GENETICS OF FAMILIAL GASTRIC CANCER:

BEYOND THE E-CADHERIN (*CDH1*) LOCUS

by

Samantha Hansford

BSc, Memorial University of Newfoundland, 2011

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES

(Pathology and Laboratory Medicine)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

April, 2014

Abstract

Importance. Familial aggregation occurs in approximately 10% of gastric cancers, which are generally sub-classified histologically as intestinal-type and diffuse gastric cancers. Though the genetic basis of familial intestinal-type gastric cancers is not known, in <50% of families clinically defined as hereditary diffuse gastric cancer, germline mutations in the E-cadherin gene, *CDH1*, are detected. This has lead to management guidelines and prevention strategies for mutation carriers.

Objectives. To determine whether pathogenic germline mutations in genes alternative to *CDH1* can be found in hereditary gastric cancer families using a multiplex panel sequencing approach.

Design, Setting and Participants. One hundred fifteen probands from families who met the International Gastric Cancer Linkage Consortium clinical criteria for hereditary diffuse gastric cancer (n=106) or familial intestinal-type gastric cancer (n=9) were included. All diffuse gastric cancer probands tested negative for *CDH1*-mutations. Germline DNA was screened against a custom panel of 55 genes, including 14 prospective gastric cancer susceptibility genes, using a multiplexed amplicon-based next generation sequencing assay. Candidate mutations were validated via Sanger sequencing. Tumours from pathogenic mutation-positive probands were evaluated by immunohistochemistry.

Results. Of 115 probands, four clearly pathogenic truncating mutations were identified in unrelated families, including two different mutations in *CTNNA1* (alpha-catenin) and two different mutations in *BRCA2*. Previously described, functionally pathogenic missense mutations in *SDHB* (2 families) and *STK11* (2 families) were also seen. Additional truncating mutations of likely lower penetrance were identified in *ATM* (4 families), *MSR1* (2 families) and

PALB2 (1 family). Cancers from carriers of *CTNNA1* truncating variants had prominent loss of protein expression, further supporting their pathogenicity.

Conclusion and Relevance. Using a multi-gene panel, families with hereditary gastric cancer were found to carry pathogenic mutations in genes commonly associated with other cancer susceptibility syndromes. In addition, this data suggests that familial gastric cancers, specifically hereditary diffuse gastric cancer syndrome, may benefit from a genetic, rather than clinical, classification. The genetic basis of the remaining families is likely attributable to mutations in genes yet to be implicated in hereditary gastric cancer or, in the diffuse gastric cancer families, atypical aberrations in the non-coding regions of *CDH1*.

Preface

The results of this thesis project have been written into a manuscript and is being prepared for submission. Dr. Hector Li-Chang, Dr. Michelle Woo, Dr. Carla Oliviera, Dr. David Huntsman and S. Hansford equally contributed manuscript writing before submission.

S. Hansford acknowledges that valid and applicable ethics certification is in place and that all requested activities comply with certification. All participants in this research project consented to genomic analysis of germline and tumour DNA samples for the purpose of novel mutation discovery.

Collection of samples, all subsequent sequencing experiments and data analysis, with exception of secondary analysis on 25 samples, was completed by myself (indicated in **Chapter 3**, Section 3.4.5.2. Registered pathologists Dr. D. Huntsman, Dr. H. LiChang or Dr. D. Schaeffer completed evaluation and interpretation of immunohistochemistry.

A version of **Chapter 2** has been published. S. Hansford. & D. Huntsman. (2013). Beyond *CDH1* Mutations: Causes of hereditary diffuse gastric cancer. In G. Corso & F. Roviello (eds.), Spotlight on Familial and Hereditary Gastric Cancer (97-110). Dordrecht, Netherlands: Springer. Text and Figures have been approved for reprint for the purpose of thesis submission and availability through associated university. S. Hansford contributed to the entirety of the book chapter and designed all figures. Figure 2.2 was adapted and reproduced from The Journal of Pathology by the permission of The Pathological Society of Great Britain and Ireland: Schrader et al. 2011). Manuscript was reviewed and edited by D. Huntsman.

In **Chapter 6**, Sub-Section 6.2: No Stomach For Cancer Grant Proposal, writing was equally contributed by Dr. M. Woo, Dr. H. Li-Chang, Dr. D. Huntsman, Dr. C. Oliviera and Hansford, S. The proposed research project was submitted for competition on March 3rd, 2014.

