:



B: CASP10



C: MAP3K6

WT:

Exon 9

AAGTAGC | gacccc

Splicing score: 0,94

Mutant:

Exon 9

AAGTAGC | gacccc

Splicing score: 0,94

Exon 9

AAGTAGCG**G** | cccc

Splicing score: 0,67

Figure 10. *Germline variants of Family 8. A: Pedigree of family 8. The proband is displayed by an arrow. B: Schematic representation of CASP10 protein, its domains, interacting molecules and germline variant found represented by a green dot (missense variants). aa: amino acids [57,58]. C: Splicing prediction of variant c.1256-2A>G in MAP3K6. In bold is represented the variant alteration.*

2.4.2 Somatic Landscape

At the somatic level, four variants were found in *ARID1A, FAT4* and *PTEN* (Figure 10, Panel A). In addition, this tumor displayed an MSI phenotype and had methylation of *MLH1* gene. The *PTEN* gene might be an MSI target gene, since the variant occurred at a microsatellite region. Interestingly, F8 does not exhibit a somatic variant in *TP53* and is methylated in the promoter region of *MLH1*, characteristics of MSI tumors.

2.4.3. Characterization of genetic and epigenetic second hits

Somatic variants in *CASP10* or *MAP3K6* genes were not found in the F8 tumor, and do not constitute the second inactivation hit. Therefore, methylation status of both genes was analyzed.

The methylation results obtained for the promoter and gene body of *MAP3K6* were similar to those obtained for F7 (*i.e* the promoter was not methylated and the gene body was hemi-methylated both in tumor and normal tissues) (Figure 10, Panel B and C). These findings suggested that a second mutation and methylation in the promoter and gene body regions of MAP3K6 might not be the second inactivation hits.

To evaluate the methylation status of the *CASP10* promoter, two sets of primers were designed. Figure 10, Panel D showed that CpG sites 1-10 were methylated and CpG sites 11-18 were hemi-methylated both in F8 tumor and normal tissues, whereas all CpG sites were methylated in the HCT-116 cell line (the positive control). For the same reasons as describe above, tissue from a bariatric surgery was analyzed and the same pattern was observed. Further validation in the UCSC database could not be performed, since methylation in this region was not evaluated by microarrays. All these findings led us to hypothesize that promoter methylation may not be the second inactivation hit and do not control the expression of *CASP10*. However, further studies should be conducted.

In summary, the co-occurrence of *CASP10* and *MAP3K6* germline variants might be the underlying cause of GC in this family. In fact, *CASP10* and *MAP3K6* somatic alterations appear, independently in each gene in approximately 4% of the sporadic GC cases depicted in the TCGA database, and co-occur in a single case (0.4%) [57,58]. Furthermore, *CASP10* alterations had been associated with gastric cancer risk, highlighting the importance of this gene in gastric adenocarcinoma [133]. Nevertheless, *MAP3K6* has been identified as a genetic risk factor for FGC [31]. Thus, by combining the germline and somatic data, it may be inferred that the co-occurrence of germline variants in moderate or low risk alleles (*CASP10* and *MAP3K6*) could predispose to FIGC.

A

Gene	Chr	Position GRCh38	cDNA	Protein	Consequence	Freq cohort
ARID1A	1	26773716	c.4003C>T	p.Arg1335Ter	Stop gained, Splice region	1
FAT4	4	125449221	c.8205A>C	p.Lys2735Asn	Missense	1
PTEN	10	87952216	c.595_597delATG	p.Met199del	Inframe deletion	1
PTEN	10	87958013	c.800delA	p.Lys267ArgfsTer9	Frameshift Indels	1

B: MAP3K6

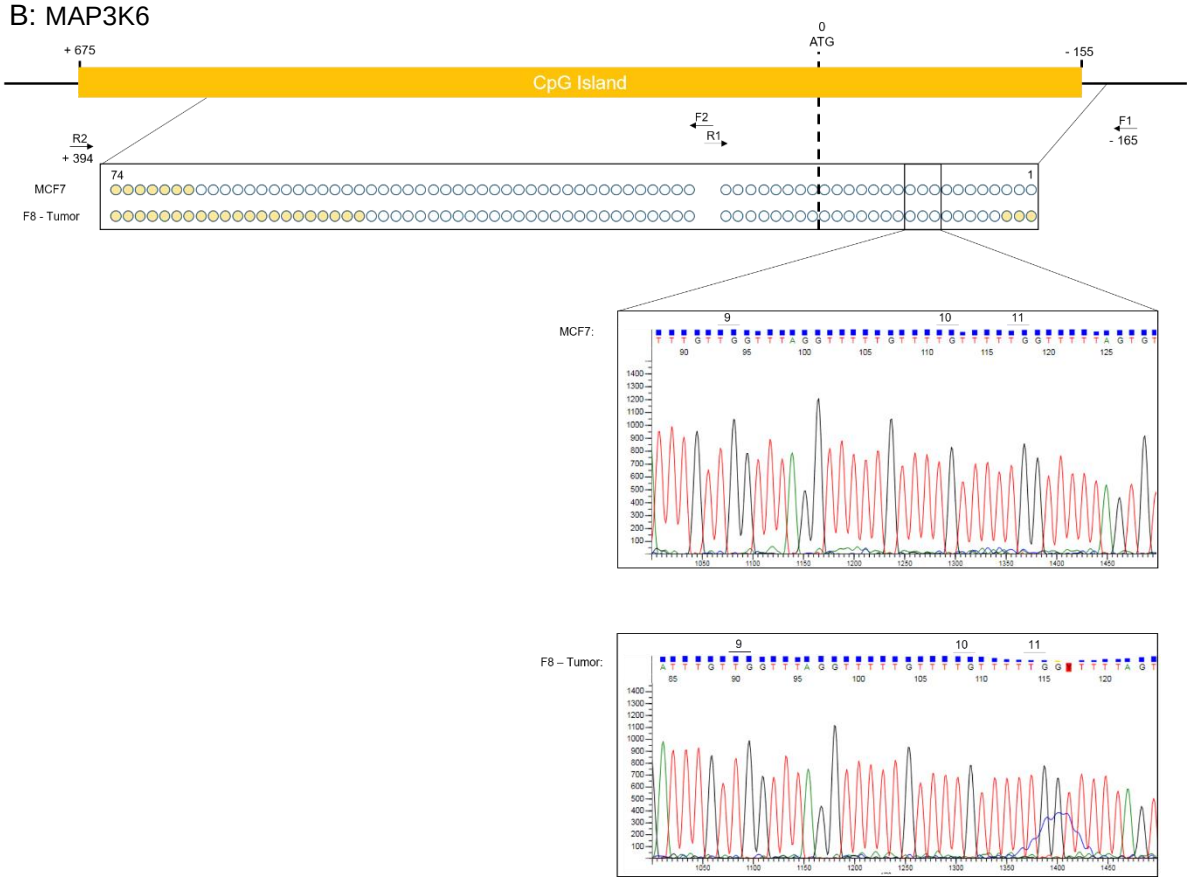


Figure 11. *Somatic Events in Family 8. A: Somatic variants found in family 8 by multiplex custom-panel based sequencing. B: MAP3K6 Promoter Methylation Analysis. Schematic representation of the CpG island located at the MAP3K6 promoter. Open circles represent a non-methylated CpG and yellow circles represent not accessed CpG sites; Representative examples of sequences of each sample analyzed: MCF7 cell line (negative control) and F8 tumor.*

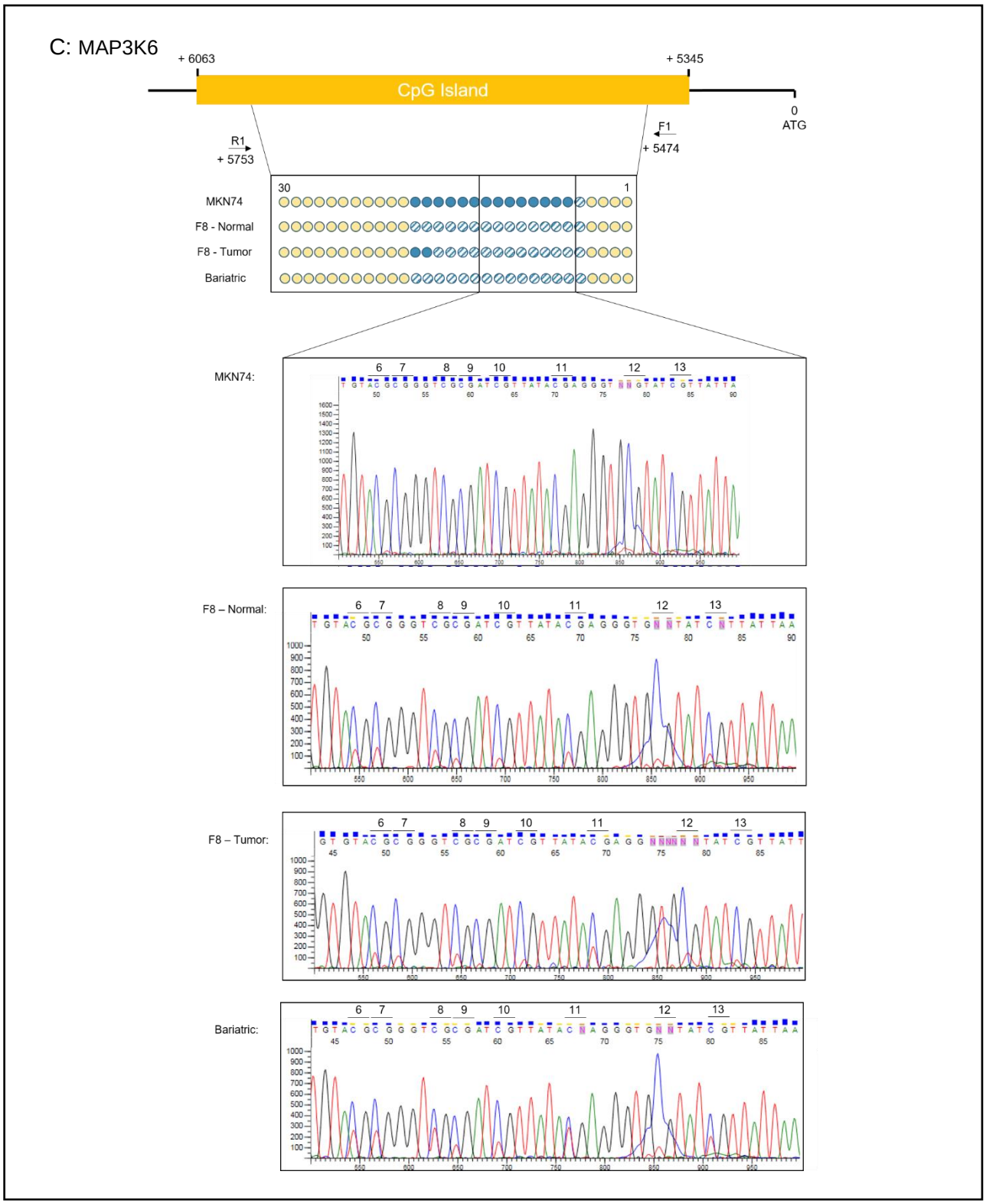


Figure 12 (cont.). *Somatic Events in Family 8. C: MAP3K6 gene body Methylation Analysis. Schematic representation of the CpG island located at the MAP3K6 gene body. Open circles represent a non-methylated CpG; blue circles represent methylated CpG, white circles with blue risks represent hemi-methylated CpG and yellow circles represent not accessed CpG; Representative examples of sequences of each sample analyzed: MKN74 cell line (negative control), F8 normal, F8 tumor and sample of bariatric surgery.*

D: CASP10

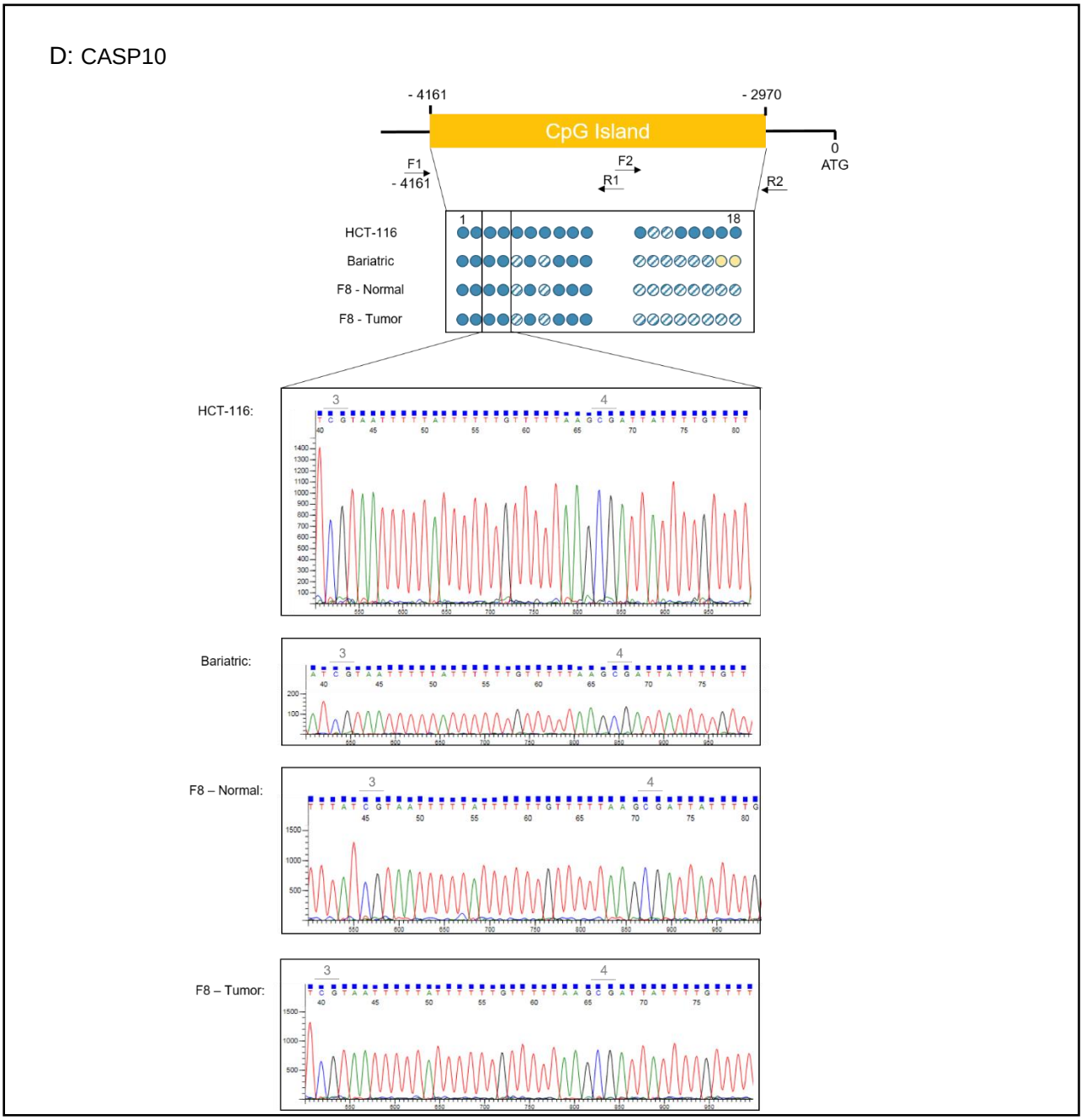


Figure 13 (cont.). *Somatic Events in Family 8. D: CASP10 gene promoter Methylation Analysis. Schematic representation of the CpG island located at the CASP10 promoter. Open circles represent a non-methylated CpG; blue circles represent methylated CpG, white circles with blue risks represent hemi-methylated CpG and yellow circles represent not accessed CpG; Representative examples of sequences of each sample analyzed: HCT-116 cell line (positive control), F8 normal, F8 tumor and sample of bariatric surgery.*

2.5. Family 11

Family 11 was represented by a 65 years old female proband (F11) with intestinal-type GC. The pedigree of this family showed that two brothers with disease: one had gastric ulcer at 66 years old and the other had GC at 47 years old. In addition, the proband's cousin had breast cancer at 44 years old (Figure 11, Panel A).

2.5.1. Germline Landscape

F11 had two germline variants in *MSH6* and *ATM* genes. The variant found in *MSH6*, (p.Val878Ala) was rare in Europe (1.3%) and Tuscany (2.34%) populations and was frequent in the cohort analyzed (8%) (Table 12). In addition, it was classified as conflicting by Annotator57, due to contradictory information regarding pathogenicity depicted in the databases (Table 12). This *MSH6* variant has been recently described in a homozygous state in two individuals, but was not associated with Lynch Syndrome (data available from collaborators and not published). Therefore, it is likely that this specific variant is not associated with Lynch, but may contribute to increase FIGC susceptibility. This variant was present in MutS domain III (Figure 11, Panel B), that suffers a large conformational change for the DNA molecule to bind to domain I [119]. It is possible that a single amino acid change at this position of the protein interferes with this conformational change, but that remains to be proved.

ATM is a tumor suppressor gene that encodes for a cell cycle checkpoint kinase that regulates downstream proteins, such as *TP53*. The variant p.His1436Tyr in *ATM* is very rare in Europe (0.1%) and in Tuscany (0.47%) and was classified as VUS (Table 12). This variant was not located in any relevant domain of the *ATM* protein (Figure 11, Panel C). Furthermore, the cousin of the proband was diagnosed with breast cancer, which might be due to the presence of a germline variant in *ATM* (a moderate penetrance gene for hereditary breast cancer) [134].

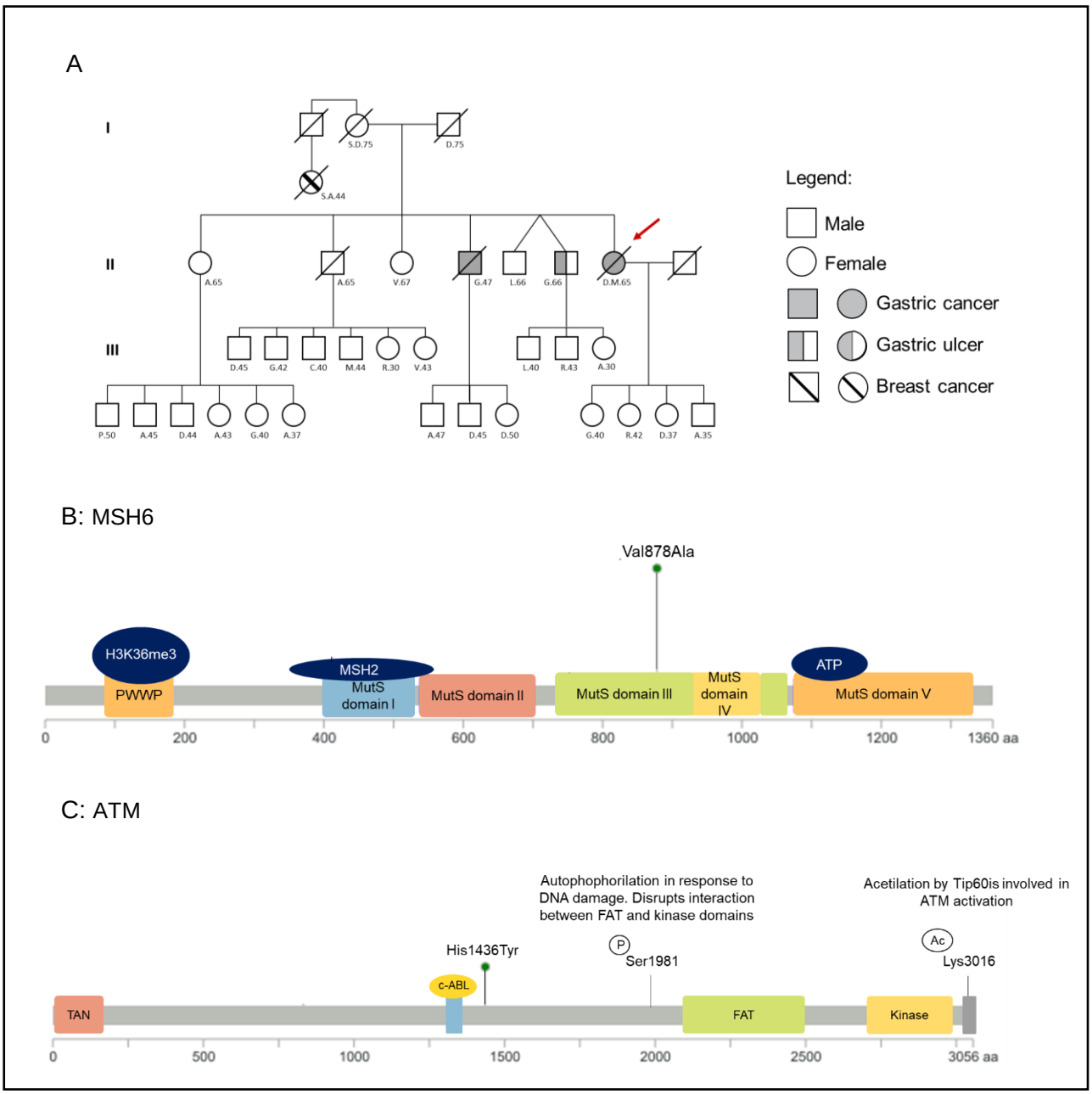


Figure 14. *Pedigree and germline variants of Family 11. A: Pedigree of family 11. The proband is displayed by an arrow. B: Schematic representation of MSH6 protein, its domains, interacting molecules and germline variant found represented by a green dot (missense variants). aa: amino acids [57,58]. C: Schematic representation of ATM protein, its domains, interacting molecules and germline variant found represented by a green dot (missense variants [57,58]).*

2.5.2 Somatic Landscape

At the somatic level, two variants were found in *CDH1* and *TP53* (Figure 12, Panel A). In addition, this tumor displayed a phenotype of MSI and a characteristic methylation of *MLH1* promoter. However, a *TP53* variant at the somatic level is not characteristic of MSI tumors.

2.5.3 Characterization of genetic and epigenetic second hits

Somatic variants in *MSH6* or *ATM* genes were not found in the F11 tumor and do not constitute the second inactivation hit. Therefore, the methylation status of both genes was assessed.

The CpG island of *MSH6* was analyzed as previously described and it was observed that CpG sites 1 to 59 were not methylated in the tumor of F11 (Figure 12, Panel B).

Three sets of primers were designed to analyze the majority of the CpG island of *ATM*, according to Pal, R et al. [135]. Similarly to *MSH6*, the CpG island of *ATM* (CpG sites 1 to 29 and 33 to 63) was not methylated in the F11 tumor, as well as, the negative control, HCT-116 (Figure 12, Panel C). These findings indicated that CpG island methylation was not the second inactivation hit in both genes.

A

Gene	Chr	Position GRCh38	cDNA	Protein	Consequence	Freq cohort
CDH1	16	68819423	c.1709A>G	p.Asn570Ser	Missense, Splice region	1
TP53	17	7674947	c.584T>C	p.Ile195Thr	Missense	1

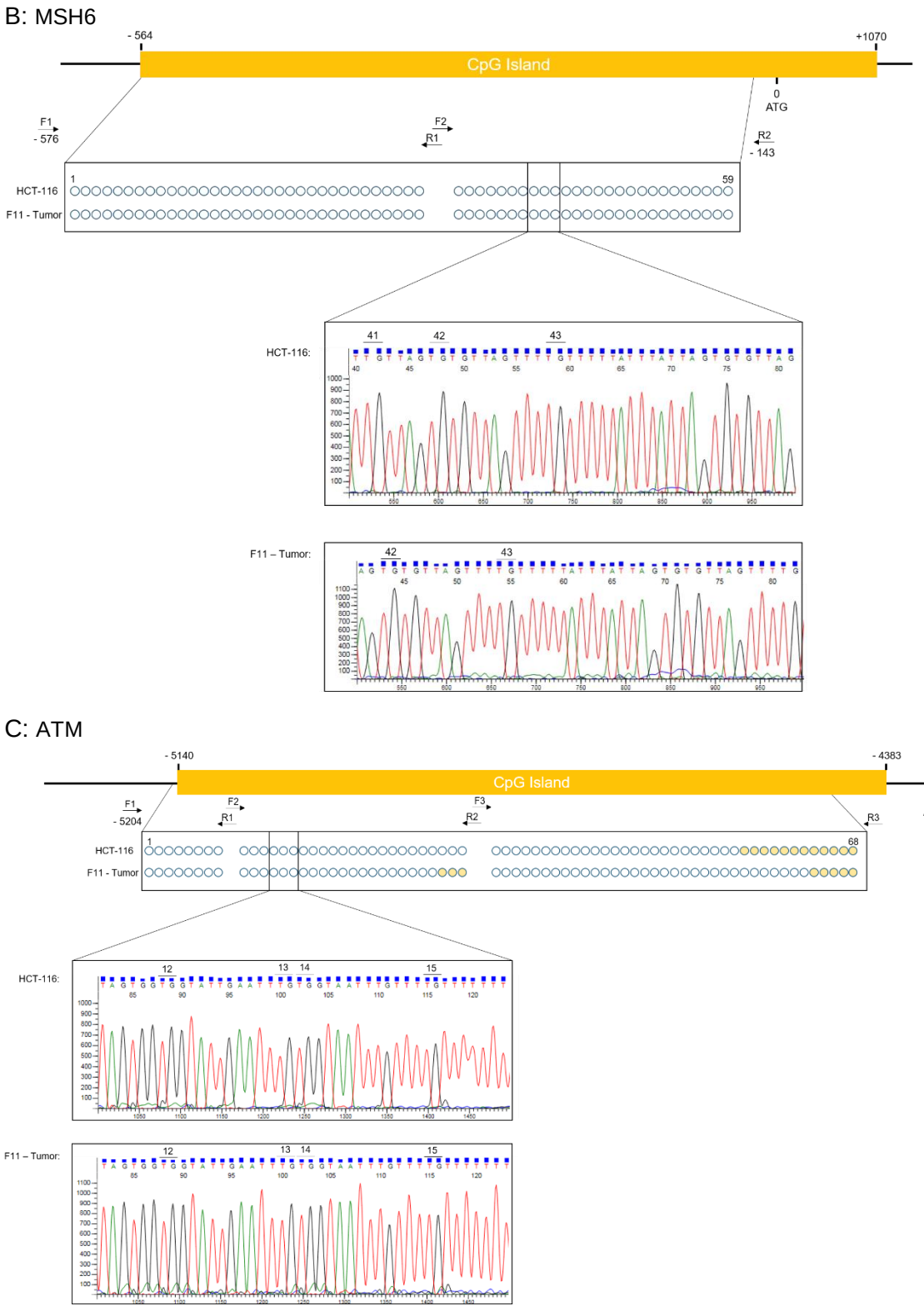


Figure 15. *Somatic Events in Family 11. A: Somatic variants found in family 11 by multiplex custom-panel based sequencing. B: MSH6 Promoter Methylation Analysis. Schematic representation of the CpG island located at the MSH6 promoter. Open circles represent a non-methylated CpG and yellow circles represent not accessed*

CpG sites; Representative examples of sequences of each sample analyzed: HCT-116 cell line (negative control) and F11 tumor. C. ATM gene promoter Methylation Analysis. Schematic representation of the CpG island located at the ATM promoter. Open circles represent a non-methylated CpG and yellow circles represent not accessed CpG; Representative examples of sequences of each sample analyzed: HCT-116 cell line (positive control) and F11 tumor.

Notwithstanding, *MSH6* is not usually a driver of the MSI phenotype [120], the association with *ATM* could be the underlying cause of FIGC in this family. Furthermore, F11 proband carried alterations in four genes that participate in DNA damage repair: *MSH6* and *ATM* at the germline level (blue) and *MLH1* and *TP53* at the somatic level (yellow), that together could predispose to FIGC (Figure 13). According to the TCGA database [57,58], *MSH6* and *ATM* somatic variants co-occur in 6 tumors (2%) with sporadic GC. Thus, it can be inferred that co-occurrence of germline low or moderate risk variants at *MSH6* and *ATM* could predispose to FIGC in this family, by interfering with DNA repair pathways.

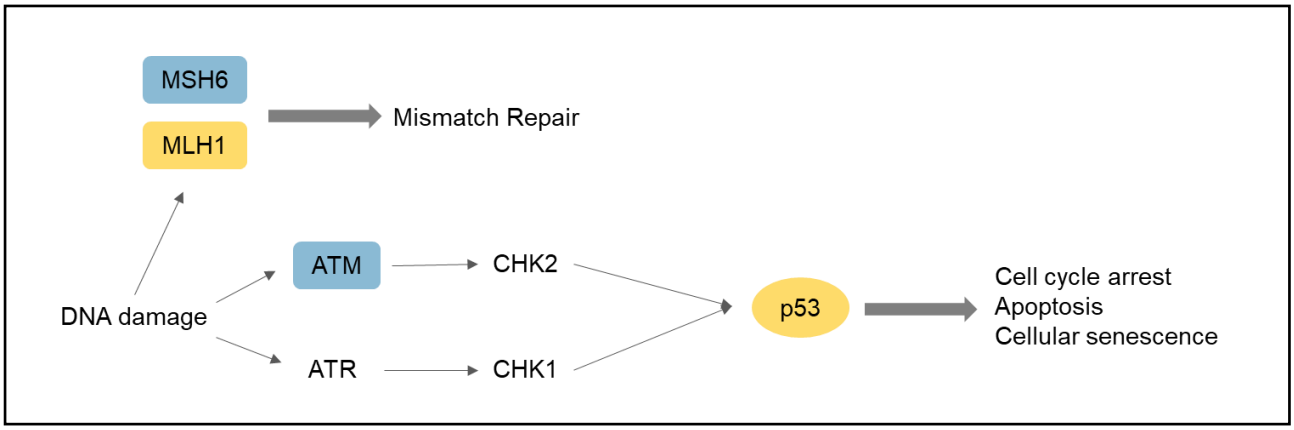


Figure 16. *DNA damage repair pathways – MRR and HR. Blue represents germline variants and yellow represents somatic variants.*

2.6. Family 12

Family 12 was represented by a 71 years old male proband (F12) with intestinal-type GC. The pedigree of this family showed a proband's brother with lung cancer, a sister with breast cancer and a cousin with GC (Figure 14, Panel A).

2.6.1. Germline Landscape

F12 had two germline variants in genes *BRCA2* and *CTHRC1* (Table 12).

The variant found in *BRCA2,* p.Thr939SerfsTer7, is novel and is predicted to generate a premature termination codon 7 amino acids downstream of the frameshifted nucleotide. This BRCA2 truncating variant was classified as likely pathogenic since numerous likely

pathogenic variants were already described in the same region, particularly in the same codon, with similar consequences. The frequency of this variant was not described in ExAC and in the 1000Genome project and was unique in the FIGC cohort (Table 12). Additionally, this variant was not located in any known domain (Figure 14, Panel B). Interestingly, a sister of the proband had breast cancer (not tested for the variant so far), suggesting that this *BRCA2* germline variant may represent a risk allele also for breast cancer.

F12 also carried a novel germline in-frame duplication (p.Leu20_Leu21dup) at the *CTHRC1* gene that results in the insertion of two additional Leucine residues in the protein. This gene is involved in Planar Cell Polarity (PCP), through stabilization of the complex Wnt-Fzd-Ror2 and activation of RhoA. Additionally, *CTHRC1* is responsible for the preferential activation of PCP in alternative to the canonical WNT-signaling pathway [(136)]. This unique variant was classified as VUS, since no information was not found, including frequency in ExAC and in the 1000Genomes project (Table 12). Only one domain has been identified in this gene (Collagen triple helix repeat) and variant p.Leu20_Leu21dup was not in the described domain (Figure 14, Panel C). However, residues 20 and 21 of CTHRC1 are part of a Leucine-rich domain which, is conserved at least in mouse, and is likely important for the function of this protein. Moreover, a close variant - Q44P -, has been associated with disruption of the protein's ability to form secondary structure [(137)].

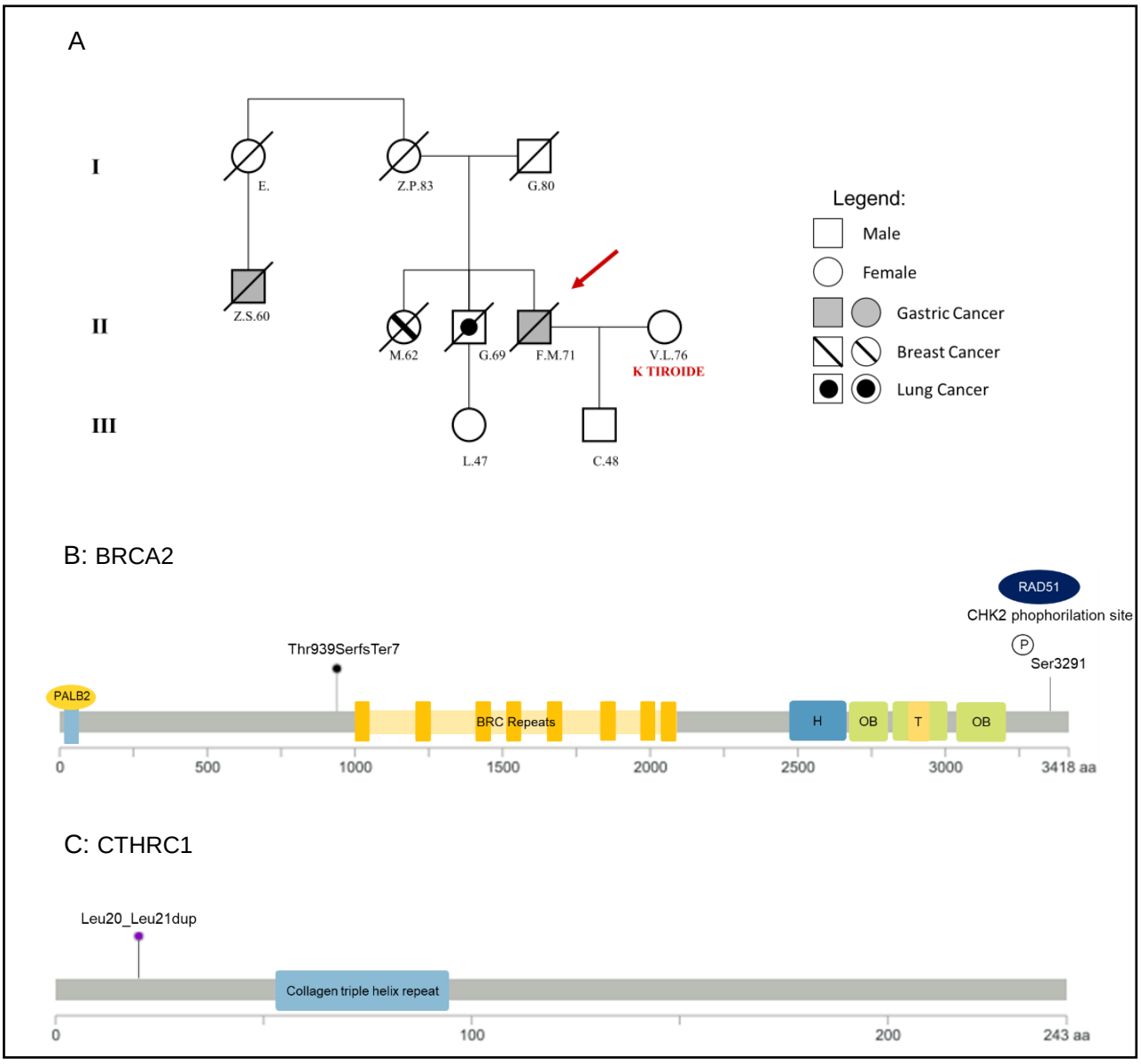


Figure 17. *Pedigree and germline variants of Family 12. A: Pedigree of family 12. The proband is displayed by an arrow. B: Schematic representation of BRCA2 protein, its domains, interacting molecules and germline variants found represented by blue dots (missense variants). aa: amino acids [57,58]. C: Schematic representation of SDHD protein, its domains, interacting molecules and germline variant found represented by purple dots (frameshift variants [57,58]).*

2.6.2 Somatic Landscape

At the somatic level, one variant was found in *APC*, a tumor suppressor gene associated with cell polarity and proliferation – WNT-signaling pathway (Figure 15, Panel A). From previous analysis of this family, the tumor has LOH of *CDH1* and displayed a phenotype of MSI.

2.6.3. Characterization of genetic and epigenetic second hits

Somatic variants in *BRCA2* and *CTHRC1* were not found, therefore the following step was to evaluate promoter methylation of these genes to determine if promoter methylation could be the somatic second inactivation hit.

In order to evaluate promoter methylation status in *BRCA2*, a pair of primers were designed according to Price RJ, et al. [138]. Tumor of F12 was not methylated in the CpG sites analyzed (1-26), as well as, the negative control, HCT-116 (Figure 15, Panel B).

For the analysis of *CTHRC1* CpG island, 3 sets of primers were designed to analyze the entire island. Since Hela cells (positive control), were not methylated in the regions analyzed with the first and second sets of primers, the third region was prioritized. In fact, *CTHRC1* was hemi-methylated (CpG sites 59-73) in Hela cells and was fully methylated in MKN74 cells (CpG sites 57-73). Both tumor and normal tissue of F12 were hemi-methylated (59-72 and 57-73, respectively). Since normal tissue derived from tumor adjacent mucosa could be affected by the tumor, methylation status from a normal patient who underwent a bariatric surgery was analyzed. This tissue was also hemi-methylated (CpG sites 57-72), Figure 16, Panel A. In addition, UCSC database [131] was used to evaluate the methylation status of normal stomach and found that the region amplified by the third set of primers was not methylated, using microarrays (Figure 16, Panel B). In fact, it can be speculated that the gastric mucosa of the normal individual could be affected by environmental factors, such as *H. pylori* infection or ingestion of salted food, leading to inflammation or other tumor-related phenomena. Nevertheless, the possibility that different techniques might also influence the methylation status evaluation could not be excluded.

2.6.4. RNA Expression

Then, quantitative real-time PCR was performed, to evaluate the impact of germline and somatic variants found at *CTHRC1* and *APC* genes, respectively. Expression of *β-Catenin* was also quantified since it is a downstream target of *APC*. Of notice, RNAs from a commercial normal stomach, commercial normal breast and HCT-116 cell line were used as positive controls and MKN74 as negative control.

Observing the results, F12 tumor did not express *APC*, but expressed *CTHRC1* and *β-Catenin* and the normal counterpart did not express *APC* and *CTHRC1*, but expressed *β-Catenin* (Figure 16, Panel C). Since *APC* is essential for the embryonic development [139], the absence of its expression in F12 normal tissue may be explained by the poor quality of the RNA. This assumption is based on the fact that the endogenous control (*18S*) was

amplified very late. Therefore, commercial normal stomach was used as a control, instead of the normal counterpart of tumor F12.

CTHRC1 is not expressed in normal stomach and is hemi-methylated in the region analyzed (Figure 16, Panel B and C). It can be inferred that repressors might bind to this region in order to suppress the expression of the RNA, that alongside with hemi-methylation of the other allele might explain absent RNA expression (Figure 16, Panel B and C). In F12 tumor, hemi-methylation is conserved and *CTHRC1* RNA expression is increased (Figure 16, Panel A and C). This may be due to the absence of repressors in the unmethylated allele that is sufficient to increase RNA expression. Furthermore, MKN74 cell line is methylated and *CTHRC1* RNA is not expressed, revealing that the action of *CTHRC1* repressors might not act in this context, or could act synergically with promoter methylation (Figure 16, Panel B and C). Moreover, expression of *CTHRC1* is decreased in comparison with positive control, normal breast tissue. According to Wang et al [(140)], protein expression of CTHRC1 was absent in gastric normal mucosa and was progressively increased with stage of gastric adenocarcinoma [(140)], which was in accordance with our RNA expression results. However, Wang et al. [(140)] also stated that promoter demethylation is responsible for increased protein expression. It can be inferred that, in tumor context, repressors that inactivate *CTHRC1* in the stomach might be absent and methylation might not be sufficient to decrease RNA expression. Thus, it can also be inferred that *CTHRC1* could act as a pro-tumorous gene and is activated with a germline variant p.Leu20_Leu21dup. In summary, since *CTHRC1* is not expressed in the stomach, an activating mutation might not impact protein expression in this tissue. In gastric tumor context, repression of this gene is inhibited and *CTHRC1* is expressed.

The absence of *APC* mRNA expression in the tumor may be due to the presence of a somatic mutation at the *APC* gene. In addition, mRNA expression levels of *β-Catenin* of F12 tumor were decreased in comparison with commercial normal stomach. These findings are in accordance with *CTHRC1* being responsible for the preferential activation of PCP in alternative to the canonical WNT-signaling pathway [(136)], decreasing expression of *β-Catenin,* a key protein in WNT-signaling pathway.

A

Gene	Chr	Position GRCh38	cDNA	Protein	Consequence	Freq cohort
APC	5	112842961	c.7367T>A	p.Leu2456Ter	Stop gained	1

B: BRCA2

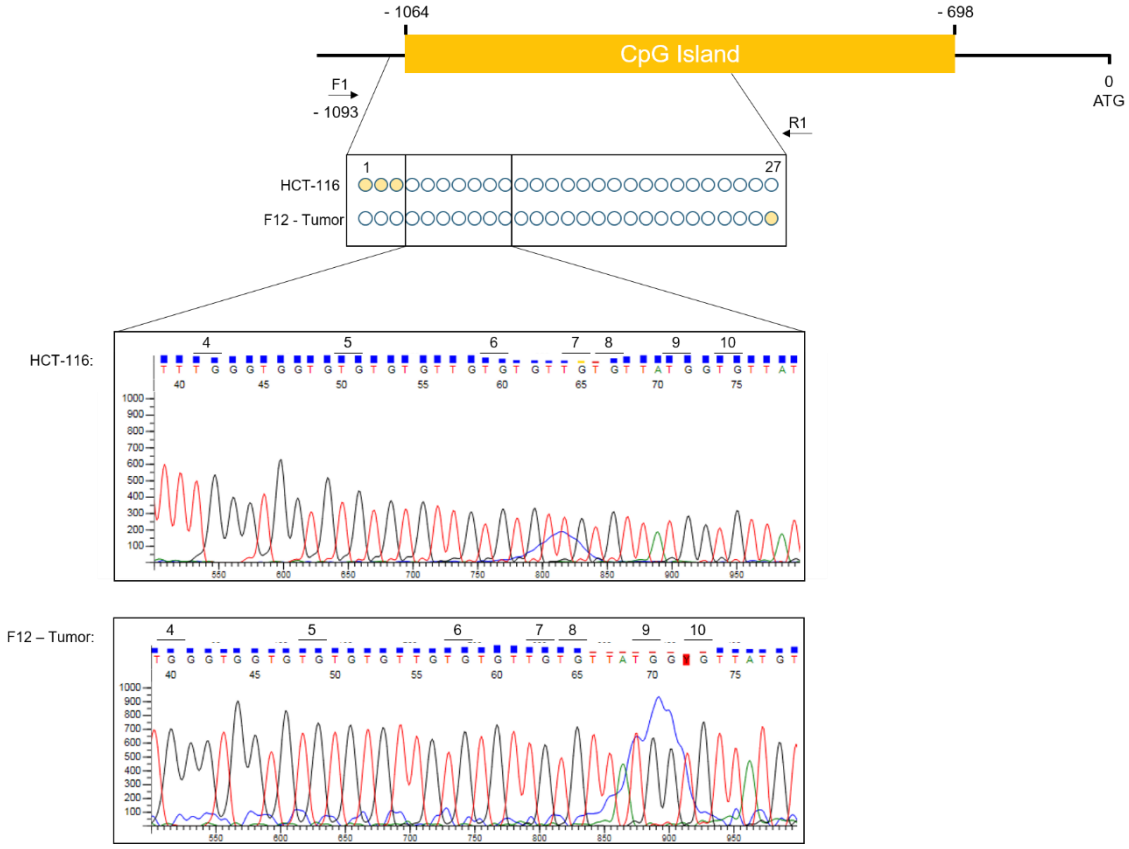


Figure 18. *Somatic Events in Family 12. A: Somatic variants found in family 12 by multiplex custom-panel based sequencing. B: BRCA2 Promoter Methylation Analysis. Schematic representation of the CpG island located at the BRCA2 promoter. Open circles represent a non-methylated CpG and yellow circles represent not accessed CpG sites; Representative examples of sequences of each sample analyzed: HCT-116 cell line (negative control) and F12 tumor.*

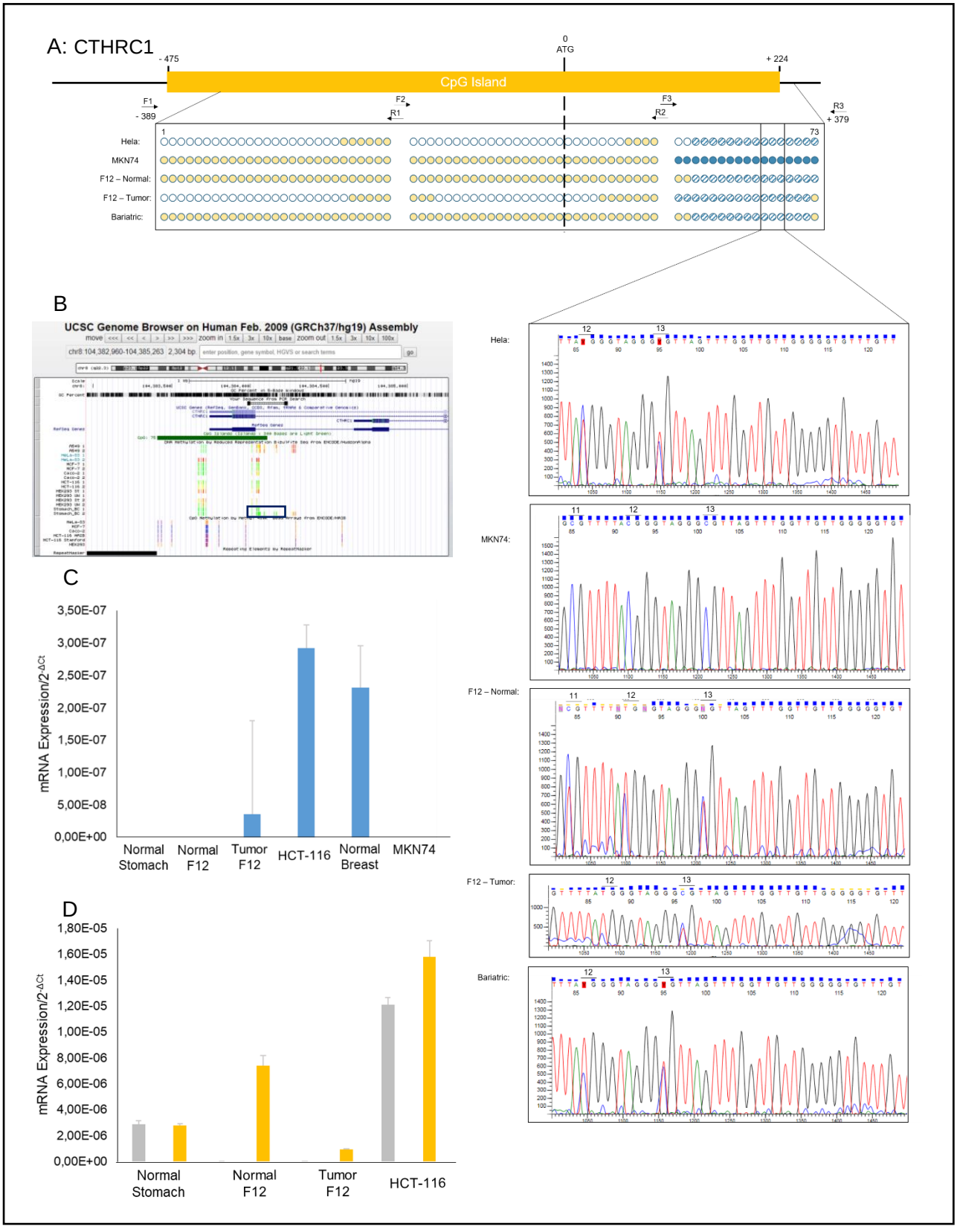


Figure 19. *Somatic Events in Family 12, cont. A: CTHRC1 Promoter Methylation Analysis. Schematic representation of the CpG island located at the CTHRC1 promoter. Open circles represent a non-methylated CpG; blue circles represent methylated CpG, white circles with blue risks represent hemi-methylated CpG and yellow circles represent not accessed CpG; Representative examples of sequences of each sample analyzed: Hela and MKN74 cell lines (positive controls), F12 normal, F12 tumor and sample of bariatric surgery. B: UCSC genome browser, displaying promoter CpG island of CTHRC1. The blue rectangle represents the region amplified by the third set of primers in normal stomach and are represented in green, code for not methylated*

region. C: Expression of CTHRC1 (blue) in normal stomach, normal F12, tumor F12, cell line HCT-116 (positive control), normal breast (positive control) and MKN74 (negative control). D: Expression of APC (grey) and β-Catenin (yellow) in normal stomach, normal F12, tumor F12 and cell line HCT-116 (positive control).

In summary, two germline variants were found in *CTHRC1* and *BRCA2* and one somatic variant in *APC*. In fact, *CTHRC1* and *APC* are involved in WNT signaling pathway, non-canonical (PCP) and canonical, respectively [136]. *CTHRC1* enhances the formation of Wnt-Fzd-Ror2 complex and activates RhoA [136]. Furthermore, *CTHRC1* selectively activates PCP signaling pathway, alternatively to the canonical WNT-signaling pathway [136]. In normal stomach, *CTHRC1* is not expressed, thus, PCP is not the preferential signaling pathway and stabilization of Wnt-Fzd-Ror2 and activation of RhoA is not completed, impairing this pathway. At the somatic level, *APC* variant leads to an imbalance of the canonical WNT-signaling pathway and *CTHRC1* is expressed, leading to a preferential activation of PCP signaling pathway. Additionally, expression of *APC* is absent in tumor F12 and expression of *CTHRC1* is increased in comparison with commercial RNA derived from normal stomach. The importance of *BRCA2* germline variant could not be excluded, due to the affected sister with breast cancer. In fact, the co-occurrence of both germline variants might predispose to breast cancer as well, since *CTHRC1* is expressed in normal breast. According to the TCGA database [57,58], 4 (1.4%) tumors displayed sporadic GC and co-occurring of somatic variants in *CTHRC1* and *BRCA2*. Thus, co-occurrence with *BRCA2* could lead to an impairment of two important signaling pathways: WNT and HR, predisposing this family to FIGC.

2.7. Family 13

Family 13 was represented by a male proband (F13) with 64 years old and intestinal-type GC. By analyzing the pedigree, it was observed that both parents of the proband were affected, the father with GC at 66 years old and the mother with laryngotracheal cancer at 62 years old. Two cousins from different parental lines were also affected, an uncle from the mother side with gastric ulcer at 68 years old and an aunt diagnosed with hepatobiliary cancer at 70 years old. A cousin of the proband was diagnosed with laryngotracheal cancer at 58 years old. A nephew and a niece of the proband were also affected, being diagnosed with a gastric ulcer at 61 years old and a gynaecological cancer at 54 years old, respectively.

2.7.1. Germline Landscape

F13 has two germline variants, in *SDHB* and *MTUS1* (Table 12).

SDHB is a subunit of Succinate Dehydrogenase Complex, responsible for the oxidation of succinate. Variant p.Ser163Pro in *SDHB* was rare in both Europe (1.7%) and Tuscany (1.87%) and was classified as conflicting, since contradictory information was found in the three databases analyzed (Table 12). Additionally, this variant recurrently appeared in F22, being enriched in this FIGC cohort (4%) (Table 12). A recent study revealed that this variant was present in one case of HDGC, being associated with AKT and MAPKT pathways increased activity [(141)]. As represented in Figure 17, Panel A, this variant was not located in any known domain.

MTUS1 encodes for a variety of proteins, such as ATIP1 and ATIP2, involved in AT2 signaling pathway, responsible for cell proliferation inhibition, via MAPK inhibition [(142)]. The variant in *MTUS1* (p.Glu999Gly) was rare in both Europe (1.3%) and Tuscany (1.87%), and was unique in the FIGC cohort (Table 12). This variant was classified as VUS and located in a coiled-coil domain, responsible for dimerization of *MTUS1* (Figure 17, Panel B) with AT2R, a pró-apoptotic and anti-proliferative receptor [(143)]. It is possible that dimerization of *MTUS1* with AT2 receptor could not occur in F13 tumor, inhibiting AT1 intercellular response and leading to activation of AT2R pathway independent of ligand/AT2 receptor interaction [(143)] and, consequently, to activation of MAPK pathway and increased proliferation (Figure 17, Panel A).

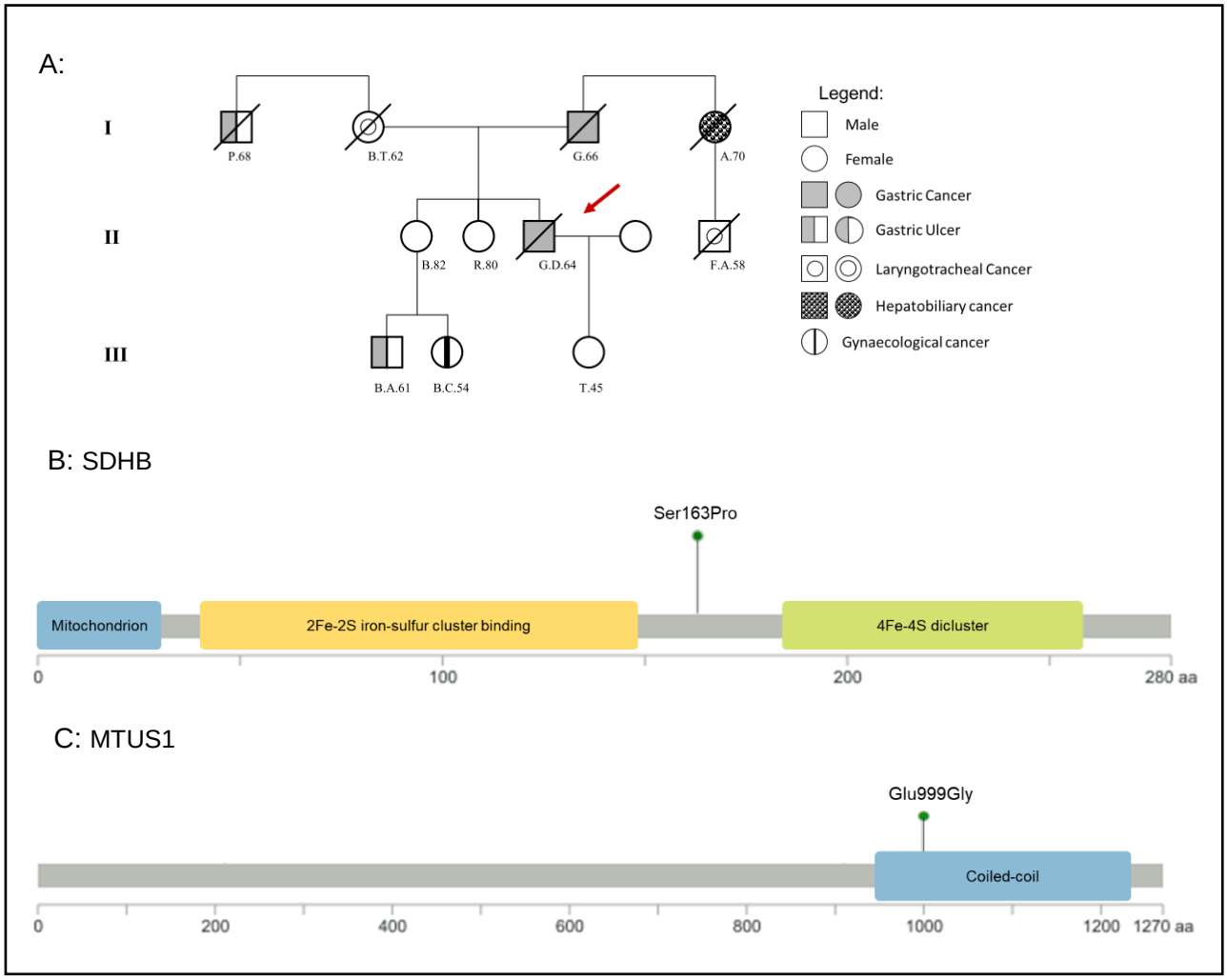


Figure 20. *Germline variants of Family 13. A: Pedigree of family 13. The proband is displayed by an arrow. B: Schematic representation of SDHB protein, its domains, interacting molecules and germline variant found represented by a green dot (missense variant). aa: amino acids [57,58]. C: Schematic representation of MTUS1 protein, its domains, interacting molecules and germline variant found represented by a green dot (missense variant) [57,58].*

2.7.2 Somatic Landscape

Somatic variants were not found in the panel of genes analyzed. However, previous analysis of this family, revealed that the tumor is MSS.

2.7.3. Characterization of genetic and epigenetic second hits

Somatic variants were not found in either *SDHB* or *MTUS1* and did not constitute the second inactivation hit. Therefore, methylation analysis in the promoter of *SDHB* was performed.

Although, in the Ensembl genome browser, a CpG island in the *SDHB* was not annotated, primers were designed to cover the CpG island as described in Westermeier, F. et al. [144],

analyzing almost the entire CpG island. As shown in Figure 18, Panel A, the F13 tumor was not methylated in all CpG sites (23), as well as, the negative control (Hela).

Regarding *MTUS1* promoter methylation, technical issues did not allow the analysis of the methylation status in the CpG island of this gene.

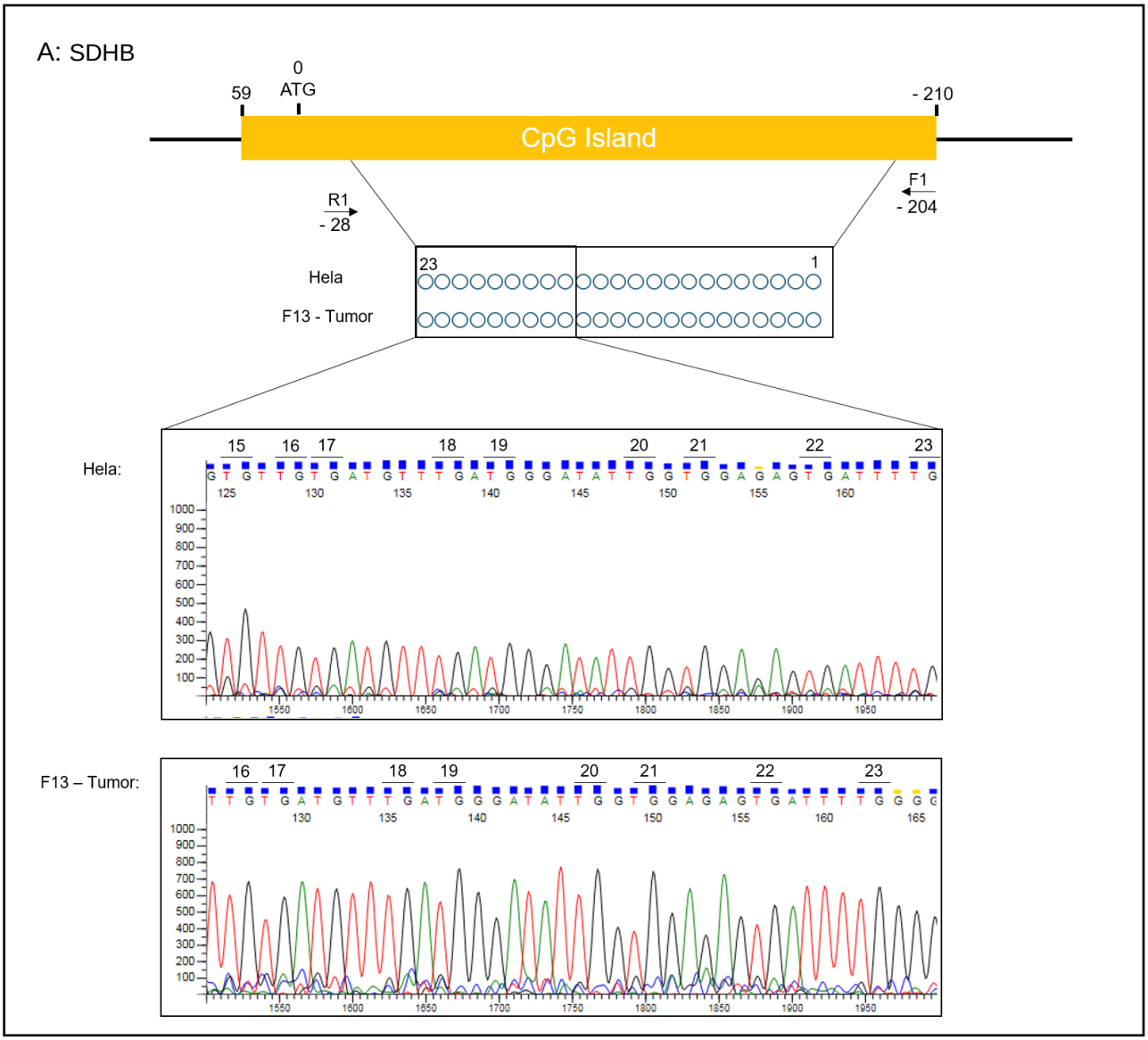


Figure 21. *Somatic Events in Family 13. A: SDHB Promoter Methylation Analysis. Schematic representation of the CpG island located at the SDHB promoter. Open circles represent a non-methylated CpG; Representative examples of sequences of each sample analyzed: Hela cell line (negative control) and F13 tumor.*

MTUS1 was found to be altered in 9% of sporadic GC [57,58]. Additionally, downregulation and association with proliferation and metastasis in gastric cancer cells have been described [145], revealing the importance of this gene in GC. In fact, *SDHB* was found to be mutated in patients with HDGC and associated with AKT and MAPK pathways increased activity. AKT and MAPK pathways have been linked with cell growth, proliferation and

survival [146], important hallmarks of cancer [147]. In fact, MTUS1 and SDHB were found to be somatically mutated in co-occurrence, in only one tumor of a subset of sporadic GCs (0.3% - TCGA data [57,58]). Variant p.Ser163Pro in *SDHB* may be a low or moderate risk allele that predisposes to familial gastric cancer that in association with *MTUS1* variant increases the risk for intestinal type gastric cancer (FIGC).

2.8. Family 14

Family 14 was represented by a female proband (F14) of 74 years old with intestinal-type GC. By analyzing the pedigree of this family, it was observed that the mother of the proband was diagnosed with GC at 52 years old, as well as, an uncle from the mother side, who was diagnosed with GC at 80 years old (Figure 19, Panel A)

2.8.1. Germline Landscape

The F14 harbored two germline variants, in *ATM* and *MTUS1* (Table 12).

The *ATM* variant, p.Arg832Cys, was absent in both Europe and Tuscany populations and unique in this FIGC cohort (2%), being enriched (Table 12). Furthermore, p.Arg832Cys was classified as VUS and was not located in any known domain of the *ATM* gene (Figure 19, Panel A).

Variant p.Ser211Ala in *MTUS1* was classified as VUS and was rare in both Europe (1.7%) and Tuscany (1.87%) (Table 12). Further, this variant was unique in the FIGC cohort (2%) and was not located in any domain discovered so far (Table 12, Figure 19, Panel A).

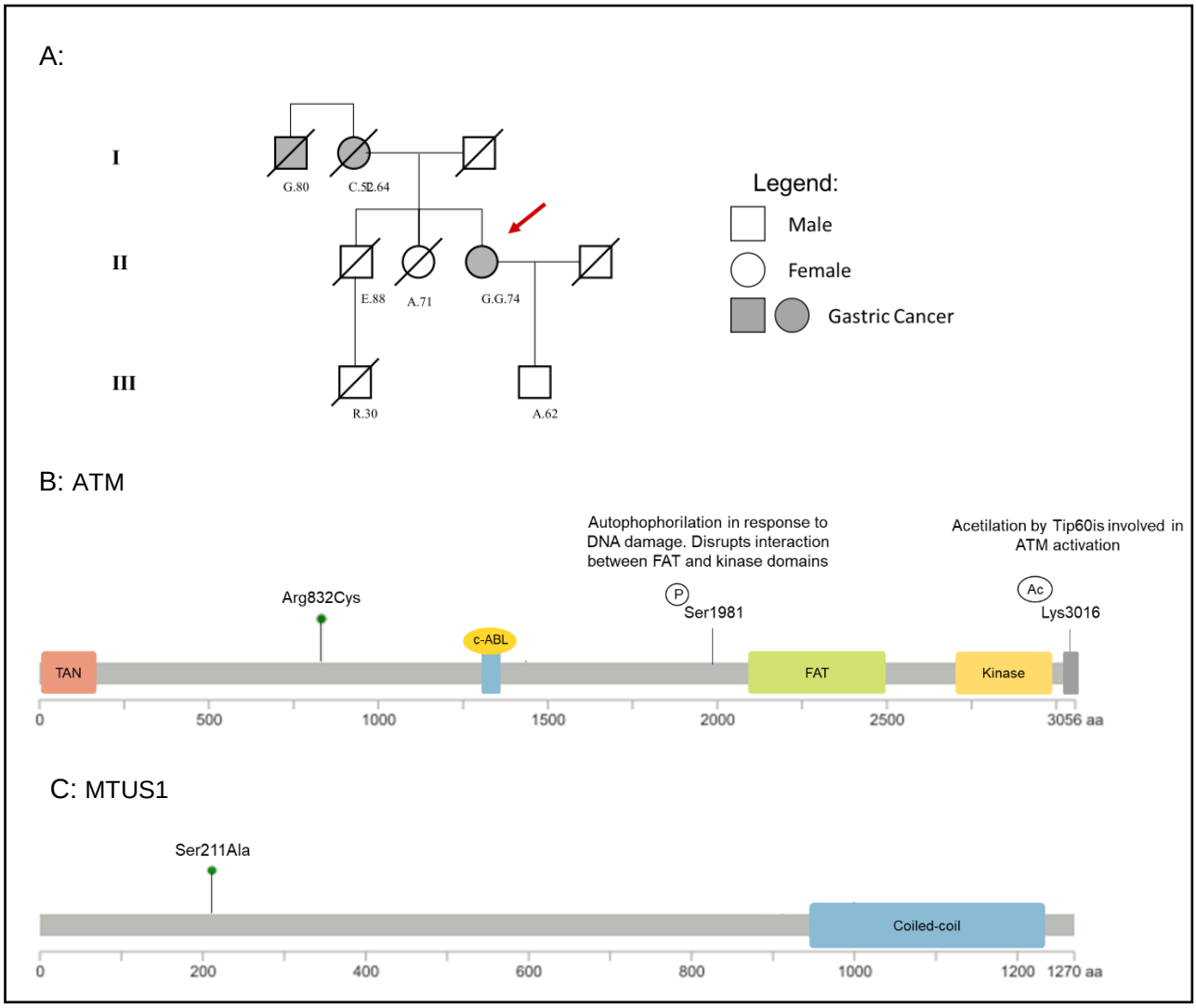


Figure 22. *Germline variants of Family 14. A: Pedigree of family 14. The proband is displayed by an arrow. B: Schematic representation of ATM protein, its domains, interacting molecules and germline variant found represented by a green dot (missense variant). aa: amino acids [57,58]. C: Schematic representation of MTUS1 protein its domains, interacting molecules and germline variant found represented by a green dot (missense variant) [57,58].*

2.8.2 Somatic Landscape

At the somatic level, variants were not found in the panel of genes analyzed. However, previous analysis classified this tumor as MSS.

2.8.3. Characterization of genetic and epigenetic second hits

ATM and *MTUS1* were not found to be mutated in the tumor of F14. Therefore, methylation in the promoter of *ATM* was analyzed to determine if methylation could be the somatic second hit. Using the same strategy as in F11, tumor of F14 was not methylated, CpG sites 1 to 27 (Figure 20, Panel A).

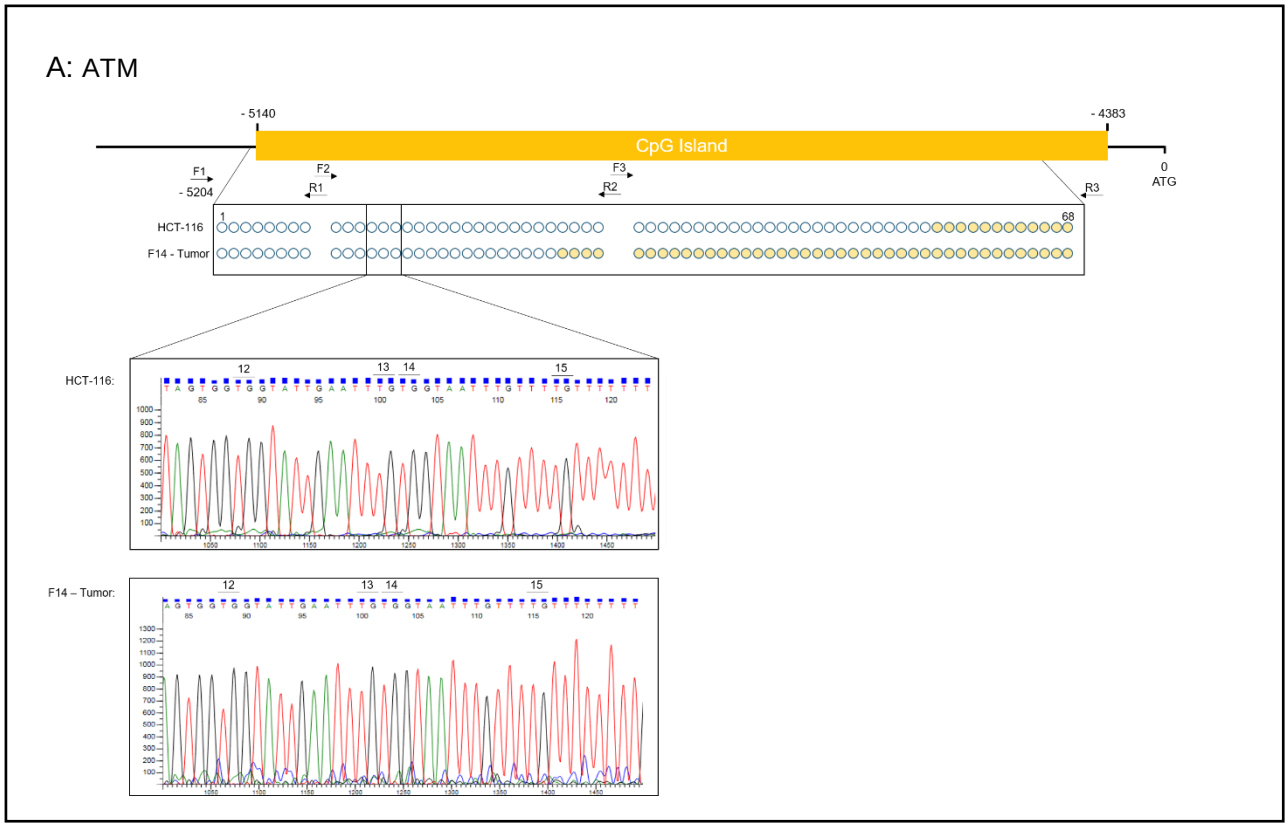


Figure 23. *Somatic Events in Family 14. A: ATM Promoter Methylation Analysis. Schematic representation of the CpG island located at the ATM promoter. Open circles represent a non-methylated CpG and yellow circles represent not accessed CpG sites; Representative examples of sequences of each sample analyzed: HCT-116 cell line (negative control) and F14 tumor.*

ATM and *MTUS1* have been reported to be frequently altered in sporadic GC, being 16% and 9%, respectively [57,58]. Furthermore, according to TCGA database [57,58], 7 (2.4%) sporadic GC tumors had co-occurrence of mutations in ATM and MTUS1. In fact, *MTUS1* downregulation has been associated with proliferation and metastasis in solid tumors, such as gastric cancer [145]. In co-occurrence with *ATM* variant, important gene in repairing double strand breaks, *MTUS1* could predispose to FIGC in this family.

2.9. Family 16

Family 16 was represented by a proband (F16) of the male sex with 78 years old and intestinal-type GC. It was observed that the proband's brother was diagnosed with GC at 97 years old. The wife of the proband had GC at 55 years old, as well as, the wife's sister of the proband, at 70 years old.

2.9.1. Germline Landscape

F16 had three germline variants in *MSH2*, *TGFβR2* and *FAT*4 (Table 12).

Variant Asn596Ser in *MSH2* was very rare globally (0.02%) and classified as Conflicting, since contradictory information was found in the three databases analyzed (Table 12). According to InSiGHT database [123,124], this variant was classified as a class III variant with uncertain significance, due to insufficient evidence. MSH2 protein expression was evaluated in Lynch Syndrome patients with p.Asn596Ser germline variant and revealed to be absent in 2 tumors [148] and present in 1 [149]. Therefore, it is not clear if variant p.Asn596Ser has an impact in protein expression. Furthermore, this variant was unique in the FIGC cohort and located in the MutS domain III, region responsible for the interaction with *MSH6* and *MSH3* [118], and may impact the heterodimer formation and, consequently, the repair in microsatellite regions (Table 12, Figure 21, Panel A)

TGFβR2 encodes for a serine/threonine protein kinase that forms a complex with TGFβR1 and, upon TGFβ binding is able to regulate the transcription of genes associated with cell proliferation, through a signaling pathway [150]. Variant p.Arg215Cys was unique in the FIGC cohort and very rare globally (0.0008%), being classified as VUS (Table 12). Additionally, this variant was not located in any known domain, although present in the cytoplasmatic region of the protein (Figure 21, Panel B).

FAT4 belongs to the proto-cadherin family, responsible for adhesion and may regulate planar cell polarity (PCP), as well as, hippo signaling pathway [151]. Variant in *FAT4* was very rare in Europe (0.5%) and was classified as Likely Benign (Table 12). Variant p.Arg175Leu was unique in this cohort and located in the first cadherin domain (Table 12, Figure 21, Panel C)

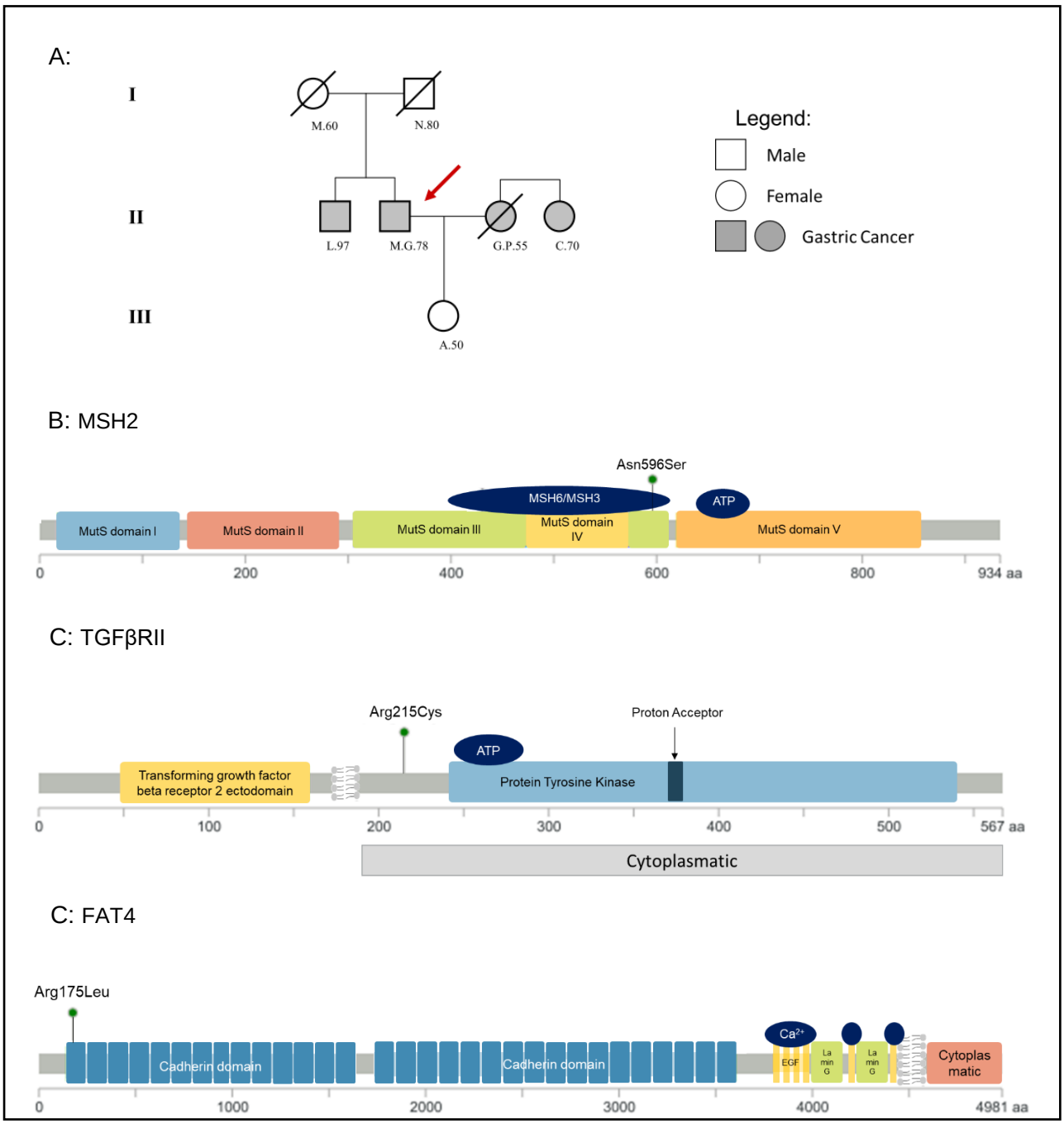


Figure 24. *Germline variants of Family 16. A: Pedigree of family 16. The proband is displayed by an arrow. B: Schematic representation of MSH2 protein, its domains, interacting molecules and germline variant found represented by a green dot (missense variant). aa: amino acids [57,58]. C: Schematic representation of TGFβRII protein, its domains, interacting molecules and germline variant found represented by a green dot (missense variant) [57,58]. C: Schematic representation of FAT4 protein, its domains, interacting molecules and germline variant found represented by a green dot (missense variant) [57,58].*

2.9.2. Somatic Landscape

At the somatic level, six variants were found in *ARID1A, FAT4, MSH3, ATM* and two variants in *PTEN* (Figure 22, Panel A). Additionally, promoter of *MLH1* and *CDH1* was methylated and the MSI phenotype was present in this tumor. *ARID1A, MSH3* and a *PTEN* variant were

targets of the MSI phenotype, being located in microsatellite regions. In fact, MSI tumors do not usually carry TP53 somatic mutations, as observed in this case.

2.9.3. Characterization of genetic and epigenetic second hits

According to section 2.9.2, a second mutation was found in *FAT4*, p.Thr3616Met (Figure 22, Panel A). This variant, which is novel, was located immediately after the last cadherin domain. Due to the lack of information, it may only be speculated that this somatic variant may work as a second hit (Figure 22, Panel B).

Despite this finding, methylation status was assessed in *MSH2, TGFβR2* and *FAT4*.

Methylation in the promoter of *MSH2* was performed as described above. It was observed that F16 tumor was not methylated before and after ATG (CpG sites 1 to 51 and 53 to 66), Figure 22, Panel C.

Promoter methylation of *TGFβR2* was assessed through the design of two sets of primers. F16 tumor, CpG sites 1 to 9 and 20 to 40 is not methylated in the region analyzed, as well as, MCF7 cell line, negative control (Figure 22, Panel D).

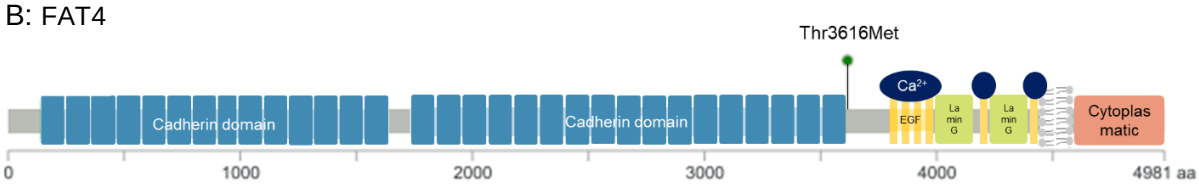
Regarding *FAT4* promoter methylation, three sets of primers were designed based on Yoshida S et al and Pilehchian Langroudi M et al [40,152]. In the first set of primers, NCI-N87 cell line was not methylated. In fact, there was a difference between normal and tumor tissue, being the first not methylated and the second hemi-methylated. To compare with normal stomach, methylation status was analyzed in an individual who underwent bariatric surgery, being not methylated, as well. Regarding the second set of primers, NCI-N87 was methylated and no differences in methylation was observed between normal, F16 tumor and bariatric surgery, being hemi-methylated. The same was observed in the third set of primers (Figure 22, Panel E). However, FAT4 promoter hypermethylation was found in sporadic gastric cancer for the same region [40], in accordance to the methylation pattern disclosed by the first set of primers.

In fact, differences between normal and tumor tissue has only been observed in *FAT4* CpG island region amplified by the first set of primers. However, NCI-87, F16 normal tissue and bariatric surgery sequencing was controversial. Sequencing with primer forward revealed to be hemi-methylated and with primer reverse were not methylated. Nevertheless, F16 tumor was hemi-methylated with both primers. This suggests that methylation analysis of *FAT4* CpG island is contradictory for this region, therefore, second mutation might be the second inactivation hit in F16.

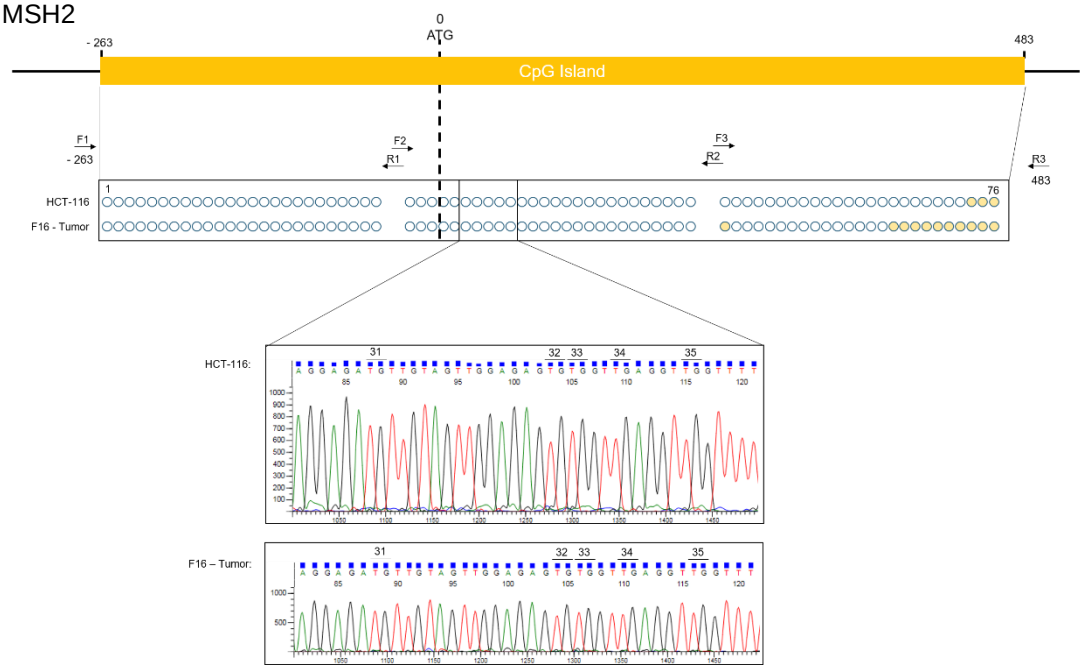
A

Gene	Chr	Position GRCh38	cDNA	Protein	Consequence	Freq cohort
ARID1A	1	26771131	c.3216delA	p.Lys1072AsnfsTer21	Frameshift Indels	1
FAT4	4	125451167	c.10151C>T	p.Ala3384Val	Missense	1
MSH3	5	80675096	c.1148delA	p.Lys383ArgfsTer32	Frameshift Indels	13
PTEN	10	87933147	c.388C>T	p.Arg130Ter	Stop gained	1
PTEN	10	87961042	c.955_958delACTT	p.Thr319Ter	Frameshift Indels	1
ATM	11	108244846	c.721A>T	p.Lys241Ter	Stop gained	1

B: FAT4



C: MSH2



D: TFGβRII

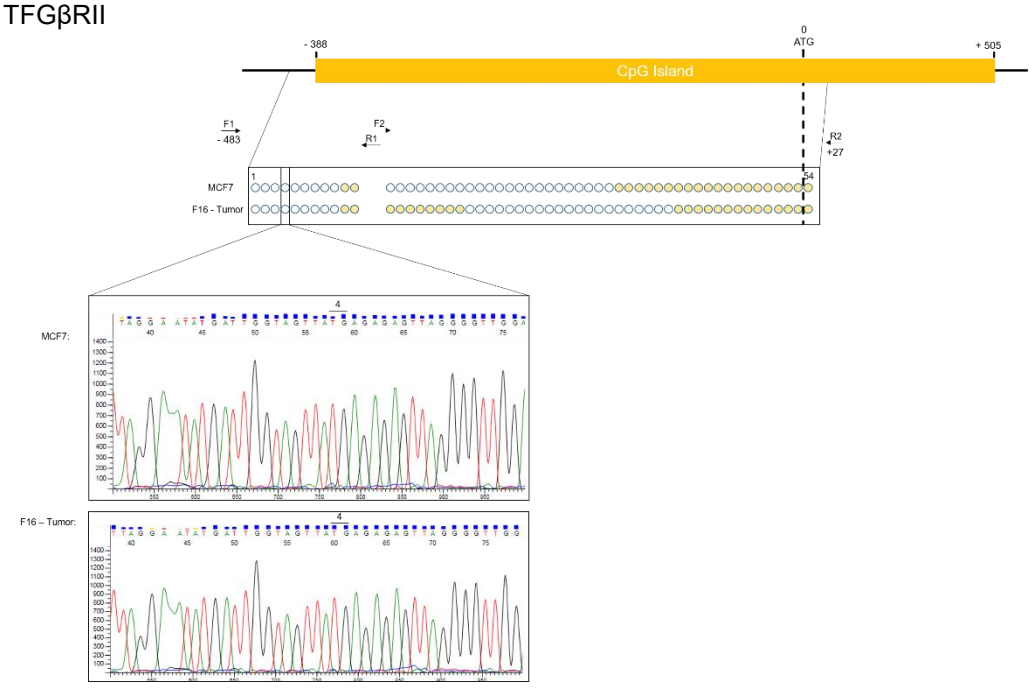


Figure 25. *Somatic Events in Family 16. A: Somatic variants found in family 16 by multiplex custom-panel based sequencing. B: Schematic representation of FAT4 protein, its domains, interacting molecules and somatic variant found represented by a green dot (missense variant). aa: amino acids [57,58]. C: MSH2 Promoter Methylation Analysis. Schematic representation of the CpG island located at the MSH2 promoter. Open circles represent a non-methylated CpG; Representative examples of sequences of each sample analyzed: HCT-116 cell line (negative control) and F16 tumor. D: TGFβRII Promoter Methylation Analysis. Schematic representation of the CpG island located at the TGFβRII promoter. Open circles represent a non-methylated CpG and yellow circles represent not accessed CpG sites; Representative examples of sequences of each sample analyzed: MCF7 cell line (negative control) and F16 tumor.*

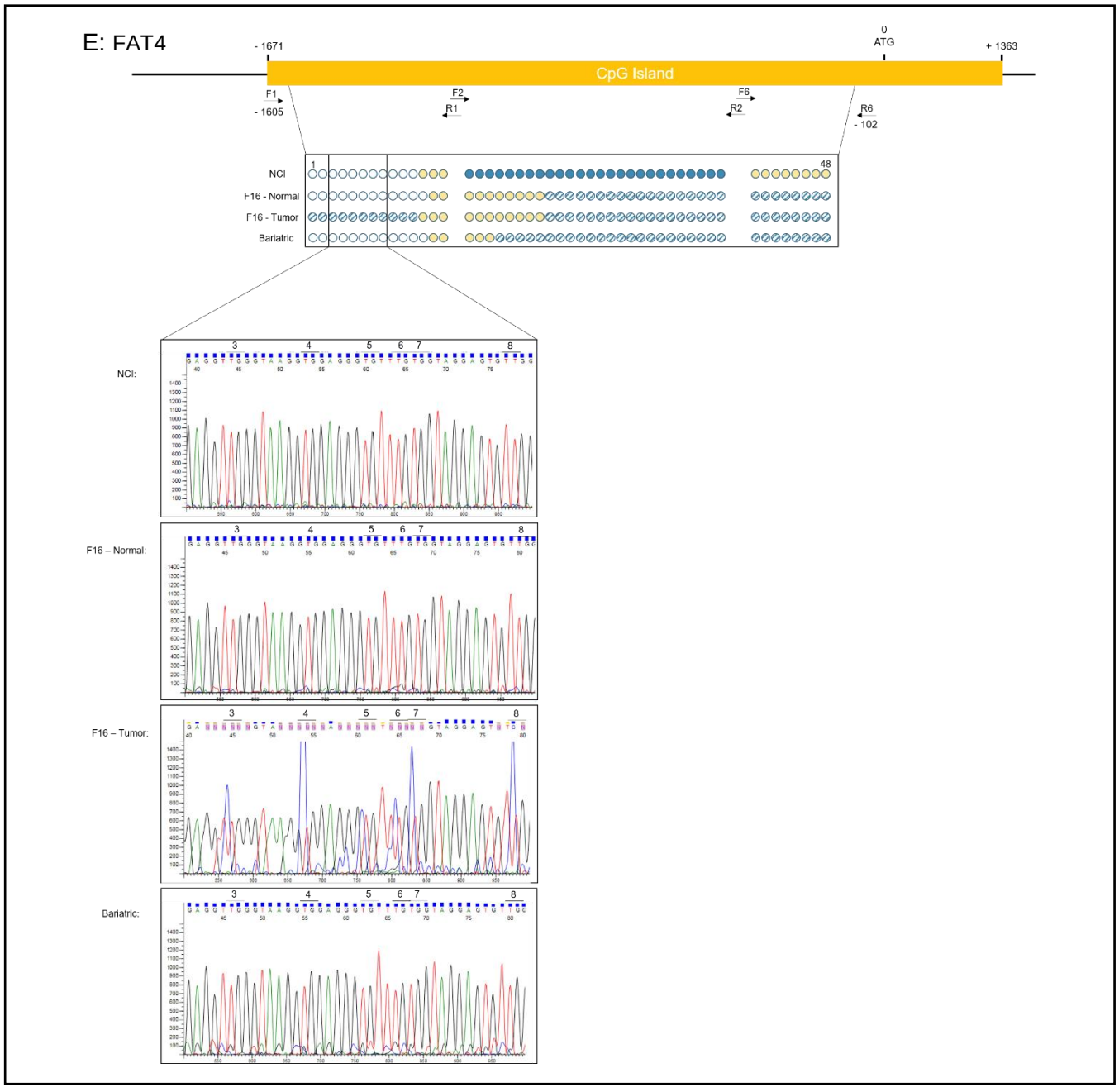


Figure 26 (cont.). *Somatic Events in Family 16. E: FAT4 Promoter Methylation Analysis. Schematic representation of the CpG island located at the FAT4 promoter. Open circles represent a non-methylated CpG blue circles represent methylated CpG, white circles with blue risks represent hemi-methylated CpG and yellow circles represent not accessed CpG sites; Representative examples of sequences of each sample analyzed: NCI cell line (negative control), F16 normal, F16 tumor and an individual who underwent a bariatric surgery.*

In summary, this family harbored a germline and a somatic variant in *FAT4*, suggesting that both variants could lead to a certain level of inactivation of the FAT4 protein. In fact, recent exome studies in sporadic GC [55] revealed *FAT4* mutations throughout the gene, particularly in the EGF domain. These results were confirmed by TCGA sporadic GC [57,58], in which *FAT4* is altered in 24% of the cases in the complete gene. Furthermore, according to the TCGA database [57,58], *TGFβRII* is altered in 7% of individuals with sporadic GC and in 8 individuals (3%) with co-occurring variants in MSH2. Moreover, a germline variant in *MSH2* was found in the binding domain of *MSH3* and *MSH6* and a somatic variant was present in its partner *MSH3*, which could lead to an impairment of the pathway, explaining the MSI phenotype. In fact, *MSH2* is altered in 2.4% of sporadic TCGA GC cases, in which 2 individuals (approximately 1%) display co-occurrence of mutations in *FAT4* and *MSH2* [57,58]. Overall, these results suggest that the co-occurrence of germline variants in *FAT4, MSH2* and *TGFβRII* might predispose this family to the development of FIGC.

2.10. Family 19

Family 19 was represented by a proband (F19) of 82 years old and male sex with intestinal-type GC and colon cancer. Analyzing the pedigree, it was observed that one of his sisters had colorectal cancer at the age of 70 and other had GC at 39 years old (Figure 23, Panel A).

2.10.1. Germline Landscape

At the germline level, two variants were found in F19 in genes *MAP3K6* and *MSR1* (Table 12).

Variant p.Pro1161Ala in *MAP3K6* was classified as VUS and was very rare globally (0.002%). This variant was unique in this FIGC cohort (2%), being enriched and located immediately before the second coiled-coil domain (Table 12, Figure 23, Panel B).

MSR1 gene encodes for a receptor that mediates inflammation [137]. The variant in *MSR1* (p.Arg293X) originates a premature stop codon, was very rare in Tuscany (0.93%), unique in the FIGC cohort (2%), being enriched and classified as conflicting (Table 12). In fact, this variant was located in the highly conserved Collagen Triple Helix Repeat domain (Figure 23, Panel C), being expected to disrupt the function of the protein [69,137]. Additionally, this variant was described previously in four individuals with HDGC [69].

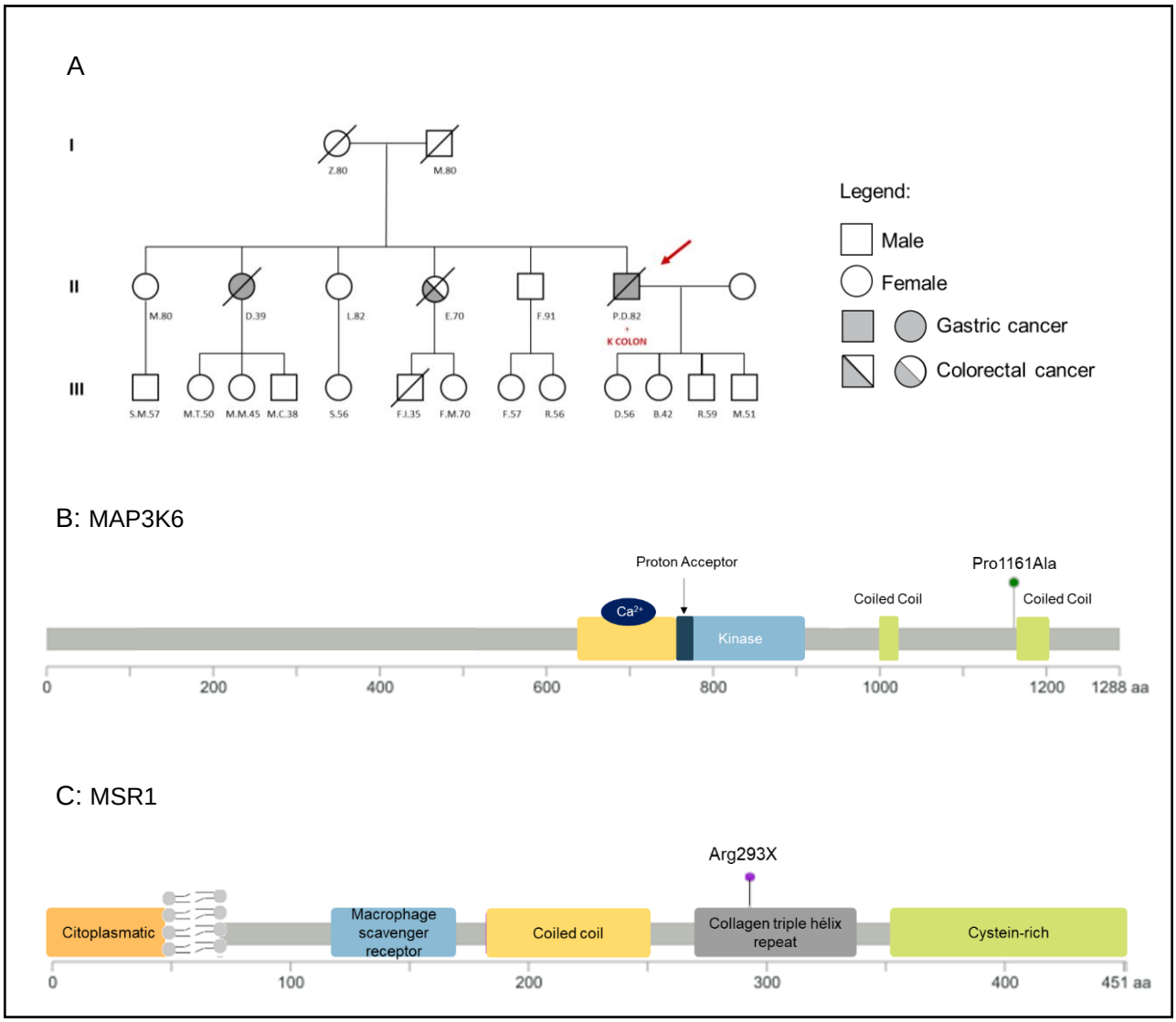


Figure 27. *Pedigree and germline variants of Family 21. A: Pedigree of family 21. The proband is displayed by an arrow. B: Schematic representation of MAP3K6 protein, its domains, interacting molecules and germline variant found represented by a green dot (missense variant). aa: amino acids [57,58]. C: Schematic representation of MSR1 protein, its domains, interacting molecules and germline variant found represented by a green dot (missense variant) [57,58].*

2.10.2 Somatic Landscape

The tumor of this family had three variants in *ARID1A, MSH6* and *MSH3* (Figure 24, Panel A). Additionally, this tumor presented a MSI phenotype and methylation of *MLH1*. Promoter methylation of *MLH1* and the absence of *TP53* mutation in F19 tumor are characteristic of MSI tumors. In fact, *MSH6* and *MSH3* variants are targets of MSI phenotype, due to the microsatellite regions in which the variants are contained.

2.10.3. Characterization of genetic and epigenetic second hits

Somatic variants in *MAP3K6* and *MSR1* were not found, therefore, promoter methylation of these genes was assessed to determine if methylation of the wild-type allele could be the somatic second inactivation hit.

The methylation results obtained for the promoter and gene body of *MAP3K6* were similar to those obtained for F7 and F8 (*i.e* the promoter was not methylated and the gene body was hemi-methylated both in tumor and normal tissues; Figure 23, Panel B and C). These findings suggest that second mutation and methylation of the promoter and gene body regions of *MAP3K6* might not be the second inactivation hit.

Regarding *MSR1* promoter methylation, technical difficulties did not allow the analysis of the methylation status, since no CpG island was found to regulate this gene.

A

Gene	Chr	Position GRCh38	cDNA	Protein	Consequence	Freq cohort
ARID1A	1	26779533	c.5635C>T	p.Arg1879Trp	Missense	1
MSH6	2	47803500	c.3261dupC	p.Phe1088LeufsTer5	Frameshift Indels	2
MSH3	5	80675096	c.1148delA	p.Lys383ArgfsTer32	Frameshift Indels	13

B: MAP3K6

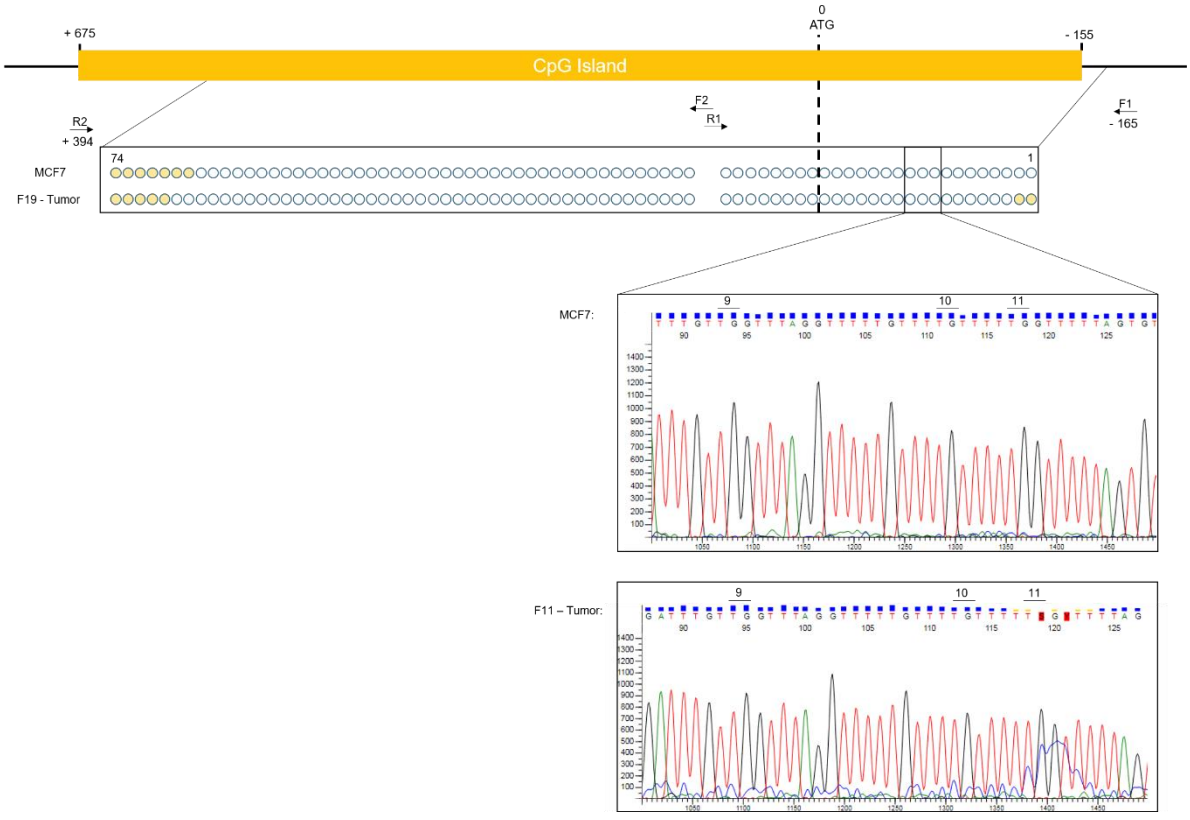


Figure 28. *Somatic Events in Family 19. A: Somatic variants found in family 19 by multiplex custom-panel based sequencing. B: MAP3K6 Promoter Methylation Analysis. Schematic representation of the CpG island located at the MAP3K6 promoter. Open circles represent a non-methylated CpG and yellow circles represent not accessed CpG sites; Representative examples of sequences of each sample analyzed: MCF7 cell line (negative control)*

and F19 tumor. C: MAP3K6 gene body Methylation Analysis. Schematic representation of the CpG island located at the MAP3K6 gene body. Open circles represent a non-methylated CpG; blue circles represent methylated CpG, white circles with blue risks represent hemi-methylated CpG and yellow circles represent not accessed CpG; Representative examples of sequences of each sample analyzed: MKN74 cell line (negative control), F19 normal, F19 tumor and sample of bariatric surgery.

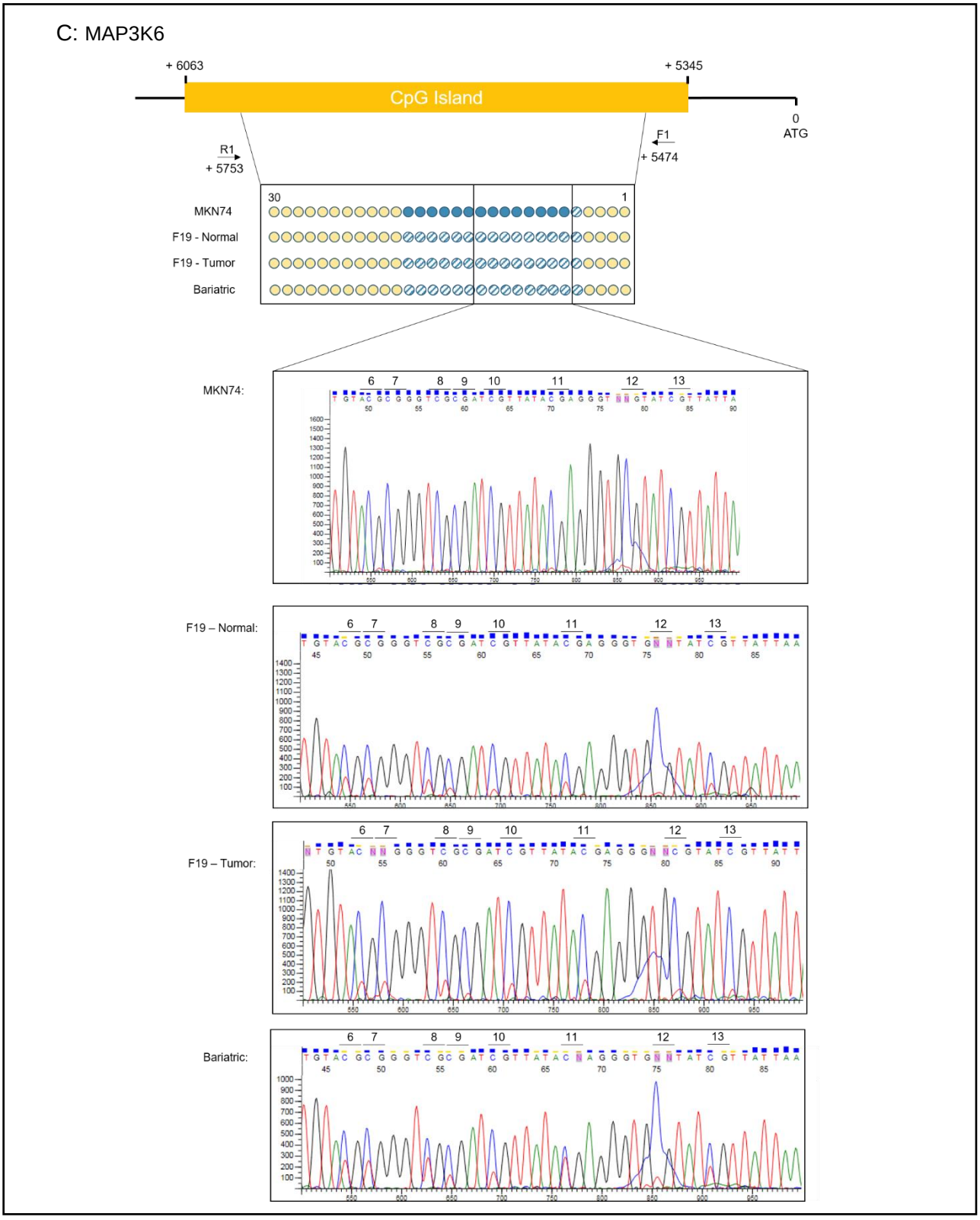


Figure 29 (cont.). *Somatic Events in Family 19. C: MAP3K6 gene body Methylation Analysis. Schematic representation of the CpG island located at the MAP3K6 gene body. Open circles represent a non-methylated*

CpG; blue circles represent methylated CpG, white circles with blue risks represent hemi-methylated CpG and yellow circles represent not accessed CpG; Representative examples of sequences of each sample analyzed: MKN74 cell line (negative control), F19 normal, F19 tumor and sample of bariatric surgery.

MSR1 variant R293X has been identified in patients with HDGC [(69)] and Barrett Esophagus and Esophageal Adenocarcinoma [(137)], suggesting a correlation with the presence of this variant and susceptibility to gastrointestinal tract cancers. In fact, *MSR1*, as a macrophage scavenger receptor, has been associated with inflammation [(137)], a hallmark of Intestinal GC [(50)]. Furthermore, *MSR1* is altered in 9% of sporadic GC individuals depicted in the TCGA database, with co-occurrence with *MAP3K6* in 2 individuals (approximately 1%) [(57,58)]. Thus, suggesting that co-occurrence of germline variants in MSR1 and *MAP3K6*, a gene associated with familial gastric cancer [(31)], could predispose to FIGC in this family.

After a detailed analysis of FIGC families carrying more than one germline variant, characterization of FIGC families that harbored single germline variants was performed. It is important to state that the possibility of occurrence of other germline variants affecting genes that were not included in the custom-panel could not be excluded.

Analysis of Families with single germline variants

It was observed that 14 out of the 24 families carrying germline variants, displayed a single variant (i.e. one variant in a single gene). These 14 germline variants were re-classified as: conflicting (5), likely benign (2), and VUS (7), using the Annotator57's software (Table 14).

The most frequently mutated gene was *MSH6* (in 3 families). Interestingly, these families (F9, 10 and 18) harbored exactly the same variant - p.Val878Ala -, which was located in MutS domain III and already detected and classified in F11 (section 2.5.1.). In addition, F9, F10 and F18 families displayed a similar somatic landscape, carrying variants in *TP53* (F9 and F10) and *ARID1A* (F10 and F18). Interestingly, F18 also carried a somatic mutation in *MSH3* gene, which may indicate a defect on the MMR pathway, since both *MSH6* and *MSH3* bind to *MSH2* (Table 14, Figure 28, 29, 30).

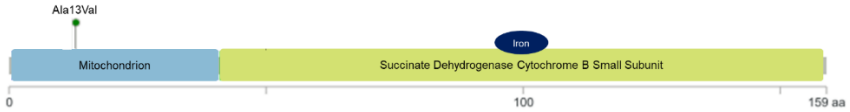
Other families were found to display single germline variants, such as: *SDHD* p.Ala13Val in family 3 (F3) and *SDHD* p.Gly12Ser in family 21 (F21); *ITIH2* p.Ala547Val in family 6 (F6); *SDHB* p.Ser163Pro in family 22 (F22); and *MSR1* p.Thr161Asn in family 24 (F24). The somatic landscape of these families revealed variants in *TP53* (F2, F3, F21 and F22), in *BRCA2* (F3), and in *FAT4* (F22) (Table 14, Figure 25, 26, 27, 31, 32, 33).

Focusing on the somatic second hit, second mutation and promoter methylation was not found in these families.

F2

Germline

A: SDHD



Somatic

B

Gene	Chr	Position GRCh38	cDNA	Protein	Consequence
TP53	17	7674953	c.578A>T	p.His193Leu	Missense

Figure 30. *Germline variant and somatic variants of F2. A: Schematic representation of SDHD protein, its domains, interacting molecules and germline variant found represented by a green dot (missense variant); B: Somatic variants in F2.*

F3

Germline

MAP3K6 splice acceptor

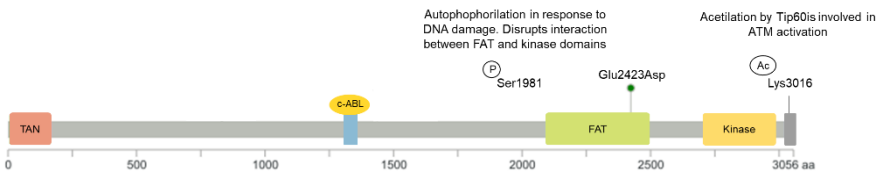
Somatic

Gene	Chr	Position GRCh38	cDNA	Protein	Consequence
BRCA2	13	32332779	c.1310_1313delAAGA	p.Lys437IlefsTer22	Frameshift Indels
TP53	17	7673764	c.856G>T	p.Glu286Ter	Stop gained

Figure 31. *Somatic variants found in F3.*

F5

Germline



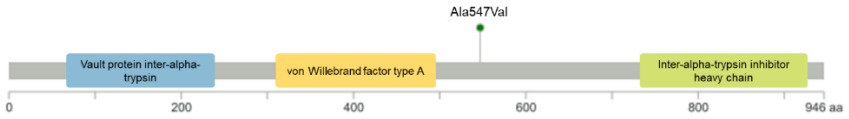
Somatic

-

Figure 32. *Schematic representation of ATM protein, its domains, interacting molecules and germline variant found represented by a green dot (missense variant). Somatic variants were not found.*

F6

Germline



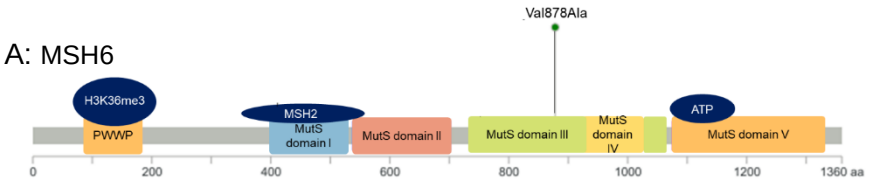
Somatic

-

Figure 33. *Schematic representation of ITIH2 protein, its domains, interacting molecules and germline variant found represented by a green dot (missense variant). Somatic variants were not found.*

F9

Germline



Somatic

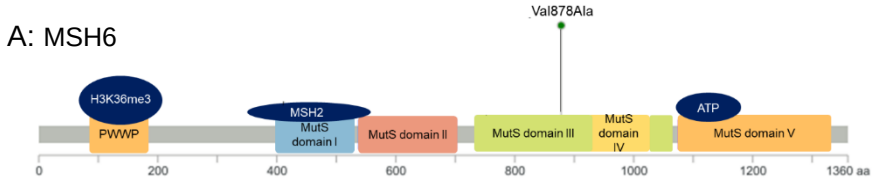
B

Gene	Chr	Position GRCh38	cDNA	Protein	Consequence
TP53	17	7674238	c.725G>T	p.Cys242Phe	Missense

Figure 34. *Germline and somatic variants of F9. A: Schematic representation of MSH6 protein, its domains, interacting molecules and germline variants found represented by a green dot (missense variant); B: Somatic variants in F9.*

F10

Germline



Somatic

B

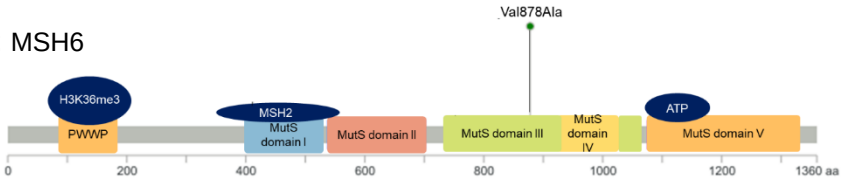
Gene	Chr	Position GRCh38	cDNA	Protein	Consequence
ARID1A	1	26780111	c.6214delG	p.Asp2072ThrfsTer63	Frameshift Indels
CTNNA1	5	138824612	c.671delC	p.Ala224AspfsTer20	Frameshift Indels
TP53	17	7674221	c.742C>T	p.Arg248Trp	Missense

Figure 35. *Germline and somatic variants of F10. A: Schematic representation of MSH6 protein, its domains, interacting molecules and germline variant found represented by a green dot (missense variant); B: Somatic variants in F10.*

F18

Germline

A: MSH6



Somatic

B

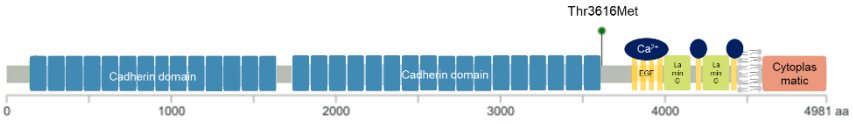
Gene	Chr	Position GRCh38	cDNA	Protein	Consequence
ARID1A	1	26774926	c.4703delC	p.Pro1568LeufsTer44	Frameshift Indels
MSH3	5	80675096	c.1148delA	p.Lys383ArgfsTer32	Frameshift Indels

Figure 36. *Germline and somatic variants of F18. A: Schematic representation of MSH6 protein, its domains, interacting molecules and germline variant found represented by a green dot (missense variant); B: Somatic variants in F18.*

F20

Germline

A: FAT4



Somatic

B

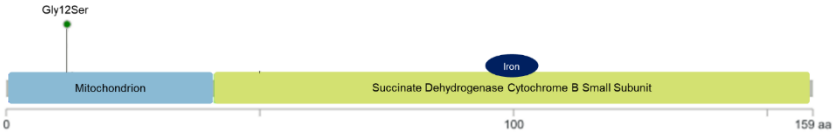
Gene	Chr	Position GRCh38	cDNA	Protein	Consequence
TP53	17	7675228	c.380_384delCCCCT	p.Ser127CysfsTer20	Frameshift Indels

Figure 37. *Germline and somatic variants of F20. A: Schematic representation of FAT4 protein, its domains, interacting molecules and germline variant found represented by a green dot (missense variant); B: Somatic variants in F20.*

F21

Germline

A: SDHD



Somatic

B

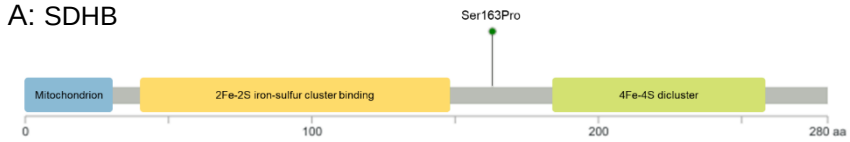
Gene	Chr	Position GRCh38	cDNA	Protein	Consequence
TP53	17	7674250	c.713G>T	p.Cys238Phe	Missense

Figure 38. *Germline and somatic variants of F21. A: Schematic representation of SDHD protein, its domains, interacting molecules and germline variant found represented by a green dot (missense variant); B: Somatic variants in F21.*

F22

Germline

A: SDHB



Somatic

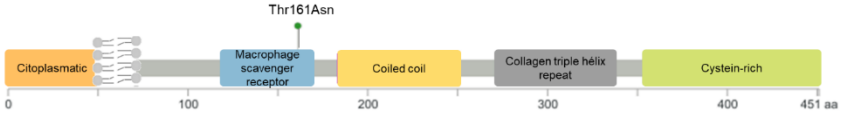
B

Gene	Chr	Position GRCh38	cDNA	Protein	Consequence
FAT4	4	125450920	c.9904C>A	p.Leu3302Ile	Missense
TP53	17	7674208	c.743_755delGGAGGCCCATCCT	p.Arg248ProfsTer93	Frameshift Indels

Figure 39. *Germline and somatic variants of F22. A: Schematic representation of SDHB protein, its domains, interacting molecules and germline variant found represented by a green dot (missense variant); B: Somatic variants in F21.*

F24

Germline



Somatic

-

Figure 40. *Schematic representation of MSR1 protein, its domains, interacting molecules and germline variant found represented by a green dot (missense variant). Somatic variants were not found*

Table 14. *Germline Variants identified in FIGC families with single germline variants by multiplex custom-panel based sequencing.*

Family	Gene	Chr	Position GRCh38	cDNA	Protein	Consequence	Freq cohort	Population Frequency	Illumina's Classification	Annotator57's Classification	Integrated Classification	Potential Risk Allele
2	SDHD	11	112086945	c.38C>T	p.Ala13Val	Missense	2%	Very Rare Global (0,006%), Very Rare Eur (0%)	VUS	NA	VUS	Yes
3	MAP3K6	1	27362252	c.1256-2A>G	-	Splice acceptor	4%	Very Rare Global (0,4%), Rare Eur (1,4%), Very Rare Tusc (0,9%)	Likely Pathogenic	NA	VUS	Yes
5	ATM	11	108329200	c.7269A>T	p.Glu2423Asp	Missense	2%	NA	VUS	Uncertain Significance - 1	VUS	Yes
6	ITIH2	10	7731989	c.1640C>T	p.Ala547Val	Missense	2%	Very Rare Global (0,02%), Very Rare Eur (0,004%)	VUS	NA	VUS	Yes
9	MSH6	2	47800616	c.2633T>C	p.Val878Ala	Missense	8%	Very Rare Global (0,8%), Rare Eur (2,6%), Rare Tusc (4,7%)	Pathogenic	Likely Pathogenic - 2; Likely Benign - 1	Conflicting	Yes
10	MSH6	2	47800616	c.2633T>C	p.Val878Ala	Missense	8%	Very Rare Global (0,8%), Rare Eur (2,6%), Rare Tusc (4,7%)	Pathogenic	Likely Pathogenic - 2; Likely Benign - 1	Conflicting	Yes
15	MTUS1	8	17684434	c.2732A>C	p.Lys911Thr	Missense	2%	Very Rare Global (0,9%), Rare Eur (1,8%), Very Rare Tusc (0,9%)	VUS	Likely Benign - 1	Likely Benign	Yes
17	APC	5	112839078	c.3484T>C	p.Tyr1162His	Missense	2%	NA	VUS	NA	VUS	Yes
18	MSH6	2	47800616	c.2633T>C	p.Val878Ala	Missense	8%	Very Rare Global (0,8%), Rare Eur (2,6%), Rare Tusc (4,7%)	Pathogenic	Likely Pathogenic - 2; Likely Benign - 1	Conflicting	Yes
20	FAT4	4	125451863	c.10847C>T	p.Thr3616Met	Missense	2%	Very Rare Global (0,3%), Very Rare Eur (0,6%), Very Rare Tusc (0,9%)	VUS	Likely Benign - 1	Likely Benign	Yes
21	SDHD	11	112086941	c.34G>A	p.Gly12Ser	Missense	4%	Very Rare Global (1%), Rare Eur (2%), Rare Tusc (4,7%)	Pathogenic	Likely Pathogenic - 2; Uncertain Significance - 1	Conflicting	Yes
22	SDHB	1	17027802	c.487T>C	p.Ser163Pro	Missense	4%	Rare Global (1,9%), Rare Eur (3,4%), Rare Tusc (3,7%)	Pathogenic	Likely Pathogenic – 2; Uncertain Significance - 1	Conflicting	Yes
23	CDH1	16	68813323	c.1148A>G	p.Gln383Arg	Missense	2%	NA	VUS	NA	VUS	Yes
24	MSR1	8	16168606	c.482C>A	p.Thr161Asn	Missense	2%	Very Rare Global (0,08%), Very Rare Eur (0,2%) Very Rare Tusc (0,9%)	VUS	NA	VUS	Yes
25-52	Germline variants were not found											

Description of germline variants, including both frequency in the cohort (freq cohort: % of mutated families), ExAC and 1000genomes and Clinical Significance classification from Illumina's Variant Interpreter, Annotator (in 3 databases) and integrated classification. Eur: Europe; Tusc: Tuscany; VUS: Variant of Unknown Significance; NA: Not available; Very Rare: <1%; Rare: >1, <5; Not Rare: >5.

Table 15 represents a characterization of germline variants and corresponding potential somatic second hits.

Table 15. *Characterization of germline variants found in each family and somatic second hit found.*

* * * * * * * * *

Gene	Gene Freq	Alteration	Alt Freq	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12	F13	F14	F15	F16	F17	F18	F19	F20	F21	F22	F23	F24
MSH6	17%	p.Lys412Asn	3%	#																							
		p.Val878Ala	11%																								
		p.Ser753Phe	3%	#																							
SDHD	11%	p.Gly12Ser	6%																								
		p.His50Arg	3%																								
		p.Ala13Val	3%																								
ATM	8%	p.Glu2423Asp	3%																								
		p.His1436Tyr	3%																								
		p.Arg832Cys	3%																								
MSH2	6%	c.1275A>G(p.=)	3%																								
		p.Asn596Ser	3%																								
TGFBR2	3%	p.Arg215Cys	3%																								
FAT4	6%	p.Arg175Leu	3%																#								
		p.Thr3616Met	3%																								
MAP3K6	11%	c.1256-2A>G	6%																								
		p.Pro946Leu	3%																								
		p.Pro1161Ala	3%																								
BRCA1	3%	c.-20+2dupT	3%																								
CASP10	3%	p.Cys169Phe	3%																								
MSR1	6%	p.Arg293Ter	3%																								
		p.Thr161Asn	3%																								
MTUS1	8%	p.Lys911Thr	3%																								
		p.Glu999Gly	3%																								
		p.Ser211Ala	3%																								
SDHB	6%	p.Ser163Pro	6%																								
BRCA2	3%	p.Thr939SerfsTer7	3%																								
CTHRC1	3%	p.Leu20_Leu21dup	3%																								
ITIH2	3%	p.Ala547Val	3%																								
APC	3%	p.Tyr1162His	3%																								
CDH1	3%	p.Gln383Arg	3%																								

Gene Freq: Frequency of each gene's variants in a total of 36 variants; Alt Freq: Frequency of each variant in a total of 36; Blue: Likely Pathogenic variant; Red: Conflicting Variant; Green: Likely Benign variant; Grey: Variant of Unknown Significance. * represents families with co-occurrence of germline variants.# represents potential second mutation as second inactivating hit.

3. Somatic Landscape of Familial Intestinal Gastric Cancer

Despite the already described similarities between FIGC and Sporadic GC, such as the MSI phenotype and their target genes [59], very little is known concerning the somatic landscape of FIGC tumors. Therefore, the fourth aim of this thesis was to identify other somatic events in FIGC tumors by using a panel-based sequencing approach.

Sequencing data from the tumors of the 52 FIGC families revealed the presence of 115 somatic variants, affecting 23 genes in 36 families (table 16). These 115 somatic variants were classified by the Illumina´s software as: pathogenic (21 variants), likely pathogenic (52 variants), and variants of unknown significance (42 variants) (Table 16). Using Annotator57, following re-classification was obtained: Pathogenic (12 variants), Likely Pathogenic (42 variants), Conflicting (1) and VUS (60 variants) (table 16). Furthermore, 22 out of the 115 variants were very rare in the global population and one very rare variant in the Tuscany population (Table 16). For the remaining 92 information regarding frequency was not found (Table 16). Detailed information on the classification of variants is described in Supplementary Table 2.

Considering the Illumina's Variant Interpreter, 42 variants were classified as VUS. From these, Annotator57 re-classified eight as Likely Pathogenic, one as Conflicting and 33 as VUS, allowing an increased knowledge of the pathogenicity of these variants. Concerning the 52 variants classified with Illumina's Variant Interpreter as Likely Pathogenic, 22 were re-classified as Likely Pathogenic, three as Pathogenic and 27 as VUS. This may be due to the actualization of the databases. Additionally, 21 variants were classified with Illumina's Variant Interpreter as Pathogenic and 9 were re-classified with Annotator57 as Pathogenic and 12 as Likely pathogenic (Table 16).

In Figure 34 is represented an oncoprint, with all the genes somatically altered in each family. In fact, tumor suppressor genes were the most frequently mutated genes, including ***TP53*** **(18%)**, *MSH3* (11%), ***ARID1A*** **(10%)**, ***FAT4*** **(7%)**, *APC* (8%), *MSH6* (7%), *ATM* (7%), *BRCA2* (6%) and *CTNNA1* (6%).

Interestingly, the TCGA research network had published a comprehensive molecular characterization of gastric adenocarcinoma [54], showing that the most frequently mutated genes in intestinal type sporadic GC were ***TP53*** **(54%)**, *LRP1B* (33%), *SYNE1* (31%), ***ARID1A*** **(30%)**, *PCLO* (29%) and ***FAT4*** **(28%)** [54,57,58]. However, the *LRP1B*, *SYNE1* and *PCLO* genes were not included in the custom-panel, which could explain their absence in this cohort. In addition, other studies based on next generation sequencing have also showed *FAT4* and *ARID1A* mutations in sporadic GC [55]. Of notice, variants in the common

genes between FIGC cohort and the TCGA sporadic GC cohort (*TP53, ARID1A, FAT4*) are more frequent in the sporadic intestinal type GC subset, in comparison with the intestinal familial subset. This may be another argument that FIGC is an independent entity with clear differences from sporadic intestinal type GC.

A similar somatic landscape was observed between FIGC families with and without germline variants (Table 17). In fact, *TP53* was the most frequently mutated gene in both type of families (18%), which is not in accordance with frequency of *TP53* mutations found in the sporadic setting (54%) [57,58]. In addition, *ARID1A* (10%) and *APC* (8%) were the second most frequently mutated genes in families with and without germline variants, respectively). Interestingly, *ARID1A* mutations have been described in gastric adenocarcinoma, predominantly intestinal type. Additionally, the majority of families carrying germline variants (63%) harbored more frequently somatic variants in more than one gene in comparison with families without germline variants (35%). This may be due to the fact that the MSI phenotype is more frequent in families with germline variants (50%), than without (32%), Table 16. According to the TCGA database [57,58], in sporadic intestinal-type GC, MSI phenotype is present in 25% of the cases, a reduced number in comparison to our results (MSI is present in 22/52 (42%) families). Thus, the MSI phenotype is more frequent in FIGC tumors than in sporadic intestinal-type GC tumors, not in accordance with previous studies [79,80].

Table 16. *Somatic Variants identified in FIGC families by multiplex custom-panel based sequencing.*

Family	Gene	Chr	Position GRCh38	cDNA	Protein	Consequence	Freq cohort	Frequency	Illumina's Classification	Annotator's Classification	Integrated Classification
1	ARID1A	1	26771139	c.3219G>A	p.Trp1073Ter	Stop gained	2%	-	Likely Pathogenic	NA	VUS
	MSH6	2	47801146	c.3163G>A	p.Ala1055Thr	Missense	2%	Very Rare Global (0,002%)	VUS	Uncertain Significance - 1	VUS
	FAT4	4	125406935	c.5363G>A	p.Arg1788His	Missense	2%	-	VUS	NA	VUS
	AKAP12	6	151351864	c.3479dupC	p.Asp1161Ter	Frameshift Indels	2%	-	Likely Pathogenic	NA	VUS
	ATM	11	108229266	c.278delA	p.Lys93ArgfsTer23	Frameshift Indels	2%	-	Likely Pathogenic	NA	VUS
	BRCA2	13	32339422	c.5073delA	p.Lys1691AsnfsTer15	Frameshift Indels	2%	-	Pathogenic	Likely Pathogenic - 1	Likely Pathogenic
	IDH2	15	90088686	c.435delG	p.Thr146LeufsTer15	Frameshift Indels	2%	-	Likely Pathogenic	NA	VUS
	PALB2	16	23635707	c.839delA	p.Asn280ThrfsTer8	Frameshift Indels	2%	-	Likely Pathogenic	NA	VUS
2	TP53	17	7674953	c.578A>T	p.His193Leu	Missense	2%	-	Likely Pathogenic	Likely Pathogenic - 2	Likely Pathogenic
3	BRCA2	13	32332779	c.1310_1313delAAGA	p.Lys437IlefsTer22	Frameshift Indels	2%	Very Rare Global (0,002%)	Likely Pathogenic	Likely Pathogenic - 1	Likely Pathogenic
	TP53	17	7673764	c.856G>T	p.Glu286Ter	Stop gained	2%	-	Likely Pathogenic	Pathogenic - 1	Pathogenic
4-6	Somatic variants were not found										
7	TP53	17	7675109	c.503A>T	p.His168Leu	Missense	2%	-	VUS	Likely Pathogenic - 1	Likely Pathogenic
8	ARID1A	1	26773716	c.4003C>T	p.Arg1335Ter	Stop gained, Splice region	2%	-	Pathogenic	Likely Pathogenic - 2	Likely Pathogenic
	FAT4	4	125449221	c.8205A>C	p.Lys2735Asn	Missense	2%	-	VUS	NA	VUS
	PTEN	10	87952216	c.595_597delATG	p.Met199del	Inframe deletion	2%	-	VUS	Uncertain Significance - 1	VUS
	PTEN	10	87958013	c.800delA	p.Lys267ArgfsTer9	Frameshift Indels	2%	-	Pathogenic	Pathogenic - 1	Pathogenic
9	TP53	17	7674238	c.725G>T	p.Cys242Phe	Missense	2%	-	VUS	Likely Pathogenic – 2	Likely Pathogenic
10	ARID1A	1	26780111	c.6214delG	p.Asp2072ThrfsTer63	Frameshift Indels	2%	-	Likely Pathogenic	NA	VUS
	CTNNA1	5	138824612	c.671delC	p.Ala224AspfsTer20	Frameshift Indels	2%	-	Likely Pathogenic	NA	VUS
	TP53	17	7674221	c.742C>T	p.Arg248Trp	Missense	4%	Very Rare Global (0,002%)	Pathogenic	Likely Pathogenic - 3	Likely Pathogenic
11	CDH1	16	68819423	c.1709A>G	p.Asn570Ser	Missense, Splice region	2%	-	VUS	NA	VUS
	TP53	17	7674947	c.584T>C	p.Ile195Thr	Missense	2%	Very Rare Global (0,002%)	VUS	Likely Pathogenic - 2	Likely Pathogenic
12	APC	5	112842961	c.7367T>A	p.Leu2456Ter	Stop gained	2%	-	Likely Pathogenic	NA	VUS
13-15	Somatic variants were not found										
16	ARID1A	1	26771131	c.3216delA	p.Lys1072AsnfsTer21	Frameshift Indels	2%	-	Likely Pathogenic	NA	VUS
	FAT4	4	125451167	c.10151C>T	p.Ala3384Val	Missense	2%	-	VUS	NA	VUS
	MSH3	5	80675096	c.1148delA	p.Lys383ArgfsTer32	Frameshift Indels	25%	-	Likely Pathogenic	Likely Pathogenic - 2	Likely Pathogenic
	PTEN	10	87933147	c.388C>T	p.Arg130Ter	Stop gained	2%	Very Rare Global (0,002%)	Pathogenic	Pathogenic - 1; Likely Pathogenic - 1	Likely Pathogenic
	PTEN	10	87961042	c.955_958delACTT	p.Thr319Ter	Frameshift Indels	2%	-	Pathogenic	Pathogenic - 1	Pathogenic
	ATM	11	108244846	c.721A>T	p.Lys241Ter	Stop gained	2%	-	Likely Pathogenic	NA	VUS
17	Somatic variants were not found										
18	ARID1A	1	26774926	c.4703delC	p.Pro1568LeufsTer44	Frameshift Indels	2%	-	Likely Pathogenic	NA	VUS
	MSH3	5	80675096	c.1148delA	p.Lys383ArgfsTer32	Frameshift Indels	25%	-	Likely Pathogenic	Likely Pathogenic - 2	Likely Pathogenic
19	ARID1A	1	26779533	c.5635C>T	p.Arg1879Trp	Missense	2%	Very Rare Global (0,008%)	VUS	NA	VUS
	MSH6	2	47803500	c.3261dupC	p.Phe1088LeufsTer5	Frameshift Indels	4%	-	Pathogenic	Pathogenic - 1	Pathogenic
	MSH3	5	80675096	c.1148delA	p.Lys383ArgfsTer32	Frameshift Indels	25%	-	Likely Pathogenic	Likely Pathogenic - 2	Likely Pathogenic
20	TP53	17	7675228	c.380_384delCCCCT	p.Ser127CysfsTer20	Frameshift Indels	2%	-	Likely Pathogenic	NA	VUS
21	TP53	17	7674250	c.713G>T	p.Cys238Phe	Missense	2%	-	VUS	Likely Pathogenic - 2	Likely Pathogenic
22	FAT4	4	125450920	c.9904C>A	p.Leu3302Ile	Missense	2%	-	VUS	NA	VUS
	TP53	17	7674208	c.743_755delGGAGGCCCATCCT	p.Arg248ProfsTer93	Frameshift Indels	2%	-	Likely Pathogenic	NA	VUS

Table 16 (cont.). *Somatic Variants identified in FIGC families by multiplex custom-panel based sequencing.*

Family	Gene	Chr	Position GRCh38	cDNA	Protein	Consequence	Freq cohort	Frequency	Illumina's Classification	Annotator's Classification	Integrated Classification
23	ARID1A	1	26766552	c.2974G>T	p.Glu992Ter	Stop gained	2%	-	Likely Pathogenic	NA	VUS
	ARID1A	1	26772565	c.3472G>T	p.Gly1158Ter	Stop gained	2%	-	Likely Pathogenic	NA	VUS
	MSH2	2	47480873	c.2634+4delT	-	Splice donor	2%	-	Likely Pathogenic	Likely Pathogenic - 1	Pathogenic
	TGFBR2	3	30688506	c.1594C>A	p.His532Asn	Missense	2%	-	VUS	NA	VUS
	MLH1	3	37040259	c.1632A>T	p.Gln544His	Missense	2%	-	VUS	NA	VUS
	FAT4	4	125451730	c.10714A>G	p.Ile3572Val	Missense	2%	-	VUS	NA	VUS
	MSH3	5	80675096	c.1148delA	p.Lys383ArgfsTer32	Frameshift Indels	25%	-	Likely Pathogenic	Likely Pathogenic - 2	Likely Pathogenic
	APC	5	112837681	c.2087A>G	p.Glu696Gly	Missense	2%	-	VUS	NA	VUS
	CTNNA1	5	138783232	c.161G>A	p.Arg54His	Missense	2%	Very Rare Global (0,005%)	VUS	NA	VUS
	CTNNA1	5	138824597	c.656C>T	p.Pro219Leu	Missense	2%	Very Rare Global (0,003%)	VUS	NA	VUS
	AKAP12	6	151350826	c.2435C>A	p.Pro812His	Missense	2%	-	VUS	NA	VUS
	ATM	11	108259061	c.2452A>G	p.Ile818Val	Missense	2%	-	VUS	NA	VUS
	ATM	11	108271249	c.2922-2A>C	-	Splice acceptor	2%	-	Likely Pathogenic	NA	Likely Pathogenic
	ATM	11	108304802	c.5624G>A	p.Arg1875Gln	Missense	2%	Very Rare Global (0,002%)	VUS	Uncertain Significance – 1	VUS
	BRCA2	13	32398361	c.9848T>A	p.Val3283Asp	Missense	2%	-	VUS	NA	VUS
	PALB2	16	23629845	c.2309C>A	p.Ala770Asp	Missense	2%	-	VUS	NA	VUS
24	Somatic variants were not found										
25	APC	5	112838478	c.2884G>C	p.Asp962His	Missense	2%	-	VUS	NA	VUS
	TP53	17	7675139	c.473G>A	p.Arg158His	Missense	2%	Very Rare Global (0,002%)	Pathogenic	Likely Pathogenic - 2	Likely Pathogenic
26	TP53	17	7675235	c.377A>G	p.Tyr126Cys	Missense, Splice region	2%	-	VUS	Likely Pathogenic - 1	Likely Pathogenic
27	MSH6	2	47803501	c.3261delC	p.Phe1088SerfsTer2	Frameshift Indels	6%	-	Pathogenic	Pathogenic - 1	Pathogenic
	MSH3	5	80675096	c.1148delA	p.Lys383ArgfsTer32	Frameshift Indels	25%	-	Likely Pathogenic	Likely Pathogenic - 2	Likely Pathogenic
	CTNNA1	5	138781963	c.39G>A	p.Trp13Ter	Stop gained	2%	-	Likely Pathogenic	NA	VUS
	CTNNA1	5	138930610	c.2150dupA	p.Gln718AlafsTer16	Frameshift Indels	2%	-	Likely Pathogenic	NA	VUS
28	TP53	17	7673802	c.818G>A	p.Arg273His	Missense	2%	Very Rare Eur (0,2%), Very Rare Tusc (0%)	Pathogenic	Likely Pathogenic - 3	Likely Pathogenic
29	TP53	17	7674240	c.723delC	p.Cys242AlafsTer5	Frameshift Indels	2%	-	Likely Pathogenic	NA	VUS
30	TP53	17	7674246	c.716dupA	p.Asn239LysfsTer25	Frameshift Indels	2%	-	Likely Pathogenic	Likely Pathogenic - 1	Likely Pathogenic
31	ARID1A	1	26762296	c.2402dupG	p.Gln802SerfsTer15	Frameshift Indels	2%	-	Likely Pathogenic	Likely Pathogenic - 1	Likely Pathogenic
	ARID1A	1	26774648	c.4424delA	p.Asn1475ThrfsTer6	Frameshift Indels	2%	-	Likely Pathogenic	Likely Pathogenic - 1	Likely Pathogenic
	MLH1	3	36993588	c.41C>T	p.Thr14Ile	Missense	2%	Very Rare Global (0,002%)	VUS	Uncertain significance - 1	VUS
	FAT4	4	125316934	c.523C>T	p.Arg175Cys	Missense	2%	Very rare Global (0,002%)	VUS	NA	VUS
	APC	5	112842672	c.7078G>T	p.Gly2360Cys	Missense	2%	-	VUS	NA	VUS
	HSPA5	9	125240909	c.123-2A>G	-	Splice acceptor	2%	-	Likely Pathogenic	NA	VUS
32	TP53	17	7674250	c.713G>C	p.Cys238Ser	Missense	2%	-	VUS	Likely Pathogenic – 2	Likely Pathogenic
33	MSH6	2	47795898	c.465delA	p.Lys155AsnfsTer19	Frameshift Indels	2%	-	Likely Pathogenic	NA	VUS
	MSH3	5	80675096	c.1148delA	p.Lys383ArgfsTer32	Frameshift Indels	25%	-	Likely Pathogenic	Likely Pathogenic - 2	Likely Pathogenic
	APC	5	112838594	c.3000C>A	p.Tyr1000Ter	Stop gained	2%	-	Likely Pathogenic	NA	Likely Pathogenic
	APC	5	112839979	c.4393_4394delAG	p.Ser1465TrpfsTer3	Frameshift Indels	2%	-	Pathogenic	Pathogenic - 1; Likely Pathogenic - 1	Likely Pathogenic
	BRCA2	13	32339700	c.5351delA	p.Asn1784ThrfsTer7	Frameshift Indels	2%	-	Pathogenic	Pathogenic - 1	Pathogenic
34	TP53	17	7675161	c.451C>T	p.Pro151Ser	Missense	2%	-	Pathogenic	Likely Pathogenic - 3	Likely Pathogenic

Table 16 (cont.). *Somatic Variants identified in FIGC families by multiplex custom-panel based sequencing.*

Family	Gene	Chr	Position GRCh38	cDNA	Protein	Consequence	Freq cohort	Frequency	Illumina's Classification	Annotator's Classification	Integrated Classification
35	ARID1A	1	26771197	c.3281delA	p.Lys1094SerfsTer67	Frameshift Indels	2%	-	Likely Pathogenic	NA	VUS
	MSH2	2	47412449	c.687delA	p.Ala230LeufsTer16	Frameshift Indels	2%	-	Pathogenic	Likely Pathogenic - 1	Likely Pathogenic
	MSH3	5	80675096	c.1148delA	p.Lys383ArgfsTer32	Frameshift Indels	25%	-	Likely Pathogenic	Likely Pathogenic - 2	Likely Pathogenic
	APC	5	112839978	c.4393_4394dupAG	p.Ser1465ArgfsTer9	Frameshift Indels	2%	-	Likely Pathogenic	Likely Pathogenic - 1	Likely Pathogenic
	APC	5	112840367	c.4778delA	p.Lys1593SerfsTer57	Frameshift Indels	2%	-	Likely Pathogenic	NA	VUS
	CTNNA1	5	138924599	c.1636C>T	p.Arg546Ter	Stop gained	2%	-	Likely Pathogenic	NA	VUS
	CTHRC1	8	103378048	c.394C>T	p.Arg132Cys	Missense	2%	Very Rare Global (0,08%)	VUS	NA	VUS
	ATM	11	108227807	c.104G>A	p.Arg35Gln	Missense	2%	-	VUS	Uncertain Significance - 1	VUS
	CDH1	16	68801883	c.377C>T	p.Pro126Leu	Missense	2%	Very Rare Global (0,004%)	VUS	Uncertain Significance - 1	VUS
	STK11	19	1220399	c.491T>C	p.Leu164Pro	Missense	2%	-	VUS	NA	VUS
36	MSR1	8	16155129	c.833C>T	p.Pro278Leu	Missense	2%	-	VUS	NA	VUS
	TP53	17	7674221	c.742C>T	p.Arg248Trp	Missense	4%	Very Rare Global (0,002%)	Pathogenic	Likely Pathogenic - 3	Likely Pathogenic
37	MSH3	5	80675096	c.1148delA	p.Lys383ArgfsTer32	Frameshift Indels	25%	-	Likely Pathogenic	Likely Pathogenic - 2	Likely Pathogenic
38	MSH3	5	80675096	c.1148delA	p.Lys383ArgfsTer32	Frameshift Indels	25%	-	Likely Pathogenic	Likely Pathogenic - 2	Likely Pathogenic
	CTNNA1	5	138824800	c.858+1G>A	-	Splice donor	2%	-	Likely Pathogenic	NA	VUS
	TP53	17	7674230	c.733G>A	p.Gly245Ser	Missense	2%	Very Rare Global (0,002%)	Pathogenic	Likely Pathogenic - 3	Likely Pathogenic
39	APC	5	112839702	c.4108A>T	p.Lys1370Ter	Stop gained	2%	-	Likely Pathogenic	NA	Likely Pathogenic
40	MSH6	2	47800177	c.2194C>T	p.Arg732Ter	Stop gained	2%	Very Rare Global (0,002%)	Pathogenic	Pathogenic - 1	Pathogenic
	MSH6	2	47803500	c.3261dupC	p.Phe1088LeufsTer5	Frameshift Indels	4%	-	Pathogenic	Likely Pathogenic - 1	Pathogenic
	FAT4	4	125321047	c.4636G>A	p.Val1546Ile	Missense	2%	Very Rare Global (0,002%)	VUS	NA	VUS
	MSH3	5	80675096	c.1148delA	p.Lys383ArgfsTer32	Frameshift Indels	25%	-	Likely Pathogenic	Likely Pathogenic - 2	Likely Pathogenic
	ATM	11	108229320	c.329_330delGA	p.Arg110LysfsTer4	Frameshift Indels, Splice region	2%	-	Likely Pathogenic	NA	VUS
	BRCA2	13	32394815	c.9383G>A	p.Arg3128Gln	Missense	2%	Very Rare Global (0,002%)	VUS	Uncertain Significance - 1	Likely Pathogenic
	TP53	17	7675157	c.455delC	p.Pro152ArgfsTer18	Frameshift Indels	2%	-	Likely Pathogenic	NA	VUS
41	TP53	17	7673787	c.833C>T	p.Pro278Leu	Missense	2%	-	VUS	Likely Pathogenic - 2	Likely Pathogenic
42	MCCC1	3	183037239	c.1573A>T	p.Thr525Ser	Missense	2%	-	VUS	NA	VUS
	FAT4	4	125448714	c.7698A>C	p.Lys2566Asn	Missense	2%	-	VUS	NA	VUS
	MSH3	5	80675096	c.1148delA	p.Lys383ArgfsTer32	Frameshift Indels	25%	-	Likely Pathogenic	Likely Pathogenic - 2	Likely Pathogenic
	ITIH2	10	7705109	c.86T>G	p.Phe29Cys	Missense, Splice region	2%	-	VUS	NA	VUS
	ATM	11	108247071	c.1009C>T	p.Arg337Cys	Missense	2%	Very Rare Global (0,002%)	VUS	Likely Pathogenic - 1; Uncertain Significance - 1	Conflicting
	BRCA2	13	32337305	c.2957dupA	p.Asn986LysfsTer2	Frameshift Indels	2%	-	Likely Pathogenic	Pathogenic - 1	Pathogenic
	BRCA2	13	32356499	c.7507G>A	p.Val2503Ile	Missense	2%	Very Rare Global (0,002%)	VUS	Uncertain Significance - 1	VUS
43	MSH6	2	47803501	c.3261delC	p.Phe1088SerfsTer2	Frameshift Indels	6%	-	Pathogenic	Likely Pathogenic - 1	Pathogenic
	MSH3	5	80675096	c.1148delA	p.Lys383ArgfsTer32	Frameshift Indels	25%	-	Likely Pathogenic	Likely Pathogenic - 2	Likely Pathogenic
44	MSH6	2	47803501	c.3261delC	p.Phe1088SerfsTer2	Frameshift Indels	6%	-	Pathogenic	Likely Pathogenic - 1	Pathogenic
	MSH3	5	80675096	c.1148delA	p.Lys383ArgfsTer32	Frameshift Indels	25%	-	Likely Pathogenic	Likely Pathogenic - 2	Likely Pathogenic
45	TP53	17	7674220	c.743G>A	p.Arg248Gln	Missense	2%	Very Rare Global (0,01%)	Pathogenic	Likely Pathogenic – 3	Likely Pathogenic
46	PTEN	10	87952122	c.499dupA	p.Thr167AsnfsTer13	Frameshift Indels	2%	-	Likely Pathogenic	NA	VUS
47-52	Somatic variants were not found										

Description of somatic variants, including both frequency in the cohort (freq cohort: % of mutated families), ExAC and 1000genomes and Clinical Significance classification from Illumina's Variant Interpreter, Annotator (in 3 databases) and integrated classification. Eur: Europe; Tusc: Tuscany; VUS: Variant of Unknown Significance; NA: Not available; Very Rare: <1%; Rare: >1, <5; Not Rare: >5.

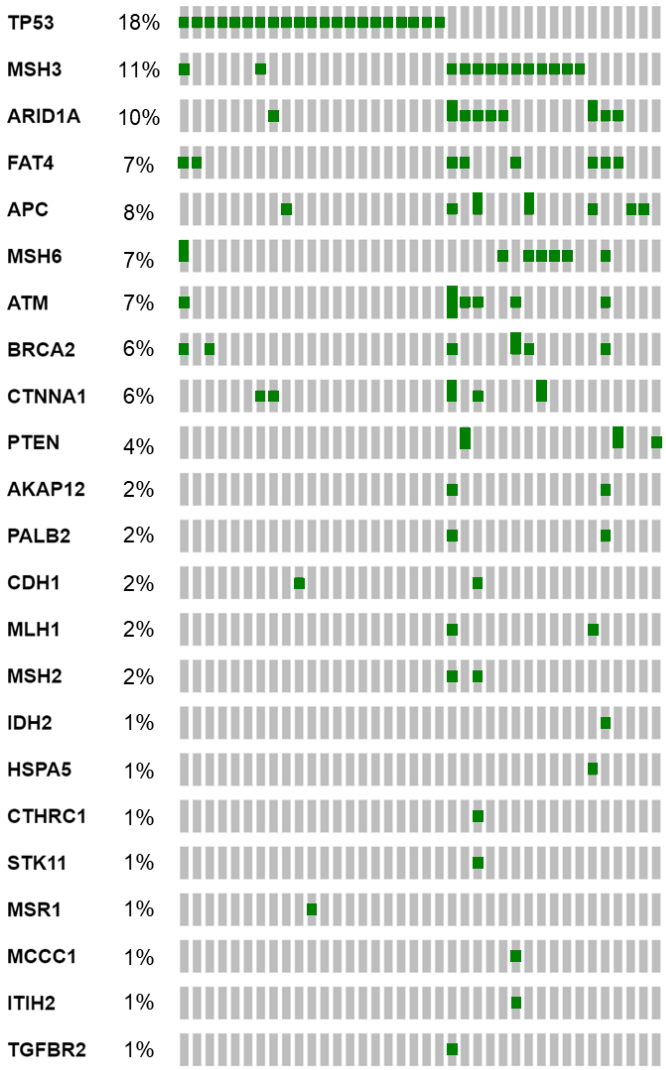


Figure 41. *Oncoprint of somatic variants. Columns represent families that harbored somatic variants. Grey: somatic variant not found; 1/3 of the rectangle green: 1 somatic variant; 2/3 of the rectangle green: 2 somatic variants; complete rectangle green: 3 somatic variants. % of variants in each gene/115.*

Table 17*. Differences of somatic landscape between families with and without germline variants.*

With germline variants	Without germline variants
24 families	28 families
16 (67%) families harbored somatic variants	20 (71%) families harbored somatic variants
TP53 (18%) and ARID1A (15%) were the most frequently altered genes	TP53 (20%) and APC (12%) were the most frequently altered genes
10 (63%) families harbored more than one somatic variant	7 (35%) families harbored more than one somatic variant
12 (50%) families displayed MSI phenotype	9 (32%) families displayed MSI phenotype

V. Conclusions

During the last decade, next generation sequencing technologies have revealed the genetic landscape of classic predisposition syndromes and contributed to the definition of cancer risks [93].

Whilst germline defects were found for HDGC and GAPPS, FIGC remains without a genetic risk factor identified.

Working under the hypothesis that the rare FIGC syndrome is caused by germline co-occurrence of moderate-risk alleles and represent a polygenic, rather than a classical monogenic disease, the work described in this master thesis was an attempt to dissect the germline and somatic landscapes of the largest FIGC cohort ever studied.

A panel of genes was used to sequence both the normal (germline) and tumor (somatic) DNA from the probands of 52 FIGC families. The choice of using a panel of genes was made based on a clinical perspective by directing to genes associated with gastrointestinal tract cancers or syndromes in which GC is predominant. Using a custom panel-based sequencing approach instead of whole genome or whole exome sequencing had also several advantages, such as: less false discovery rate, i.e, fewer variants discovered with no implication in the clinic; less bioinformatics analysis complexity and less data volume and storage.

One of the main findings of this thesis was the co-occurrence of germline variants in 10 out of 24 (42%) FIGC families that carried germline variants. However, it should not be discard that other important variants located in other genes (absent in the panel), could also occur in concomitance with the variants found, thus potentially increasing the number of FIGC families with co-occurrence of variants.

Main Findings

- Thirty-six **germline variants,** affecting 17 genes, were found in 24 families;
- The 36 germline variants were classified as: Likely Pathogenic (1), Conflicting (11), Likely Benign (3), and Variants of Unknown Significance (21);
- The most frequently mutated genes were *MSH6* (17%), *MAP3K6* (11%), *SDHD* (11%), *ATM* (8%) and *MTUS1* (8%);
- 50% (12/24) of the families had germline variants in DNA repair genes;
- 25% (6/24) of the families had germline variants in genes associated with metabolism;

Main Findings (cont.)

- 10/ 24 FIGC families (42%) carried **co-occurrence of germline variants**;
- DNA repair genes were frequently altered in co-occurrence with metabolism associated genes (F1, F4 and F11);
- Three families harbored germline variants in chromosomes 2 and 11 (F1, F4 and F11);
- Two families displayed, potentially, a second inactivation hit in the wild-type allele: second somatic mutation (in F1 – *MSH6* and F16 – *FAT4*);
- One family displayed increased *CTHRC1* mRNA expression in the tumor, in comparison with normal stomach and a decreased *APC* mRNA expression in the tumor associated with *APC* somatic variant;
- Impairment of the same pathways recurrently appeared in families with co-occurrence of germline variants: MMR (*MSH6* and *MSH2*), HR (*ATM*, *BRCA1* and *BRCA2*), MAPK (*MAP3K6)* and PCP (*CTHRC1* and *FAT4*), as well as, same genes, *MTUS1 SDHB* and *SDHD*;

- One hundred and fifteen **somatic variants,** affecting 23 genes, were found in 36 families;
- The 115 somatic variants were classified as: Pathogenic (12), Likely Pathogenic (54), and Variants of Unknown Significance (49);
- The somatic landscape of families with and without germline variants was similar;
- The most frequently mutated genes were *TP53* (18%), *MSH3* (11%), *ARID1A* (10%), *FAT4* (7%), *APC* (8%), *MSH6* (7%), *ATM* (7%), *BRCA2* (6%) and *CTNNA1* (6%);
- The MSI phenotype was more frequent in families with germline variants;
- The majority of families with germline variants (63%) harbored somatic variants in more than one gene, comparing with families without germline variants (35%).

In conclusion, the work described in this thesis pinpointed FIGC as a polygenic rather than a monogenic disease, where co-occurrence of low or moderate risk alleles that interact with family history and other non-genetic factors (environmental) can affect the risk of cancer of each individual.

VI. Future Directions

The work developed during this master thesis was the first attempt to dissect the genetic cause underlying FIGC using a relatively large and homogeneous cohort. Nevertheless, several issues and unanswered questions still need to be clarified and future experiments should be done for a better understanding of the genetic risk factors underlying FIGC and to prove its potential polygenic nature.

Further experiments may include:

1) Loss of heterozygosity analysis as a potential somatic second inactivation hit, in families that harbored germline variants, but neither second mutation nor promoter methylation revealed to be the inactivation mechanisms;

2) Analysis of the impact of somatic second hits by quantifying the mRNA and protein levels of potentially inactivated genes, using real time PCR and immunohistochemistry, respectively;

3) Immunohistochemistry analysis of protein partners of genes carrying germline variants;

4) Dissecting the pathogenicity of the germline variants found to be in co-occurrence in each family. For example, intestinal GC cell lines could be transfected with expression vectors carrying those variants and further analyzed for proliferation, invasion, apoptosis and other cancer-related properties;

5) In parallel, segregation analysis in all families, to better understand if the germline variants found in the proband are reflected in the affected family members.

These approaches may shine a light on the underlying genetic causes of FIGC and envision the establishment of a genetic screening protocol for a better management of FIGC families.

VII. References

1. McIntosh, J. R. (2016) Mitosis. *Cold Spring Harbor perspectives in biology* **8**
2. Vinay Kumar, A. K. A., Jon C. Aster. (2014) *Robbins & Cotran Pathologic Basis of Disease*, Elsevier Health
3. David L. Nelson, M. M. C. (2008) *Lehninger, Principles of Biochemistry*, W. H. Freeman and Company, New York
4. Romero-Laorden, N., and Castro, E. (2017) Inherited mutations in DNA repair genes and cancer risk. *Current problems in cancer* **41**, 251-264
5. Jang, S. J., Gardner, J. M., and Ro, J. Y. (2011) Diagnostic approach and prognostic factors of cancers. *Advances in anatomic pathology* **18**, 165-172
6. Ferlay, J., Soerjomataram, I., Dikshit, R., Eser, S., Mathers, C., Rebelo, M., Parkin, D. M., Forman, D., and Bray, F. (2015) Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *International journal of cancer* **136**, E359-386
7. Hanahan, D., and Weinberg, R. A. (2000) The hallmarks of cancer. *Cell* **100**, 57-70
8. Provencher, P. A., and Love, J. A. (2015) Synthesis and Performance of a Biomimetic Indicator for Alkylating Agents. *The Journal of organic chemistry* **80**, 9603-9609
9. D'Orazio, J., Jarrett, S., Amaro-Ortiz, A., and Scott, T. (2013) UV radiation and the skin. *International journal of molecular sciences* **14**, 12222-12248
10. Figueiredo, C., Costa, S., Karameris, A., and Machado, J. C. (2015) Pathogenesis of Gastric Cancer. *Helicobacter* **20 Suppl 1**, 30-35
11. Macleod, K. (2000) Tumor suppressor genes. *Current opinion in genetics & development* **10**, 81-93
12. Liu, Z., Liu, L., Li, M., Wang, Z., Feng, L., Zhang, Q., Cheng, S., and Lu, S. (2011) Epidermal growth factor receptor mutation in gastric cancer. *Pathology* **43**, 234-238
13. Tan, C., and Du, X. (2012) KRAS mutation testing in metastatic colorectal cancer. *World journal of gastroenterology* **18**, 5171-5180
14. Boxer, L. M., and Dang, C. V. (2001) Translocations involving c-myc and c-myc function. *Oncogene* **20**, 5595-5610
15. Malumbres, M., and Barbacid, M. (2009) Cell cycle, CDKs and cancer: a changing paradigm. *Nature reviews. Cancer* **9**, 153-166
16. Guo, X. E., Ngo, B., Modrek, A. S., and Lee, W. H. (2014) Targeting tumor suppressor networks for cancer therapeutics. *Current drug targets* **15**, 2-16
17. Westcott, P. M., and To, M. D. (2013) The genetics and biology of KRAS in lung cancer. *Chinese journal of cancer* **32**, 63-70

18. Gajria, D., and Chandarlapaty, S. (2011) HER2-amplified breast cancer: mechanisms of trastuzumab resistance and novel targeted therapies. *Expert review of anticancer therapy* **11**, 263-275
19. Dasgupta, S., Ray, U. K., Mitra, A. G., Bhattacharyya, D. M., Mukhopadhyay, A., Das, P., Gangopadhyay, S., Roy, S., and Mukhopadhyay, S. (2017) Evaluation of a new flow cytometry based method for detection of BCR-ABL1 fusion protein in chronic myeloid leukemia. *Blood research* **52**, 112-118
20. Knudson, A. G. (2001) Two genetic hits (more or less) to cancer. *Nature reviews. Cancer* **1**, 157-162
21. Zilfou, J. T., and Lowe, S. W. (2009) Tumor suppressive functions of p53. *Cold Spring Harbor perspectives in biology* **1**, a001883
22. Lane, D., and Levine, A. (2010) p53 Research: the past thirty years and the next thirty years. *Cold Spring Harbor perspectives in biology* **2**, a000893
23. Caponigro, G., and Sellers, W. R. (2011) Advances in the preclinical testing of cancer therapeutic hypotheses. *Nature reviews. Drug discovery* **10**, 179-187
24. (2012) Comprehensive molecular characterization of human colon and rectal cancer. *Nature* **487**, 330-337
25. Munoz, N. M., Upton, M., Rojas, A., Washington, M. K., Lin, L., Chytil, A., Sozmen, E. G., Madison, B. B., Pozzi, A., Moon, R. T., Moses, H. L., and Grady, W. M. (2006) Transforming growth factor beta receptor type II inactivation induces the malignant transformation of intestinal neoplasms initiated by Apc mutation. *Cancer research* **66**, 9837-9844
26. Carneiro, P., Fernandes, M. S., Figueiredo, J., Caldeira, J., Carvalho, J., Pinheiro, H., Leite, M., Melo, S., Oliveira, P., Simoes-Correia, J., Oliveira, M. J., Carneiro, F., Figueiredo, C., Paredes, J., Oliveira, C., and Seruca, R. (2012) E-cadherin dysfunction in gastric cancer--cellular consequences, clinical applications and open questions. *FEBS letters* **586**, 2981-2989
27. Carethers, J. M., and Jung, B. H. (2015) Genetics and Genetic Biomarkers in Sporadic Colorectal Cancer. *Gastroenterology* **149**, 1177-1190.e1173
28. Calon, A., Espinet, E., Palomo-Ponce, S., Tauriello, D. V., Iglesias, M., Cespedes, M. V., Sevillano, M., Nadal, C., Jung, P., Zhang, X. H., Byrom, D., Riera, A., Rossell, D., Mangues, R., Massague, J., Sancho, E., and Batlle, E. (2012) Dependency of colorectal cancer on a TGF-beta-driven program in stromal cells for metastasis initiation. *Cancer cell* **22**, 571-584
29. McBride, K. A., Ballinger, M. L., Killick, E., Kirk, J., Tattersall, M. H., Eeles, R. A., Thomas, D. M., and Mitchell, G. (2014) Li-Fraumeni syndrome: cancer risk

assessment and clinical management. *Nature reviews. Clinical oncology* **11**, 260-271

30. Boland, C. R., and Goel, A. (2010) Microsatellite instability in colorectal cancer. *Gastroenterology* **138**, 2073-2087.e2073
31. Gaston, D., Hansford, S., Oliveira, C., Nightingale, M., Pinheiro, H., Macgillivray, C., Kaurah, P., Rideout, A. L., Steele, P., Soares, G., Huang, W. Y., Whitehouse, S., Blowers, S., LeBlanc, M. A., Jiang, H., Greer, W., Samuels, M. E., Orr, A., Fernandez, C. V., Majewski, J., Ludman, M., Dyack, S., Penney, L. S., McMaster, C. R., Huntsman, D., and Bedard, K. (2014) Germline mutations in MAP3K6 are associated with familial gastric cancer. *PLoS genetics* **10**, e1004669
32. Chial, H. (2008) Tumor Suppressor (TS) Genes and the Two-Hit Hypothesis. *Nature Education* **1**, 177
33. Worm, J., and Guldberg, P. (2002) DNA methylation: an epigenetic pathway to cancer and a promising target for anticancer therapy. *Journal of oral pathology & medicine : official publication of the International Association of Oral Pathologists and the American Academy of Oral Pathology* **31**, 443-449
34. Clouaire, T., and Stancheva, I. (2008) Methyl-CpG binding proteins: specialized transcriptional repressors or structural components of chromatin? *Cellular and molecular life sciences : CMLS* **65**, 1509-1522
35. Oliveira, C., Sousa, S., Pinheiro, H., Karam, R., Bordeira-Carrico, R., Senz, J., Kaurah, P., Carvalho, J., Pereira, R., Gusmao, L., Wen, X., Cipriano, M. A., Yokota, J., Carneiro, F., Huntsman, D., and Seruca, R. (2009) Quantification of epigenetic and genetic 2nd hits in CDH1 during hereditary diffuse gastric cancer syndrome progression. *Gastroenterology* **136**, 2137-2148
36. Shi, J., Qu, Y. P., and Hou, P. (2014) Pathogenetic mechanisms in gastric cancer. *World journal of gastroenterology* **20**, 13804-13819
37. Maconi, G., Manes, G., and Porro, G. B. (2008) Role of symptoms in diagnosis and outcome of gastric cancer. *World journal of gastroenterology* **14**, 1149-1155
38. Ang, T. L., and Fock, K. M. (2014) Clinical epidemiology of gastric cancer. *Singapore medical journal* **55**, 621-628
39. Jang, B. G., and Kim, W. H. (2011) Molecular pathology of gastric carcinoma. *Pathobiology : journal of immunopathology, molecular and cellular biology* **78**, 302-310
40. Pilehchian Langroudi, M., Nikbakhsh, N., Samadani, A. A., Fattahi, S., Taheri, H., Shafaei, S., Amirbozorgi, G., Pilehchian Langroudi, R., and Akhavan-Niaki, H. (2017) FAT4 hypermethylation and grade dependent downregulation in gastric adenocarcinoma. *Journal of cell communication and signaling* **11**, 69-75

41. Corso, G., Carvalho, J., Marrelli, D., Vindigni, C., Carvalho, B., Seruca, R., Roviello, F., and Oliveira, C. (2013) Somatic mutations and deletions of the E-cadherin gene predict poor survival of patients with gastric cancer. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* **31**, 868-875
42. Park, J. Y., von Karsa, L., and Herrero, R. (2014) Prevention strategies for gastric cancer: a global perspective. *Clinical endoscopy* **47**, 478-489
43. Corso, G., Seruca, R., and Roviello, F. (2012) Gastric cancer carcinogenesis and tumor progression. *Annali italiani di chirurgia* **83**, 172-176
44. Petrovchich, I., and Ford, J. M. (2016) Genetic predisposition to gastric cancer. *Seminars in oncology* **43**, 554-559
45. Lauren, P. (1991) Histogenesis of intestinal and diffuse types of gastric carcinoma. *Scandinavian journal of gastroenterology. Supplement* **180**, 160-164
46. Oliveira, C., Seruca, R., and Carneiro, F. (2009) Hereditary gastric cancer. *Best practice & research. Clinical gastroenterology* **23**, 147-157
47. F., C. (1997) Classification of gastric carcinomas. *Curr Diagn Pathol* **4**, 51-59
48. Crew, K. D., and Neugut, A. I. (2006) Epidemiology of gastric cancer. *World journal of gastroenterology* **12**, 354-362
49. Correa, P., Haenszel, W., Cuello, C., Tannenbaum, S., and Archer, M. (1975) A model for gastric cancer epidemiology. *Lancet (London, England)* **2**, 58-60
50. Tan, P., and Yeoh, K. G. (2015) Genetics and Molecular Pathogenesis of Gastric Adenocarcinoma. *Gastroenterology* **149**, 1153-1162.e1153
51. El-Omar, E. M., Carrington, M., Chow, W. H., McColl, K. E., Bream, J. H., Young, H. A., Herrera, J., Lissowska, J., Yuan, C. C., Rothman, N., Lanyon, G., Martin, M., Fraumeni, J. F., Jr., and Rabkin, C. S. (2000) Interleukin-1 polymorphisms associated with increased risk of gastric cancer. *Nature* **404**, 398-402
52. Figueiredo, C., Machado, J. C., Pharoah, P., Seruca, R., Sousa, S., Carvalho, R., Capelinha, A. F., Quint, W., Caldas, C., van Doorn, L. J., Carneiro, F., and Sobrinho-Simoes, M. (2002) Helicobacter pylori and interleukin 1 genotyping: an opportunity to identify high-risk individuals for gastric carcinoma. *Journal of the National Cancer Institute* **94**, 1680-1687
53. Liou, J. M., Lin, J. T., Wang, H. P., Huang, S. P., Lee, Y. C., Chiu, H. M., Shun, C. T., and Wu, M. S. (2007) IL-1B-511 C-->T polymorphism is associated with increased host susceptibility to Helicobacter pylori infection in Chinese. *Helicobacter* **12**, 142-149
54. (2014) Comprehensive molecular characterization of gastric adenocarcinoma. *Nature* **513**, 202-209

55. Zang, Z. J., Cutcutache, I., Poon, S. L., Zhang, S. L., McPherson, J. R., Tao, J., Rajasegaran, V., Heng, H. L., Deng, N., Gan, A., Lim, K. H., Ong, C. K., Huang, D., Chin, S. Y., Tan, I. B., Ng, C. C., Yu, W., Wu, Y., Lee, M., Wu, J., Poh, D., Wan, W. K., Rha, S. Y., So, J., Salto-Tellez, M., Yeoh, K. G., Wong, W. K., Zhu, Y. J., Futreal, P. A., Pang, B., Ruan, Y., Hillmer, A. M., Bertrand, D., Nagarajan, N., Rozen, S., Teh, B. T., and Tan, P. (2012) Exome sequencing of gastric adenocarcinoma identifies recurrent somatic mutations in cell adhesion and chromatin remodeling genes. *Nature genetics* **44**, 570-574

56. Wang, K., Kan, J., Yuen, S. T., Shi, S. T., Chu, K. M., Law, S., Chan, T. L., Kan, Z., Chan, A. S., Tsui, W. Y., Lee, S. P., Ho, S. L., Chan, A. K., Cheng, G. H., Roberts, P. C., Rejto, P. A., Gibson, N. W., Pocalyko, D. J., Mao, M., Xu, J., and Leung, S. Y. (2011) Exome sequencing identifies frequent mutation of ARID1A in molecular subtypes of gastric cancer. *Nature genetics* **43**, 1219-1223

57. Cerami, E., Gao, J., Dogrusoz, U., Gross, B. E., Sumer, S. O., Aksoy, B. A., Jacobsen, A., Byrne, C. J., Heuer, M. L., Larsson, E., Antipin, Y., Reva, B., Goldberg, A. P., Sander, C., and Schultz, N. (2012) The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer discovery* **2**, 401-404

58. Gao, J., Aksoy, B. A., Dogrusoz, U., Dresdner, G., Gross, B., Sumer, S. O., Sun, Y., Jacobsen, A., Sinha, R., Larsson, E., Cerami, E., Sander, C., and Schultz, N. (2013) Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Science signaling* **6**, pl1

59. Yamamoto, H., and Imai, K. (2015) Microsatellite instability: an update. *Archives of toxicology* **89**, 899-921

60. Goldberg, R. M. (2015) Genomic Profiling in Gastrointestinal Cancer: Are We Ready To Use These Data to Make Treatment Decisions? *The oncologist* **20**, 1448-1456

61. Duval, A., and Hamelin, R. (2002) Mutations at coding repeat sequences in mismatch repair-deficient human cancers: toward a new concept of target genes for instability. *Cancer research* **62**, 2447-2454

62. Carvalho, B., Buffart, T. E., Reis, R. M., Mons, T., Moutinho, C., Silva, P., van Grieken, N. C., Grabsch, H., van de Velde, C. J., Ylstra, B., Meijer, G. A., and Carneiro, F. (2006) Mixed gastric carcinomas show similar chromosomal aberrations in both their diffuse and glandular components. *Cellular oncology : the official journal of the International Society for Cellular Oncology* **28**, 283-294

63. Garattini, S. K., Basile, D., Cattaneo, M., Fanotto, V., Ongaro, E., Bonotto, M., Negri, F. V., Berenato, R., Ermacora, P., Cardellino, G. G., Giovannoni, M., Pella, N., Scartozzi, M., Antonuzzo, L., Silvestris, N., Fasola, G., and Aprile, G. (2017)

Molecular classifications of gastric cancers: Novel insights and possible future applications. *World journal of gastrointestinal oncology* **9**, 194-208

64. Ottini, L., Falchetti, M., Lupi, R., Rizzolo, P., Agnese, V., Colucci, G., Bazan, V., and Russo, A. (2006) Patterns of genomic instability in gastric cancer: clinical implications and perspectives. *Annals of oncology : official journal of the European Society for Medical Oncology* **17 Suppl 7**, vii97-102
65. Albarello, L., Pecciarini, L., and Doglioni, C. (2011) HER2 testing in gastric cancer. *Advances in anatomic pathology* **18**, 53-59
66. Maleki, S. S., and Rocken, C. (2017) Chromosomal Instability in Gastric Cancer Biology. *Neoplasia (New York, N.Y.)* **19**, 412-420
67. Niwa, T., Tsukamoto, T., Toyoda, T., Mori, A., Tanaka, H., Maekita, T., Ichinose, M., Tatematsu, M., and Ushijima, T. (2010) Inflammatory processes triggered by Helicobacter pylori infection cause aberrant DNA methylation in gastric epithelial cells. *Cancer research* **70**, 1430-1440
68. Oliveira, C., Pinheiro, H., Figueiredo, J., Seruca, R., and Carneiro, F. (2015) Familial gastric cancer: genetic susceptibility, pathology, and implications for management. *The Lancet. Oncology* **16**, e60-70
69. Hansford, S., Kaurah, P., Li-Chang, H., Woo, M., Senz, J., Pinheiro, H., Schrader, K. A., Schaeffer, D. F., Shumansky, K., Zogopoulos, G., Santos, T. A., Claro, I., Carvalho, J., Nielsen, C., Padilla, S., Lum, A., Talhouk, A., Baker-Lange, K., Richardson, S., Lewis, I., Lindor, N. M., Pennell, E., MacMillan, A., Fernandez, B., Keller, G., Lynch, H., Shah, S. P., Guilford, P., Gallinger, S., Corso, G., Roviello, F., Caldas, C., Oliveira, C., Pharoah, P. D., and Huntsman, D. G. (2015) Hereditary Diffuse Gastric Cancer Syndrome: CDH1 Mutations and Beyond. *JAMA oncology* **1**, 23-32
70. Pinheiro, H., Bordeira-Carrico, R., Seixas, S., Carvalho, J., Senz, J., Oliveira, P., Inacio, P., Gusmao, L., Rocha, J., Huntsman, D., Seruca, R., and Oliveira, C. (2010) Allele-specific CDH1 downregulation and hereditary diffuse gastric cancer. *Human molecular genetics* **19**, 943-952
71. van Roy, F. (2014) Beyond E-cadherin: roles of other cadherin superfamily members in cancer. *Nature reviews. Cancer* **14**, 121-134
72. Colvin, H., Yamamoto, K., Wada, N., and Mori, M. (2015) Hereditary Gastric Cancer Syndromes. *Surgical oncology clinics of North America* **24**, 765-777
73. Schrader, K. A., Masciari, S., Boyd, N., Wiyrick, S., Kaurah, P., Senz, J., Burke, W., Lynch, H. T., Garber, J. E., and Huntsman, D. G. (2008) Hereditary diffuse gastric cancer: association with lobular breast cancer. *Familial cancer* **7**, 73-82

74. Fitzgerald, R. C., Hardwick, R., Huntsman, D., Carneiro, F., Guilford, P., Blair, V., Chung, D. C., Norton, J., Ragunath, K., Van Krieken, J. H., Dwerryhouse, S., and Caldas, C. (2010) Hereditary diffuse gastric cancer: updated consensus guidelines for clinical management and directions for future research. *Journal of medical genetics* **47**, 436-444

75. Brooks-Wilson, A. R., Kaurah, P., Suriano, G., Leach, S., Senz, J., Grehan, N., Butterfield, Y. S., Jeyes, J., Schinas, J., Bacani, J., Kelsey, M., Ferreira, P., MacGillivray, B., MacLeod, P., Micek, M., Ford, J., Foulkes, W., Australie, K., Greenberg, C., LaPointe, M., Gilpin, C., Nikkel, S., Gilchrist, D., Hughes, R., Jackson, C. E., Monaghan, K. G., Oliveira, M. J., Seruca, R., Gallinger, S., Caldas, C., and Huntsman, D. (2004) Germline E-cadherin mutations in hereditary diffuse gastric cancer: assessment of 42 new families and review of genetic screening criteria. *Journal of medical genetics* **41**, 508-517

76. Worthley, D. L., Phillips, K. D., Wayte, N., Schrader, K. A., Healey, S., Kaurah, P., Shulkes, A., Grimpen, F., Clouston, A., Moore, D., Cullen, D., Ormonde, D., Mounkley, D., Wen, X., Lindor, N., Carneiro, F., Huntsman, D. G., Chenevix-Trench, G., and Suthers, G. K. (2012) Gastric adenocarcinoma and proximal polyposis of the stomach (GAPPS): a new autosomal dominant syndrome. *Gut* **61**, 774-779

77. Li, J., Woods, S. L., Healey, S., Beesley, J., Chen, X., Lee, J. S., Sivakumaran, H., Wayte, N., Nones, K., Waterfall, J. J., Pearson, J., Patch, A. M., Senz, J., Ferreira, M. A., Kaurah, P., Mackenzie, R., Heravi-Moussavi, A., Hansford, S., Lannagan, T. R., Spurdle, A. B., Simpson, P. T., da Silva, L., Lakhani, S. R., Clouston, A. D., Bettington, M., Grimpen, F., Busuttil, R. A., Di Costanzo, N., Boussioutas, A., Jeanjean, M., Chong, G., Fabre, A., Olschwang, S., Faulkner, G. J., Bellos, E., Coin, L., Rioux, K., Bathe, O. F., Wen, X., Martin, H. C., Neklason, D. W., Davis, S. R., Walker, R. L., Calzone, K. A., Avital, I., Heller, T., Koh, C., Pineda, M., Rudloff, U., Quezado, M., Pichurin, P. N., Hulick, P. J., Weissman, S. M., Newlin, A., Rubinstein, W. S., Sampson, J. E., Hamman, K., Goldgar, D., Poplawski, N., Phillips, K., Schofield, L., Armstrong, J., Kiraly-Borri, C., Suthers, G. K., Huntsman, D. G., Foulkes, W. D., Carneiro, F., Lindor, N. M., Edwards, S. L., French, J. D., Waddell, N., Meltzer, P. S., Worthley, D. L., Schrader, K. A., and Chenevix-Trench, G. (2016) Point Mutations in Exon 1B of APC Reveal Gastric Adenocarcinoma and Proximal Polyposis of the Stomach as a Familial Adenomatous Polyposis Variant. *American journal of human genetics* **98**, 830-842

78. Giovanni Corso, F. R. (2013) *Spotlight on Familial and Hereditary Gastric Cancer* Springer Netherlands

79. Pedrazzani, C., Corso, G., Velho, S., Leite, M., Pascale, V., Bettarini, F., Marrelli, D., Seruca, R., and Roviello, F. (2009) Evidence of tumor microsatellite instability in gastric cancer with familial aggregation. *Familial cancer* **8**, 215-220

80. Leite, M., Corso, G., Sousa, S., Milanezi, F., Afonso, L. P., Henrique, R., Soares, J. M., Castedo, S., Carneiro, F., Roviello, F., Oliveira, C., and Seruca, R. (2011) MSI phenotype and MMR alterations in familial and sporadic gastric cancer. *International journal of cancer* **128**, 1606-1613

81. Setia, N., Clark, J. W., Duda, D. G., Hong, T. S., Kwak, E. L., Mullen, J. T., and Lauwers, G. Y. (2015) Familial Gastric Cancers. *The oncologist* **20**, 1365-1377

82. Corso, G., Roncalli, F., Marrelli, D., Carneiro, F., and Roviello, F. (2013) History, pathogenesis, and management of familial gastric cancer: original study of John XXIII's family. *BioMed research international* **2013**, 385132

83. Sereno, M., Aguayo, C., Guillen Ponce, C., Gomez-Raposo, C., Zambrana, F., Gomez-Lopez, M., and Casado, E. (2011) Gastric tumours in hereditary cancer syndromes: clinical features, molecular biology and strategies for prevention. *Clinical & translational oncology : official publication of the Federation of Spanish Oncology Societies and of the National Cancer Institute of Mexico* **13**, 599-610

84. Caldas, C., Carneiro, F., Lynch, H. T., Yokota, J., Wiesner, G. L., Powell, S. M., Lewis, F. R., Huntsman, D. G., Pharoah, P. D., Jankowski, J. A., MacLeod, P., Vogelsang, H., Keller, G., Park, K. G., Richards, F. M., Maher, E. R., Gayther, S. A., Oliveira, C., Grehan, N., Wight, D., Seruca, R., Roviello, F., Ponder, B. A., and Jackson, C. E. (1999) Familial gastric cancer: overview and guidelines for management. *Journal of medical genetics* **36**, 873-880

85. Stoffel, E. M. (2016) Heritable Gastrointestinal Cancer Syndromes. *Gastroenterology clinics of North America* **45**, 509-527

86. Beggs, A. D., Latchford, A. R., Vasen, H. F., Moslein, G., Alonso, A., Aretz, S., Bertario, L., Blanco, I., Bulow, S., Burn, J., Capella, G., Colas, C., Friedl, W., Moller, P., Hes, F. J., Jarvinen, H., Mecklin, J. P., Nagengast, F. M., Parc, Y., Phillips, R. K., Hyer, W., Ponz de Leon, M., Renkonen-Sinisalo, L., Sampson, J. R., Stormorken, A., Tejpar, S., Thomas, H. J., Wijnen, J. T., Clark, S. K., and Hodgson, S. V. (2010) Peutz-Jeghers syndrome: a systematic review and recommendations for management. *Gut* **59**, 975-986

87. Sehgal, R., Sheahan, K., O'Connell, P. R., Hanly, A. M., Martin, S. T., and Winter, D. C. (2014) Lynch syndrome: an updated review. *Genes* **5**, 497-507

88. Aihara, H., Kumar, N., and Thompson, C. C. (2014) Diagnosis, surveillance, and treatment strategies for familial adenomatous polyposis: rationale and update. *European journal of gastroenterology & hepatology* **26**, 255-262

89. Brosens, L. A., Langeveld, D., van Hattem, W. A., Giardiello, F. M., and Offerhaus, G. J. (2011) Juvenile polyposis syndrome. *World journal of gastroenterology* **17**, 4839-4844

90. Nielsen, F. C., van Overeem Hansen, T., and Sorensen, C. S. (2016) Hereditary breast and ovarian cancer: new genes in confined pathways. *Nature reviews. Cancer* **16**, 599-612

91. Chial, H. (2008) Mendelian Genetics: Patterns of Inheritance and Single-Gene Disorders. *Nature Education* **1**, 63

92. Cheadle, J. P., and Sampson, J. R. (2007) MUTYH-associated polyposis--from defect in base excision repair to clinical genetic testing. *DNA repair* **6**, 274-279

93. Cooper, D. N., Krawczak, M., Polychronakos, C., Tyler-Smith, C., and Kehrer-Sawatzki, H. (2013) Where genotype is not predictive of phenotype: towards an understanding of the molecular basis of reduced penetrance in human inherited disease. *Human genetics* **132**, 1077-1130

94. Tung, N., Domchek, S. M., Stadler, Z., Nathanson, K. L., Couch, F., Garber, J. E., Offit, K., and Robson, M. E. (2016) Counselling framework for moderate-penetrance cancer-susceptibility mutations. *Nature reviews. Clinical oncology* **13**, 581-588

95. Yurgelun, M. B., Kulke, M. H., Fuchs, C. S., Allen, B. A., Uno, H., Hornick, J. L., Ukaegbu, C. I., Brais, L. K., McNamara, P. G., Mayer, R. J., Schrag, D., Meyerhardt, J. A., Ng, K., Kidd, J., Singh, N., Hartman, A. R., Wenstrup, R. J., and Syngal, S. (2017) Cancer Susceptibility Gene Mutations in Individuals With Colorectal Cancer. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* **35**, 1086-1095

96. Pharoah, P. D., Antoniou, A. C., Easton, D. F., and Ponder, B. A. (2008) Polygenes, risk prediction, and targeted prevention of breast cancer. *The New England journal of medicine* **358**, 2796-2803

97. Pitteloud, N., Quinton, R., Pearce, S., Raivio, T., Acierno, J., Dwyer, A., Plummer, L., Hughes, V., Seminara, S., Cheng, Y. Z., Li, W. P., Maccoll, G., Eliseenkova, A. V., Olsen, S. K., Ibrahimi, O. A., Hayes, F. J., Boepple, P., Hall, J. E., Bouloux, P., Mohammadi, M., and Crowley, W. (2007) Digenic mutations account for variable phenotypes in idiopathic hypogonadotropic hypogonadism. *The Journal of clinical investigation* **117**, 457-463

98. Shawky, R. M. (2014) Reduced penetrance in human inherited disease. *The Egyptian Journal of Medical Human Genetics* **15**, 103-111

99. Van Heyningen, V., and Yeyati, P. L. (2004) Mechanisms of non-Mendelian inheritance in genetic disease. *Human molecular genetics* **13 Spec No 2**, R225-233

100. Sahasrabudhe, R., Lott, P., Bohorquez, M., Toal, T., Estrada, A. P., Suarez, J. J., Brea-Fernandez, A., Cameselle-Teijeiro, J., Pinto, C., Ramos, I., Mantilla, A., Prieto, R., Corvalan, A., Norero, E., Alvarez, C., Tapia, T., Carvallo, P., Gonzalez, L. M., Cock-Rada, A., Solano, A., Neffa, F., Della Valle, A., Yau, C., Soares, G., Borowsky, A., Hu, N., He, L. J., Han, X. Y., Taylor, P. R., Goldstein, A. M., Torres, J., Echeverry, M., Ruiz-Ponte, C., Teixeira, M. R., and Carvajal-Carmona, L. G. (2017) Germline Mutations in PALB2, BRCA1, and RAD51C, Which Regulate DNA Recombination Repair, in Patients With Gastric Cancer. *Gastroenterology* **152**, 983-986.e986

101. Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., Grody, W. W., Hegde, M., Lyon, E., Spector, E., Voelkerding, K., and Rehm, H. L. (2015) Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genetics in medicine : official journal of the American College of Medical Genetics* **17**, 405-424

102. http://www.uniprot.org/help/variant.

103. (2017) UniProt: the universal protein knowledgebase. *Nucleic acids research* **45**, D158-d169

104. http://omim.org/about.

105. Amberger, J. S., Bocchini, C. A., Schiettecatte, F., Scott, A. F., and Hamosh, A. (2015) OMIM.org: Online Mendelian Inheritance in Man (OMIM(R)), an online catalog of human genes and genetic disorders. *Nucleic acids research* **43**, D789-798

106. https://www.ncbi.nlm.nih.gov/clinvar/intro/.

107. Landrum, M. J., Lee, J. M., Benson, M., Brown, G., Chao, C., Chitipiralla, S., Gu, B., Hart, J., Hoffman, D., Hoover, J., Jang, W., Katz, K., Ovetsky, M., Riley, G., Sethi, A., Tully, R., Villamarin-Salomon, R., Rubinstein, W., and Maglott, D. R. (2016) ClinVar: public archive of interpretations of clinically relevant variants. *Nucleic acids research* **44**, D862-868

108. Forbes, S. A., Bhamra, G., Bamford, S., Dawson, E., Kok, C., Clements, J., Menzies, A., Teague, J. W., Futreal, P. A., and Stratton, M. R. (2008) The Catalogue of Somatic Mutations in Cancer (COSMIC). *Current protocols in human genetics* **Chapter 10**, Unit 10.11

109. Sherry, S. T., Ward, M. H., Kholodov, M., Baker, J., Phan, L., Smigielski, E. M., and Sirotkin, K. (2001) dbSNP: the NCBI database of genetic variation. *Nucleic acids research* **29**, 308-311

110. Lek, M., Karczewski, K. J., Minikel, E. V., Samocha, K. E., Banks, E., Fennell, T., O'Donnell-Luria, A. H., Ware, J. S., Hill, A. J., Cummings, B. B., Tukiainen, T.,

Birnbaum, D. P., Kosmicki, J. A., Duncan, L. E., Estrada, K., Zhao, F., Zou, J., Pierce-Hoffman, E., Berghout, J., Cooper, D. N., Deflaux, N., DePristo, M., Do, R., Flannick, J., Fromer, M., Gauthier, L., Goldstein, J., Gupta, N., Howrigan, D., Kiezun, A., Kurki, M. I., Moonshine, A. L., Natarajan, P., Orozco, L., Peloso, G. M., Poplin, R., Rivas, M. A., Ruano-Rubio, V., Rose, S. A., Ruderfer, D. M., Shakir, K., Stenson, P. D., Stevens, C., Thomas, B. P., Tiao, G., Tusie-Luna, M. T., Weisburd, B., Won, H. H., Yu, D., Altshuler, D. M., Ardissino, D., Boehnke, M., Danesh, J., Donnelly, S., Elosua, R., Florez, J. C., Gabriel, S. B., Getz, G., Glatt, S. J., Hultman, C. M., Kathiresan, S., Laakso, M., McCarroll, S., McCarthy, M. I., McGovern, D., McPherson, R., Neale, B. M., Palotie, A., Purcell, S. M., Saleheen, D., Scharf, J. M., Sklar, P., Sullivan, P. F., Tuomilehto, J., Tsuang, M. T., Watkins, H. C., Wilson, J. G., Daly, M. J., and MacArthur, D. G. (2016) Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285-291

111. Auton, A., Brooks, L. D., Durbin, R. M., Garrison, E. P., Kang, H. M., Korbel, J. O., Marchini, J. L., McCarthy, S., McVean, G. A., and Abecasis, G. R. (2015) A global reference for human genetic variation. *Nature* **526**, 68-74
112. http://www.umd.be/HSF3/HSF.shtml.
113. http://www.cbs.dtu.dk/services/NetGene2/.
114. http://dbcat.cgm.ntu.edu.tw/.
115. http://www.ensembl.org/index.html.
116. Kurdyukov, S., and Bullock, M. (2016) DNA Methylation Analysis: Choosing the Right Method. *Biology* **5**
117. Livak, K. J., and Schmittgen, T. D. (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods (San Diego, Calif.)* **25**, 402-408
118. Boland, C. R., and Fishel, R. (2005) Lynch syndrome: form, function, proteins, and basketball. *Gastroenterology* **129**, 751-755
119. Mukherjee, S., and Feig, M. (2009) Conformational change in MSH2-MSH6 upon binding DNA coupled to ATPase activity. *Biophysical journal* **96**, L63-65
120. Peltomaki, P. (2016) Update on Lynch syndrome genomics. *Familial cancer* **15**, 385-393
121. Akram, M. (2014) Citric acid cycle and role of its intermediates in metabolism. *Cell biochemistry and biophysics* **68**, 475-478
122. Roy, R., Chun, J., and Powell, S. N. (2011) BRCA1 and BRCA2: different roles in a common pathway of genome protection. *Nature reviews. Cancer* **12**, 68-78
123. http://insight-database.org.
124. http://insight-database.org/classifications/.

125. http://insight-database.org/classifications/index.html?gene=MSH6&variant=c.3150_3162dup&protein=.
126. Goodfellow, P. J., Buttin, B. M., Herzog, T. J., Rader, J. S., Gibb, R. K., Swisher, E., Look, K., Walls, K. C., Fan, M. Y., and Mutch, D. G. (2003) Prevalence of defective DNA mismatch repair and MSH6 mutation in an unselected series of endometrial cancers. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 5908-5913
127. Yu, W., Ni, Y., Saji, M., Ringel, M. D., Jaini, R., and Eng, C. (2017) Cowden syndrome-associated germline succinate dehydrogenase complex subunit D (SDHD) variants cause PTEN-mediated down-regulation of autophagy in thyroid cancer cells. *Human molecular genetics* **26**, 1365-1375
128. Hoekstra, A. S., van den Ende, B., Julia, X. P., van Breemen, L., Scheurwater, K., Tops, C. M., Malinoc, A., Devilee, P., Neumann, H. P., and Bayley, J. P. (2017) Simple and rapid characterization of novel large germline deletions in SDHB, SDHC and SDHD-related paraganglioma. *Clinical genetics* **91**, 536-544
129. Eto, N., Miyagishi, M., Inagi, R., Fujita, T., and Nangaku, M. (2009) Mitogen-activated protein 3 kinase 6 mediates angiogenic and tumorigenic effects via vascular endothelial growth factor expression. *The American journal of pathology* **174**, 1553-1563
130. Wiedemeyer, W. R., Beach, J. A., and Karlan, B. Y. (2014) Reversing Platinum Resistance in High-Grade Serous Ovarian Carcinoma: Targeting BRCA and the Homologous Recombination System. *Frontiers in oncology* **4**, 34
131. https://genome.ucsc.edu/.
132. Valmiki, M. G., and Ramos, J. W. (2009) Death effector domain-containing proteins. *Cellular and molecular life sciences : CMLS* **66**, 814-830
133. Hyland, P. L., Lin, S. W., Hu, N., Zhang, H., Wang, L., Su, H., Wang, C., Ding, T., Tang, Z. Z., Fan, J. H., Qiao, Y. L., Xiong, X., Wheeler, W., Giffen, C., Yu, K., Yuenger, J., Burdett, L., Wang, Z., Chanock, S. J., Tucker, M. A., Dawsey, S. M., Freedman, N. D., Goldstein, A. M., Abnet, C. C., and Taylor, P. R. (2014) Genetic variants in fas signaling pathway genes and risk of gastric cancer. *International journal of cancer* **134**, 822-831
134. van der Groep, P., van der Wall, E., and van Diest, P. J. (2011) Pathology of hereditary breast cancer. *Cellular oncology (Dordrecht)* **34**, 71-88
135. Pal, R., Srivastava, N., Chopra, R., Gochhait, S., Gupta, P., Prakash, N., Agarwal, G., and Bamezai, R. N. (2010) Investigation of DNA damage response and apoptotic

gene methylation pattern in sporadic breast tumors using high throughput quantitative DNA methylation analysis technology. *Molecular cancer* **9**, 303

136. Saling, M., Duckett, J. K., Ackers, I., Coschigano, K., Jenkinson, S., and Malgor, R. (2017) Wnt5a / planar cell polarity signaling pathway in urothelial carcinoma, a potential prognostic biomarker. *Oncotarget* **8**, 31655-31665
137. Orloff, M., Peterson, C., He, X., Ganapathi, S., Heald, B., Yang, Y. R., Bebek, G., Romigh, T., Song, J. H., Wu, W., David, S., Cheng, Y., Meltzer, S. J., and Eng, C. (2011) Germline mutations in MSR1, ASCC1, and CTHRC1 in patients with Barrett esophagus and esophageal adenocarcinoma. *Jama* **306**, 410-419
138. Price, R. J., Lillycrop, K. A., and Burdge, G. C. (2015) Folic acid supplementation in vitro induces cell type-specific changes in BRCA1 and BRCA 2 mRNA expression, but does not alter DNA methylation of their promoters or DNA repair. *Nutrition research (New York, N.Y.)* **35**, 532-544
139. Valvezan, A. J., Huang, J., Lengner, C. J., Pack, M., and Klein, P. S. (2014) Oncogenic mutations in adenomatous polyposis coli (Apc) activate mechanistic target of rapamycin complex 1 (mTORC1) in mice and zebrafish. *Disease models & mechanisms* **7**, 63-71
140. Wang, P., Wang, Y. C., Chen, X. Y., Shen, Z. Y., Cao, H., Zhang, Y. J., Yu, J., Zhu, J. D., Lu, Y. Y., and Fang, J. Y. (2012) CTHRC1 is upregulated by promoter demethylation and transforming growth factor-beta1 and may be associated with metastasis in human gastric cancer. *Cancer science* **103**, 1327-1333
141. Ni, Y., Zbuk, K. M., Sadler, T., Patocs, A., Lobo, G., Edelman, E., Platzer, P., Orloff, M. S., Waite, K. A., and Eng, C. (2008) Germline mutations and variants in the succinate dehydrogenase genes in Cowden and Cowden-like syndromes. *American journal of human genetics* **83**, 261-268
142. Westermeier, F., Bustamante, M., Pavez, M., Garcia, L., Chiong, M., Ocaranza, M. P., and Lavandero, S. (2015) Novel players in cardioprotection: Insulin like growth factor-1, angiotensin-(1-7) and angiotensin-(1-9). *Pharmacological research* **101**, 41-55
143. Nouet, S., Amzallag, N., Li, J. M., Louis, S., Seitz, I., Cui, T. X., Alleaume, A. M., Di Benedetto, M., Boden, C., Masson, M., Strosberg, A. D., Horiuchi, M., Couraud, P. O., and Nahmias, C. (2004) Trans-inactivation of receptor tyrosine kinases by novel angiotensin II AT2 receptor-interacting protein, ATIP. *The Journal of biological chemistry* **279**, 28989-28997
144. Astuti, D., Morris, M., Krona, C., Abel, F., Gentle, D., Martinsson, T., Kogner, P., Neumann, H. P., Voutilainen, R., Eng, C., Rustin, P., Latif, F., and Maher, E. R.

(2004) Investigation of the role of SDHB inactivation in sporadic phaeochromocytoma and neuroblastoma. *British journal of cancer* **91**, 1835-1841

145. Bozgeyik, I., Yumrutas, O., and Bozgeyik, E. (2017) MTUS1, a gene encoding angiotensin-II type 2 (AT2) receptor-interacting proteins, in health and disease, with special emphasis on its role in carcinogenesis. *Gene* **626**, 54-63
146. Liu, Z., Zhu, G., Getzenberg, R. H., and Veltri, R. W. (2015) The Upregulation of PI3K/Akt and MAP Kinase Pathways is Associated with Resistance of Microtubule-Targeting Drugs in Prostate Cancer. *Journal of cellular biochemistry* **116**, 1341-1349
147. Hanahan, D., and Weinberg, R. A. (2011) Hallmarks of cancer: the next generation. *Cell* **144**, 646-674
148. Woods, M. O., Hyde, A. J., Curtis, F. K., Stuckless, S., Green, J. S., Pollett, A. F., Robb, J. D., Green, R. C., Croitoru, M. E., Careen, A., Chaulk, J. A., Jegathesan, J., McLaughlin, J. R., Gallinger, S. S., Younghusband, H. B., Bapat, B. V., and Parfrey, P. S. (2005) High frequency of hereditary colorectal cancer in Newfoundland likely involves novel susceptibility genes. *Clinical cancer research : an official journal of the American Association for Cancer Research* **11**, 6853-6861
149. Barnetson, R. A., Cartwright, N., van Vliet, A., Haq, N., Drew, K., Farrington, S., Williams, N., Warner, J., Campbell, H., Porteous, M. E., and Dunlop, M. G. (2008) Classification of ambiguous mutations in DNA mismatch repair genes identified in a population-based study of colorectal cancer. *Human mutation* **29**, 367-374
150. de Miranda, N. F., van Dinther, M., van den Akker, B. E., van Wezel, T., ten Dijke, P., and Morreau, H. (2015) Transforming Growth Factor beta Signaling in Colorectal Cancer Cells With Microsatellite Instability Despite Biallelic Mutations in TGFBR2. *Gastroenterology* **148**, 1427-1437.e1428
151. Saburi, S., Hester, I., Fischer, E., Pontoglio, M., Eremina, V., Gessler, M., Quaggin, S. E., Harrison, R., Mount, R., and McNeill, H. (2008) Loss of Fat4 disrupts PCP signaling and oriented cell division and leads to cystic kidney disease. *Nature genetics* **40**, 1010-1015
152. Yoshida, S., Yamashita, S., Niwa, T., Mori, A., Ito, S., Ichinose, M., and Ushijima, T. (2017) Epigenetic inactivation of FAT4 contributes to gastric field cancerization. *Gastric cancer : official journal of the International Gastric Cancer Association and the Japanese Gastric Cancer Association* **20**, 136-145

VIII. Supplementary Material

Table S1. *Germline Variants identified in FIGC families by multiplex custom-panel based sequencing.*

Family	Gene	Chr	Position GRCh38	cDNA	Protein	Consequence	Freq cohort	Population Frequency	Illumina's Classification	UniProt Classification	OMIM Classification	ClinVar Classification	Annotator's Classification	Cosmic Classification	FATHMN prediction	HGMD	Final Classification
1	MSH6	2	47799219	c.1236G>C	p.Lys412Asn	Missense	2%	NA	VUS	NA	NA	NA	NA	-	-	-	VUS
	MSH6	2	47800241	c.2258C>T	p.Ser753Phe	Missense	2%	NA	VUS	NA	NA	Uncertain Significance	Uncertain Significance - 1	-	-	-	VUS
	SDHD	11	112086941	c.34G>A	p.Gly12Ser	Missense	4%	Very Rare Global (1%), Rare Eur (2%), Rare Tusc (4,7%)	Pathogenic	Likely Pathogenic	Likely Pathogenic	Uncertain Significance	Likely Pathogenic - 2; Uncertain Significance - 1	-	-	-	Conflicting
2	SDHD	11	112086945	c.38C>T	p.Ala13Val	Missense	2%	Very Rare Global (0,006%), Very Rare Eur (0%)	VUS	NA	NA	NA	NA	-	-	-	Likely Pathogenic
3	MAP3K6	1	27362252	c.1256-2A>G	-	Splice acceptor	4%	Very Rare Global (0,4%), Rare Eur (1,4%), Very Rare Tusc (0,9%)	Likely Pathogenic	NA	NA	NA	NA	-	-	-	Likely Pathogenic
4	MSH2	2	47429940	c.1275A>G	c.1275A>G(p.=)	Splice region, Synonymous	2%	Very Rare Global (0,02%), Very Rare Eur (0,05%)	VUS	NA	NA	Uncertain Significance	Uncertain Significance - 1	-	-	Yes	Likely Pathogenic
	SDHD	11	112087953	c.149A>G	p.His50Arg	Missense	2%	Rare Global (1,3%), Rare Eur (3,4%), Rare Tusc (2,8%)	Pathogenic	Likely Pathogenic	Likely Pathogenic	Uncertain Significance	Likely Pathogenic - 2; Uncertain Significance - 1	-	-	Yes	Conflicting
5	ATM	11	108329200	c.7269A>T	p.Glu2423Asp	Missense	2%	NA	VUS	NA	NA	Uncertain Significance	Uncertain Significance - 1	Yes	Pathogenic (0,99)		VUS
6	ITIH2	10	7731989	c.1640C>T	p.Ala547Val	Missense	2%	Very Rare Global (0,02%), Very Rare Eur (0,004%)	VUS	NA	NA	NA	NA	-	-	-	Likely Pathogenic
7	MAP3K6	1	27358259	c.2837C>T	p.Pro946Leu	Missense	2%	Very Rare Global (0,4%), Very Rare Eur (1%), Very Rare Tusc (0,9%)	VUS	NA	NA	NA	NA	-	-	-	Likely Pathogenic
	BRCA1	17	43125268	c.-20+2dupT	-	Splice region, Intron	2%	NA	VUS	NA	NA	NA	NA	-	-	-	VUS
8	MAP3K6	1	27362252	c.1256-2A>G	-	Splice acceptor	4%	Very Rare Global (0,4%), Rare Eur (1,4%), Very Rare Tusc (0,9%)	Likely Pathogenic	NA	NA	NA	NA	-	-	-	Likely Pathogenic
	CASP10	2	201193048	c.506G>T	p.Cys169Phe	Missense	2%	Very Rare Global (0,002%), Very Rare Eur (0,02%)	VUS	NA	NA	NA	NA	-	-	-	Likely Pathogenic
9	MSH6	2	47800616	c.2633T>C	p.Val878Ala	Missense	8%	Very Rare Global (0,8%), Rare Eur (2,6%), Rare Tusc (4,7%)	Pathogenic	Likely Pathogenic	Likely Pathogenic	Likely Benign	Likely Pathogenic - 2; Likely Benign - 1	-	-	-	Conflicting
10	MSH6	2	47800616	c.2633T>C	p.Val878Ala	Missense	8%	Very Rare Global (0,8%), Rare Eur (2,6%), Rare Tusc (4,7%)	Pathogenic	Likely Pathogenic	Likely Pathogenic	Likely Benign	Likely Pathogenic - 2; Likely Benign - 1	-	-	-	Conflicting
11	MSH6	2	47800616	c.2633T>C	p.Val878Ala	Missense	8%	Very Rare Global (0,8%), Rare Eur (2,6%), Rare Tusc (4,7%)	Pathogenic	Likely Pathogenic	Likely Pathogenic	Likely Benign	Likely Pathogenic - 2; Likely Benign - 1	-	-	-	Conflicting
	ATM	11	108289671	c.4306C>T	p.His1436Tyr	Missense	2%	Very Rare Global (0,04%), Very Rare Eur (0,2%), Very Rare Tusc (0,9%)	VUS	NA	NA	Uncertain Significance	Uncertain Significance - 1	-	-	-	Likely Pathogenic
12	CTHRC1	8	103371707	c.57_62dupCCTGCT	p.Leu20_Leu21dup	Inframe insertion	2%	NA	VUS	NA	NA	NA	NA	-	-	-	VUS
	BRCA2	13	32337163	c.2812_2815dupGCAA	p.Thr939SerfsTer7	Frameshift Indels	2%	NA	Likely Pathogenic	NA	NA	NA	NA	-	-	-	Likely Pathogenic
13	SDHB	1	17027802	c.487T>C	p.Ser163Pro	Missense	4%	Rare Global (1,9%), Rare Eur (3,4%), Rare Tusc (3,7%)	Pathogenic	Likely Pathogenic	Likely Pathogenic	Uncertain Significance	Likely Pathogenic – 2; Uncertain Significance - 1	-	-	-	Conflicting
	MTUS1	8	17655975	c.2996A>G	p.Glu999Gly	Missense	2%	Very Rare Global (0,6%), Rare Eur (2,6%), Rare Tusc (3,7%)	VUS	NA	NA	NA	NA	-	-	-	Likely Pathogenic
14	MTUS1	8	17755177	c.631T>G	p.Ser211Ala	Missense	2%	Very Rare Global (0,8%), Rare Eur (3,4%), Rare Tusc (3,7%)	VUS	NA	NA	NA	NA	-	-	-	Likely Pathogenic
	ATM	11	108267198	c.2494C>T	p.Arg832Cys	Missense	2%	Very Rare Global (0,04%), Very Rare Eur (0%), Very Rare Tusc (0%)	VUS	NA	NA	Uncertain Significance	Uncertain Significance - 1	Yes	Neutral	-	Likely Pathogenic
15	MTUS1	8	17684434	c.2732A>C	p.Lys911Thr	Missense	2%	Very Rare Global (0,9%), Rare Eur (1,8%), Very Rare Tusc (0,9%)	VUS	Likely Benign	NA	NA	Likely Benign - 1	-	-	-	Likely Benign
16	MSH2	2	47475052	c.1787A>G	p.Asn596Ser	Missense	2%	Very Rare Global (0,06%), Very Rare Eur (0,07%)	VUS	Likely Pathogenic	NA	Uncertain Significance	Likely Pathogenic – 1; Uncertain Significance - 1	Yes	Pathogenic (0,99)	-	Conflicting
	TGFBR2	3	30671751	c.643C>T	p.Arg215Cys	Missense	2%	Very Rare Global (0,02%), Very Rare Eur (0%)	VUS	NA	NA	NA	NA	-	-	-	Likely Pathogenic
	FAT4	4	125316935	c.524G>T	p.Arg175Leu	Missense	2%	Very Rare Global (0,2%), Very Rare Eur (1%)	VUS	NA	NA	Likely Benign	Likely Benign - 1	-	-	-	Likely Benign
17	APC	5	112839078	c.3484T>C	p.Tyr1162His	Missense	2%	NA	VUS	NA	NA	NA	NA	-	-	-	VUS
18	MSH6	2	47800616	c.2633T>C	p.Val878Ala	Missense	8%	Very Rare Global (0,8%), Rare Eur (2,6%), Rare Tusc (4,7%)	Pathogenic	Likely Pathogenic	Likely Pathogenic	Likely Benign	Likely Pathogenic - 2; Likely Benign - 1	-	-	-	Conflicting
19	MAP3K6	1	27356633	c.3481C>G	p.Pro1161Ala	Missense	2%	Very Rare Global (0,005%), Very Rare Eur (0,009%)	VUS	NA	NA	NA	NA	-	-	-	Likely Pathogenic
	MSR1	8	16155085	c.877C>T	p.Arg293Ter	Stop gained	2%	Very Rare Global (1,1%), Rare Eur (2,4%), Rare Tusc (1,7%)	Pathogenic	NA	Likely Pathogenic	Uncertain Significance	Likely Pathogenic - 1; Uncertain Significance - 1	Yes	Neutral (0,08)	-	Likely Pathogenic
20	FAT4	4	125451863	c.10847C>T	p.Thr3616Met	Missense	2%	Very Rare Global (0,3%), Very Rare Eur (0,6%), Very Rare Tusc (0,9%)	VUS	NA	NA	Likely Benign	Likely Benign - 1	-	-	-	Likely Benign
21	SDHD	11	112086941	c.34G>A	p.Gly12Ser	Missense	4%	Very Rare Global (1%), Rare Eur (2%), Rare Tusc (4,7%)	Pathogenic	Likely Pathogenic	Likely Pathogenic	Uncertain Significance	Likely Pathogenic - 2; Uncertain Significance - 1	Yes	Neutral (0,02)	-	Conflicting
22	SDHB	1	17027802	c.487T>C	p.Ser163Pro	Missense	4%	Rare Global (1,9%), Rare Eur (3,4%), Rare Tusc (3,7%)	Pathogenic	Likely Pathogenic	Likely Pathogenic	Uncertain Significance	Likely Pathogenic – 2; Uncertain Significance - 1	-	-	-	Conflicting
23	CDH1	16	68813323	c.1148A>G	p.Gln383Arg	Missense	2%	NA	VUS	NA	NA	NA	NA	-	-	-	VUS
24	MSR1	8	16168606	c.482C>A	p.Thr161Asn	Missense	2%	Very Rare Global (0,08%), Very Rare Eur (0,2%) Very Rare Tusc (0,9%)	VUS	NA	NA	NA	NA	-	-	-	Likely Pathogenic

25-52 variants were not found

Description of germline variants, including both frequency in the cohort (freq cohort: % of mutated families), ExAC and 1000genomes and Clinical Significance classification from Illumina's Variant Interpreter, UniProt, OMIM, ClinVar, Annotator57 (in the 3 databases) and Integrated classification. Eur: Europe; Tusc: Tuscany; VUS: Variant of Unknown Significance; NA: Not available; Very Rare: <1%; Rare: >1, <5; Not Rare: >5.

Table S2. *Somatic Variants identified in FIGC families by multiplex custom-panel based sequencing.*

Family	Gene	Chr	Position GRCh38	cDNA	Protein	Consequence	Freq cohort	Frequency	Illumina's Classification	UniProt Classification	OMIM Classification	ClinVar Classification	Annotator57's Classification	Cosmic Classification	FATHMN prediction	HGMD	Integrated Classification
1	ARID1A	1	26771139	c.3219G>A	p.Trp1073Ter	Stop gained	2%	-	Likely Pathogenic	NA	NA	NA	NA	Yes	Pathogenic (0,98)	-	VUS
	MSH6	2	47801146	c.3163G>A	p.Ala1055Thr	Missense	2%	Very Rare Global (0,002%)	VUS	NA	NA	Uncertain Significance	Uncertain Significance - 1	-	-	-	VUS
	FAT4	4	125406935	c.5363G>A	p.Arg1788His	Missense	2%	-	VUS	NA	NA	NA	NA	Yes	Pathogenic (0,91)	-	VUS
	AKAP12	6	151351864	c.3479dupC	p.Asp1161Ter	Frameshift Indels	2%	-	Likely Pathogenic	NA	NA	NA	NA	-	-	-	VUS
	ATM	11	108229266	c.278delA	p.Lys93ArgfsTer23	Frameshift Indels	2%	-	Likely Pathogenic	NA	NA	NA	NA	-	-	-	VUS
	BRCA2	13	32339422	c.5073delA	p.Lys1691AsnfsTer15	Frameshift Indels	2%	-	Pathogenic	NA	NA	Likely Pathogenic	Likely Pathogenic - 1	-	-	Yes	Likely Pathogenic
	IDH2	15	90088686	c.435delG	p.Thr146LeufsTer15	Frameshift Indels	2%	-	Likely Pathogenic	NA	NA	NA	NA	-	-	-	VUS
	PALB2	16	23635707	c.839delA	p.Asn280ThrfsTer8	Frameshift Indels	2%	-	Likely Pathogenic	NA	NA	NA	NA	-	-	-	VUS
2	TP53	17	7674953	c.578A>T	p.His193Leu	Missense	2%	-	Likely Pathogenic	Likely Pathogenic	NA	Likely Pathogenic	Likely Pathogenic - 2			-	Likely Pathogenic
3	BRCA2	13	32332779	c.1310_1313delAAGA	p.Lys437IlefsTer22	Frameshift Indels	2%	Very Rare Global (0,002%)	Likely Pathogenic	NA	NA	Likely Pathogenic	Likely Pathogenic - 1			Yes	Likely Pathogenic
	TP53	17	7673764	c.856G>T	p.Glu286Ter	Stop gained	2%	-	Likely Pathogenic	NA	NA	Pathogenic	Pathogenic - 1	Yes	Pathogenic (1,00)	-	Pathogenic
4-6	Somatic variants were not found																
7	TP53	17	7675109	c.503A>T	p.His168Leu	Missense	2%	-	VUS	Likely Pathogenic	NA	NA	Likely Pathogenic - 1	Yes	Pathogenic (0,97)	-	Likely Pathogenic
8	ARID1A	1	26773716	c.4003C>T	p.Arg1335Ter	Stop gained, Splice region	2%	-	Pathogenic	NA	Likely Pathogenic	Likely Pathogenic	Likely Pathogenic - 2	Yes	Pathogenic (0,87)	-	Likely Pathogenic
	FAT4	4	125449221	c.8205A>C	p.Lys2735Asn	Missense	2%	-	VUS	NA	NA	NA	NA	Yes	None (0,51)	-	VUS
	PTEN	10	87952216	c.595_597delATG	p.Met199del	Inframe deletion	2%	-	VUS	NA	NA	Uncertain Significance	Uncertain Significance - 1	Yes	None	Yes	VUS
	PTEN	10	87958013	c.800delA	p.Lys267ArgfsTer9	Frameshift Indels	2%	-	Pathogenic	NA	NA	Pathogenic	Pathogenic - 1	Yes	None	-	Pathogenic
9	TP53	17	7674238	c.725G>T	p.Cys242Phe	Missense	2%	-	VUS	Likely Pathogenic	NA	Likely Pathogenic	Likely Pathogenic – 2	Yes	Pathogenic (0,99)	-	Likely Pathogenic
10	ARID1A	1	26780111	c.6214delG	p.Asp2072ThrfsTer63	Frameshift Indels	2%	-	Likely Pathogenic	NA	NA	NA	NA	-	-	-	VUS
	CTNNA1	5	138824612	c.671delC	p.Ala224AspfsTer20	Frameshift Indels	2%	-	Likely Pathogenic	NA	NA	NA	NA	-	-	-	VUS
	TP53	17	7674221	c.742C>T	p.Arg248Trp	Missense	4%	Very Rare Global (0,002%)	Pathogenic	Likely Pathogenic	Likely Pathogenic	Likely Pathogenic	Likely Pathogenic - 3			-	Likely Pathogenic
11	CDH1	16	68819423	c.1709A>G	p.Asn570Ser	Missense, Splice region	2%	-	VUS	NA	NA	NA	NA			-	VUS
	TP53	17	7674947	c.584T>C	p.Ile195Thr	Missense	2%	Very Rare Global (0,002%)	VUS	Likely Pathogenic	NA	Likely Pathogenic	Likely Pathogenic - 2	Yes	Pathogenic (0,99)	Yes	Likely Pathogenic
12	APC	5	112842961	c.7367T>A	p.Leu2456Ter	Stop gained	2%	-	Likely Pathogenic	NA	NA	NA	NA			-	VUS
13-15	Somatic variants were not found																
16	ARID1A	1	26771131	c.3216delA	p.Lys1072AsnfsTer21	Frameshift Indels	2%	-	Likely Pathogenic	NA	NA	NA	NA	Yes	None (0,00)	-	VUS
	FAT4	4	125451167	c.10151C>T	p.Ala3384Val	Missense	2%	-	VUS	NA	NA	NA	NA			-	VUS
	MSH3	5	80675096	c.1148delA	p.Lys383ArgfsTer32	Frameshift Indels	25%	-	Likely Pathogenic	NA	Likely Pathogenic	Likely Pathogenic	Likely Pathogenic - 2	Yes	None	-	Likely Pathogenic
	PTEN	10	87933147	c.388C>T	p.Arg130Ter	Stop gained	2%	Very Rare Global (0,002%)	Pathogenic	NA	Likely Pathogenic	Pathogenic	Pathogenic - 1; Likely Pathogenic - 1			-	Likely Pathogenic
	PTEN	10	87961042	c.955_958delACTT	p.Thr319Ter	Frameshift Indels	2%	-	Pathogenic	NA	NA	Pathogenic	Pathogenic - 1	Yes	None (0,00)	-	Pathogenic
	ATM	11	108244846	c.721A>T	p.Lys241Ter	Stop gained	2%	-	Likely Pathogenic	NA	NA	NA	NA			-	VUS
17	Somatic variants were not found																
18	ARID1A	1	26774926	c.4703delC	p.Pro1568LeufsTer44	Frameshift Indels	2%	-	Likely Pathogenic	NA	NA	NA	NA	-	-	-	VUS
	MSH3	5	80675096	c.1148delA	p.Lys383ArgfsTer32	Frameshift Indels	25%	-	Likely Pathogenic	NA	Likely Pathogenic	Likely Pathogenic	Likely Pathogenic - 2	Yes	None	-	Likely Pathogenic
19	ARID1A	1	26779533	c.5635C>T	p.Arg1879Trp	Missense	2%	Very Rare Global (0,008%)	VUS	NA	NA	NA	NA	Yes	Pathogenic (0,78)	-	VUS
	MSH6	2	47803500	c.3261dupC	p.Phe1088LeufsTer5	Frameshift Indels	4%	-	Pathogenic	NA	NA	Pathogenic	Pathogenic - 1	Yes	None (0,00)	Yes	Pathogenic
	MSH3	5	80675096	c.1148delA	p.Lys383ArgfsTer32	Frameshift Indels	25%	-	Likely Pathogenic	NA	NA	NA	Likely Pathogenic - 2	-	-	-	Likely Pathogenic
20	TP53	17	7675228	c.380_384delCCCCT	p.Ser127CysfsTer20	Frameshift Indels	2%	-	Likely Pathogenic	Likely Pathogenic	NA	Likely Pathogenic	NA	Yes	Pathogenic (0,99)	-	VUS
21	TP53	17	7674250	c.713G>T	p.Cys238Phe	Missense	2%	-	VUS	NA	NA	NA	Likely Pathogenic - 2			-	Likely Pathogenic
22	FAT4	4	125450920	c.9904C>A	p.Leu3302Ile	Missense	2%	-	VUS	NA	NA	NA	NA	-	-	-	VUS
	TP53	17	7674208	c.743_755delGGAGGCCCATCCT	p.Arg248ProfsTer93	Frameshift Indels	2%	-	Likely Pathogenic	NA	NA	NA	NA	Yes	Pathogenic (1,00)	-	VUS
23	ARID1A	1	26766552	c.2974G>T	p.Glu992Ter	Stop gained	2%	-	Likely Pathogenic	NA	NA	NA	NA	Yes	Pathogenic (0,96)	-	VUS
	ARID1A	1	26772565	c.3472G>T	p.Gly1158Ter	Stop gained	2%	-	Likely Pathogenic	NA	NA	Pathogenic	NA			-	VUS
	MSH2	2	47480873	c.2634+4delT	-	Splice donor	2%	-	Likely Pathogenic	NA	NA	NA	Likely Pathogenic - 1			-	Pathogenic
	TGFBR2	3	30688506	c.1594C>A	p.His532Asn	Missense	2%	-	VUS	NA	NA	NA	NA			-	VUS
	MLH1	3	37040259	c.1632A>T	p.Gln544His	Missense	2%	-	VUS	NA	NA	NA	NA			-	VUS
	FAT4	4	125451730	c.10714A>G	p.Ile3572Val	Missense	2%	-	VUS	NA	Likely Pathogenic	Likely Pathogenic	NA	Yes	None	-	VUS
	MSH3	5	80675096	c.1148delA	p.Lys383ArgfsTer32	Frameshift Indels	25%	-	Likely Pathogenic	NA	NA	NA	Likely Pathogenic - 2			-	Likely Pathogenic
	APC	5	112837681	c.2087A>G	p.Glu696Gly	Missense	2%	-	VUS	NA	NA	NA	NA			-	VUS
	CTNNA1	5	138783232	c.161G>A	p.Arg54His	Missense	2%	Very Rare Global (0,005%)	VUS	NA	NA	NA	NA			-	VUS
	CTNNA1	5	138824597	c.656C>T	p.Pro219Leu	Missense	2%	Very Rare Global (0,003%)	VUS	NA	NA	NA	NA			-	VUS
	AKAP12	6	151350826	c.2435C>A	p.Pro812His	Missense	2%	-	VUS	NA	NA	NA	NA			-	VUS
	ATM	11	108259061	c.2452A>G	p.Ile818Val	Missense	2%	-	VUS	NA	NA	NA	NA			Yes	VUS
	ATM	11	108271249	c.2922-2A>C	-	Splice acceptor	2%	-	Likely Pathogenic	NA	NA	Uncertain Significance	NA			-	Likely Pathogenic
	ATM	11	108304802	c.5624G>A	p.Arg1875Gln	Missense	2%	Very Rare Global (0,002%)	VUS	NA	NA	NA	Uncertain Significance – 1			-	VUS
	BRCA2	13	32398361	c.9848T>A	p.Val3283Asp	Missense	2%	-	VUS	NA	NA	NA	NA			-	VUS
	PALB2	16	23629845	c.2309C>A	p.Ala770Asp	Missense	2%	-	VUS	NA	NA	NA	NA	-	-	Yes	VUS
24	Somatic variants were not found																

Table S2 (cont.). *Somatic Variants identified in FIGC families by multiplex custom-panel based sequencing.*

Family	Gene	Chr	Position GRCh38	cDNA	Protein	Consequence	Freq cohort	Frequency	Illumina's Classification	UniProt Classification	OMIM Classification	ClinVar Classification	Annotator57's Classification	Cosmic Classification	FATHMN prediction	HGMD	Integrated Classification
25	APC	5	112838478	c.2884G>C	p.Asp962His	Missense	2%	-	VUS	Likely Pathogenic	NA	Likely Pathogenic	NA	-	-	-	VUS
	TP53	17	7675139	c.473G>A	p.Arg158His	Missense	2%	Very Rare Global (0,002%)	Pathogenic	Likely Pathogenic	NA	NA	Likely Pathogenic - 2	-	-	-	Likely Pathogenic
26	TP53	17	7675235	c.377A>G	p.Tyr126Cys	Missense, Splice region	2%	-	VUS	NA	NA	Pathogenic	Likely Pathogenic - 1	Yes	None (0,00)	Yes	Likely Pathogenic
27	MSH6	2	47803501	c.3261delC	p.Phe1088SerfsTer2	Frameshift Indels	6%	-	Pathogenic	NA	Likely Pathogenic	Likely Pathogenic	Pathogenic - 1	Yes	None	-	Pathogenic
	MSH3	5	80675096	c.1148delA	p.Lys383ArgfsTer32	Frameshift Indels	25%	-	Likely Pathogenic	NA	Likely Pathogenic	Likely Pathogenic	Likely Pathogenic - 2	Yes	None	-	Likely Pathogenic
	CTNNA1	5	138781963	c.39G>A	p.Trp13Ter	Stop gained	2%	-	Likely Pathogenic	NA	NA	NA	NA			-	VUS
	CTNNA1	5	138930610	c.2150dupA	p.Gln718AlafsTer16	Frameshift Indels	2%	-	Likely Pathogenic	NA	NA	NA	NA	-	-	-	VUS
28	TP53	17	7673802	c.818G>A	p.Arg273His	Missense	2%	Very Rare Eur (0,2%), Very Rare Tusc (0%)	Pathogenic	Likely Pathogenic	Likely Pathogenic	Likely Pathogenic	Likely Pathogenic - 3			-	Likely Pathogenic
29	TP53	17	7674240	c.723delC	p.Cys242AlafsTer5	Frameshift Indels	2%	-	Likely Pathogenic	NA	NA	NA	NA	Yes	None	-	VUS
30	TP53	17	7674246	c.716dupA	p.Asn239LysfsTer25	Frameshift Indels	2%	-	Likely Pathogenic	Likely Pathogenic	NA	NA	Likely Pathogenic - 1	-	-	-	Likely Pathogenic
31	ARID1A	1	26762296	c.2402dupG	p.Gln802SerfsTer15	Frameshift Indels	2%	-	Likely Pathogenic	NA	NA	NA	Likely Pathogenic - 1			-	Likely Pathogenic
	ARID1A	1	26774648	c.4424delA	p.Asn1475ThrfsTer6	Frameshift Indels	2%	-	Likely Pathogenic	NA	NA	NA	Likely Pathogenic - 1			-	Likely Pathogenic
	MLH1	3	36993588	c.41C>T	p.Thr14Ile	Missense	2%	Very Rare Global (0,002%)	VUS	NA	NA	Uncertain Significance	Uncertain significance - 1			-	VUS
	FAT4	4	125316934	c.523C>T	p.Arg175Cys	Missense	2%	Very rare Global (0,002%)	VUS	NA	NA	NA	NA			-	VUS
	APC	5	112842672	c.7078G>T	p.Gly2360Cys	Missense	2%	-	VUS	NA	NA	NA	NA			-	VUS
	HSPA5	9	125240909	c.123-2A>G	-	Splice acceptor	2%	-	Likely Pathogenic	NA	NA	NA	NA			-	VUS
32	TP53	17	7674250	c.713G>C	p.Cys238Ser	Missense	2%	-	VUS	Likely Pathogenic	NA	Likely Pathogenic	Likely Pathogenic – 2	Yes	Pathogenic (0,99)	-	Likely Pathogenic
33	MSH6	2	47795898	c.465delA	p.Lys155AsnfsTer19	Frameshift Indels	2%	-	Likely Pathogenic	NA	NA	NA	NA	-	-	-	VUS
	MSH3	5	80675096	c.1148delA	p.Lys383ArgfsTer32	Frameshift Indels	25%	-	Likely Pathogenic	NA	Likely Pathogenic	Likely Pathogenic	Likely Pathogenic - 2	Yes	None	-	Likely Pathogenic
	APC	5	112838594	c.3000C>A	p.Tyr1000Ter	Stop gained	2%	-	Likely Pathogenic	NA	NA	NA	NA			Yes	Likely Pathogenic
	APC	5	112839979	c.4393_4394delAG	p.Ser1465TrpfsTer3	Frameshift Indels	2%	-	Pathogenic	NA	Likely Pathogenic	Pathogenic	Pathogenic - 1; Likely Pathogenic - 1	Yes	None (0,00)	Yes	Likely Pathogenic
	BRCA2	13	32339700	c.5351delA	p.Asn1784ThrfsTer7	Frameshift Indels	2%	-	Pathogenic	NA	NA	Likely Pathogenic	Pathogenic - 1	-	-	Yes	Pathogenic
34	TP53	17	7675161	c.451C>T	p.Pro151Ser	Missense	2%	-	Pathogenic	Likely Pathogenic	Likely Pathogenic	Likely Pathogenic	Likely Pathogenic - 3			-	Likely Pathogenic
35	ARID1A	1	26771197	c.3281delA	p.Lys1094SerfsTer67	Frameshift Indels	2%	-	Likely Pathogenic	NA	NA	NA	NA	Yes	None (0,00)	-	VUS
	MSH2	2	47412449	c.687delA	p.Ala230LeufsTer16	Frameshift Indels	2%	-	Pathogenic	NA	NA	Likely Pathogenic	Likely Pathogenic - 1	Yes	None (0,00)	Yes	Likely Pathogenic
	MSH3	5	80675096	c.1148delA	p.Lys383ArgfsTer32	Frameshift Indels	25%	-	Likely Pathogenic	NA	Likely Pathogenic	Likely Pathogenic	Likely Pathogenic - 2	Yes	None	-	Likely Pathogenic
	APC	5	112839978	c.4393_4394dupAG	p.Ser1465ArgfsTer9	Frameshift Indels	2%	-	Likely Pathogenic	NA	Likely Pathogenic	NA	Likely Pathogenic - 1	Yes	None (0,00)	Yes	Likely Pathogenic
	APC	5	112840367	c.4778delA	p.Lys1593SerfsTer57	Frameshift Indels	2%	-	Likely Pathogenic	NA	NA	NA	NA	-	-	-	VUS
	CTNNA1	5	138924599	c.1636C>T	p.Arg546Ter	Stop gained	2%	-	Likely Pathogenic	NA	NA	NA	NA			-	VUS
	CTHRC1	8	103378048	c.394C>T	p.Arg132Cys	Missense	2%	Very Rare Global (0,08%)	VUS	NA	NA	NA	NA			-	VUS
	ATM	11	108227807	c.104G>A	p.Arg35Gln	Missense	2%	-	VUS	NA	NA	Uncertain Significance	Uncertain Significance - 1			-	VUS
	CDH1	16	68801883	c.377C>T	p.Pro126Leu	Missense	2%	Very Rare Global (0,004%)	VUS	NA	NA	Uncertain Significance	Uncertain Significance - 1	Yes	None (0,00)	-	VUS
	STK11	19	1220399	c.491T>C	p.Leu164Pro	Missense	2%	-	VUS	NA	NA	NA	NA			-	VUS
36	MSR1	8	16155129	c.833C>T	p.Pro278Leu	Missense	2%	-	VUS	NA	NA	NA	NA	Yes	None (0,53)	-	VUS
	TP53	17	7674221	c.742C>T	p.Arg248Trp	Missense	4%	Very Rare Global (0,002%)	Pathogenic	Likely Pathogenic	Likely Pathogenic	Likely Pathogenic	Likely Pathogenic - 3			-	Likely Pathogenic
37	MSH3	5	80675096	c.1148delA	p.Lys383ArgfsTer32	Frameshift Indels	25%	-	Likely Pathogenic	NA	Likely Pathogenic	Likely Pathogenic	Likely Pathogenic - 2	Yes	None	-	Likely Pathogenic
38	MSH3	5	80675096	c.1148delA	p.Lys383ArgfsTer32	Frameshift Indels	25%	-	Likely Pathogenic	NA	Likely Pathogenic	Likely Pathogenic	Likely Pathogenic - 2	Yes	None	-	Likely Pathogenic
	CTNNA1	5	138824800	c.858+1G>A	-	Splice donor	2%	-	Likely Pathogenic	NA	NA	NA	NA	Yes	Pathogenic (0,99)	-	VUS
	TP53	17	7674230	c.733G>A	p.Gly245Ser	Missense	2%	Very Rare Global (0,002%)	Pathogenic	Likely Pathogenic	Likely Pathogenic	Likely Pathogenic	Likely Pathogenic - 3			-	Likely Pathogenic
39	APC	5	112839702	c.4108A>T	p.Lys1370Ter	Stop gained	2%	-	Likely Pathogenic	NA	NA	NA	NA	Yes	Pathogenic (0,93)	Yes	Likely Pathogenic
40	MSH6	2	47800177	c.2194C>T	p.Arg732Ter	Stop gained	2%	Very Rare Global (0,002%)	Pathogenic	NA	NA	Pathogenic	Pathogenic - 1	Yes	None (0,00)	-	Pathogenic
	MSH6	2	47803500	c.3261dupC	p.Phe1088LeufsTer5	Frameshift Indels	4%	-	Pathogenic	NA	NA	Likely Pathogenic	Likely Pathogenic - 1	Yes	None (0,00)	Yes	Pathogenic
	FAT4	4	125321047	c.4636G>A	p.Val1546Ile	Missense	2%	Very Rare Global (0,002%)	VUS	NA	NA	NA	NA	Yes	Pathogenic (0,78)	-	VUS
	MSH3	5	80675096	c.1148delA	p.Lys383ArgfsTer32	Frameshift Indels	25%	-	Likely Pathogenic	NA	Likely Pathogenic	Likely Pathogenic	Likely Pathogenic - 2	Yes	None	-	Likely Pathogenic
	ATM	11	108229320	c.329_330delGA	p.Arg110LysfsTer4	Frameshift Indels, Splice region	2%	-	Likely Pathogenic	NA	NA	NA	NA	-	-	-	VUS
	BRCA2	13	32394815	c.9383G>A	p.Arg3128Gln	Missense	2%	Very Rare Global (0,002%)	VUS	NA	NA	Uncertain Significance	Uncertain Significance - 1			-	Likely Pathogenic
	TP53	17	7675157	c.455delC	p.Pro152ArgfsTer18	Frameshift Indels	2%	-	Likely Pathogenic	NA	NA	NA	NA	-	-	-	VUS
41	TP53	17	7673787	c.833C>T	p.Pro278Leu	Missense	2%	-	VUS	Likely Pathogenic	NA	Likely Pathogenic	Likely Pathogenic - 2			-	Likely Pathogenic
42	MCCC1	3	183037239	c.1573A>T	p.Thr525Ser	Missense	2%	-	VUS	NA	NA	NA	NA			-	VUS
	FAT4	4	125448714	c.7698A>C	p.Lys2566Asn	Missense	2%	-	VUS	NA	NA	NA	NA			-	VUS
	MSH3	5	80675096	c.1148delA	p.Lys383ArgfsTer32	Frameshift Indels	25%	-	Likely Pathogenic	NA	Likely Pathogenic	Likely Pathogenic	Likely Pathogenic - 2	Yes	None	-	Likely Pathogenic
	ITIH2	10	7705109	c.86T>G	p.Phe29Cys	Missense, Splice region	2%	-	VUS	NA	NA	NA	NA			-	VUS
	ATM	11	108247071	c.1009C>T	p.Arg337Cys	Missense	2%	Very Rare Global (0,002%)	VUS	Likely Pathogenic	NA	Uncertain Significance	Likely Pathogenic - 1; Uncertain Significance - 1	Yes	Pathogenic (0,95)	-	Conflicting
	BRCA2	13	32337305	c.2957dupA	p.Asn986LysfsTer2	Frameshift Indels	2%	-	Likely Pathogenic	NA	NA	Pathogenic	Pathogenic - 1	-	-	Yes	Pathogenic
	BRCA2	13	32356499	c.7507G>A	p.Val2503Ile	Missense	2%	Very Rare Global (0,002%)	VUS	NA	NA	Uncertain Significance	Uncertain Significance - 1			-	VUS
43	MSH6	2	47803501	c.3261delC	p.Phe1088SerfsTer2	Frameshift Indels	6%	-	Pathogenic	NA	NA	Likely Pathogenic	Likely Pathogenic - 1	Yes	None (0,00)	Yes	Pathogenic
	MSH3	5	80675096	c.1148delA	p.Lys383ArgfsTer32	Frameshift Indels	25%	-	Likely Pathogenic	NA	Likely Pathogenic	Likely Pathogenic	Likely Pathogenic - 2	Yes	None	-	Likely Pathogenic
44	MSH6	2	47803501	c.3261delC	p.Phe1088SerfsTer2	Frameshift Indels	6%	-	Pathogenic	NA	NA	Likely Pathogenic	Likely Pathogenic - 1	Yes	None (0,00)	Yes	Pathogenic
	MSH3	5	80675096	c.1148delA	p.Lys383ArgfsTer32	Frameshift Indels	25%	-	Likely Pathogenic	NA	Likely Pathogenic	Likely Pathogenic	Likely Pathogenic - 2	Yes	None	-	Likely Pathogenic
45	TP53	17	7674220	c.743G>A	p.Arg248Gln	Missense	2%	Very Rare Global (0,01%)	Pathogenic	Likely Pathogenic	Likely Pathogenic	Likely Pathogenic	Likely Pathogenic – 3			-	Likely Pathogenic
46	PTEN	10	87952122	c.499dupA	p.Thr167AsnfsTer13	Frameshift Indels	2%	-	Likely Pathogenic	NA	NA	NA	NA	-	-	Yes	VUS

47-52 Somatic variants were not found

Description of germline variants, including both frequency in the cohort (freq cohort: % of mutated families), ExAC and 1000genomes and Clinical Significance classification from Illumina's Variant Interpreter, UniProt, OMIM, ClinVar, Annotator57 (in the 3 databases) and Integrated classification. Eur: Europe; Tusc: Tuscany; VUS: Variant of Unknown Significance; NA: Not available; Very Rare: <1%; Rare: >1, <5; Not Rare: >5.