Table of Contents

List of Tables

List of Figures

Abbreviations

GC- Gastric Cancer

IGC- Intestinal gastric Cancer

DGC- Diffuse Gastric Cancer

NGS- Next Generation Sequencing

FGC- Familial Gastric Cancer

LBC- Lobular Breast Cancer

FIGC- Familial Intestinal Gastric Cancer

HDGC- Hereditary Diffuse Gastric Cancer

IGCLC- International Gastric Cancer Linkage Consortium

CI- Confidence Interval

e-cadherin- Epithelial Cadherin

TSG- Tumour Suppressor Gene

LOH- Loss of Heterozygosity

GI- Gastrointestinal

PCR- Polymerase Chain Reaction

WGS- Whole Genome Sequencing

ES- Exome Sequencing

IHC- Immunohistochemistry

EAC- Esophageal Adenocarcinoma

BE- Barrett's Esophagous

Oligos- Oligonucleotides

ul- microlitres

SNP- Single Nucleotide Polymorphism

Indel- Insertion/Deletion Mutation

PROVEAN- Protein Variation Effect Analyzer

SIFT- Sorting Intolerant From Tolerant

PolyPhen- Polymorphism Phenotyping

FFPE- Formalin Fixed Paraffin Embedded

TSCA- TruSeq Custom Amplicon

VH1/2/3- Vinculin Domains 1/2/3

NTR- Netrin Domain

H- Helical Domain

OB1/2/3- Oligonucleotide Binding Domains 1/2/3

NLS- Nuclear Localisation Sequence

SRCR- Scavenger Receptor Cysteine-Rich Domain

TAN- Tel1/ATM N-Terminal Motif

FAT/FATC- FRAP/ATM/TRRAP Motif

PI3-K- phosphatidylinositol 3-kinase

VUS- Variant of Unknown Significance

SNV- Single Nucleotide Variant

CDH1- Cadherin 1

CTNNA1-Catennin, Alpha-1

BRCA2- Breast Cancer 2

PALB2- Partner and localizer of *BRCA2*

STK11- Serine/Threonine Kinase 11

SDHB- Succinate Dehydrogenase-B

MSR1- Macrophage Scavenger Receptor-1

CFTR- Cystic Fibrosis Transmembrane

MAP3K6- Mitogen Activated Protein Kinase Kinase Kinase 6

ATM- Ataxia Telangiectasia-Mutated

PJS- Peutz-Jeghers Syndrome

CS- Cowden Syndrome

CLS- Cowden-Like Syndrome

PTEN- Phosphotase Tensin

A-T- Ataxia Telangiecasia

GAPPs- Gastric Adenocarcinoma and Proximal Polyposis

Acknowledgements

To my family,

Thank you for your continuing support, encouragement and unconditional love.

To my grandmother, Laura Goss,

A two-time cancer survivor.

Your strength and courage will never cease to amaze me.

Above all, to my late aunt, Arleen A. Goss,

who succumb to diffuse gastric cancer December 10th, 2002.

Your perseverance and relentless battle inspired me to conduct this research.

Chapter 1: Introduction

1.1 Background – Gastric Cancer

Gastric cancer (GC) is the second-leading cause of cancer mortality world wide, accounting for over 700,000 deaths annually (Global Cancer Stats, 2011). This high mortality rate has remained relatively unchanged for the past thirty years, despite its decreasing worldwide incidence (Dicken et al. 2005; Howsen et al. 1986). Histopathologically, greater than 90% of gastric cancer diagnoses are classified as adenocarcinomas, with mucosa-associated lymphoid tissue (MALT) lymphomas and carcinoids comprising the other 10% of cases. The Lauren Classification has proven effective in diagnosing and defining gastric adenocarcinomas as intestinal-type and diffuse-type GC (Lauren P. 1965; Bosman et al. 2010), which differ in both tumour growth patterns and microscopic configuration.

Intestinal gastric cancers (IGC) take on a glandular, tubular-type microscopic appearance that mimics the appearance of colonic and intestinal mucosa (Dicken et al. 2005). Tumours grow in a unifocal, expanding fashion rather than infiltrative and are more often a secondary response to chronic inflammation, caused by such environmental factors as *heliobacter pylori* (*h. pylori*) infection and chronic gastritis (Dicken et al. 2005; Gore et al. 1997). These environmental risk factors, and others such as poor diet, infection and socioeconomic influence, play a strong role in the development and progression of GC and are more often associated with sporadic form of disease. As a result of the increased awareness and detection of environmental factors, the overall incidence of intestinal type GC has decreased in the Western world over the past 20 years (Borch et al 2000).

Diffuse-type gastric adenocarcinomas have multi-focal, signet-ring cell precursor lesions and are not limited to any particular part of gastric mucosa (Lauren, P. 1965) These tumours do not frequently exhibit glandular formation or intestinal metaplasia and are often associated with deep infiltration of the stomach wall (Dicken et al. 2005). Unlike intestinal adenocarcinomas of the stomach, diffuse-type GC (DGC) is not strongly associated with environmental influences. Earlier average age of onset with worse prognosis and often with relative incidence, strongly suggests a genetic influence and predisposition to disease. Nonspecific symptoms and poor detection, despite endoscopic screening, contribute to the delayed diagnosis in Western countries and overall high mortality rate.

Overall, greater than 65% of GC diagnosis will occur in late stage (T3/T4), contributing to a mere 20-30% 5-year survival rate (Dicken et al. 2005; Hundahl et al. 2000). A better understanding of risk factors and etiology of GC development are important for lowering the mortality rate and detection at its most early stages. As mentioned, the discovery and improved awareness of environmental risk factors have slowly contributed to a decrease in mortality of intestinal gastric cancers, which more often occur sporadically (without family history or inherited predisposition to disease). However, approximately 10% of gastric adenocarcinomas (more often diffuse-type) have a familial pattern of disease (La Vecchia et al. 1992; Fitzgerald 2010), occurring because of heritable genetic mutations that increase an individual's risk from birth. GC also occasionally presents in other familial syndromes with known genetic basis and management guidelines in place (Lynch et al. 1993; Carneiro, F. 2012). Identifying heritable genetic abnormalities and the risk they pose to carriers can drastically change the understanding of GC and allow for risk-stratification in more families. Improvements in genetic sequencing

technologies have changed the way we study heritable diseases. Using the latest next generation sequencing (NGS) techniques, we can effectively an efficiently identify new heritable genetic aberrations to uncover the relationship between genotypes and unexplained phenotypes, such as GC.

1.2 Familial Gastric Cancer

Familial (or hereditary) cancers present with aggregates of the same form of disease in a single family and can occur for a number of factors. For example, when family members have particular environmental cancer-risk factors in common (such as smoking or obesity), cancer may develop across multiple relatives. However, in most cases familial cancers are caused by genetic abnormalities that are inherited from one generation to the next, increasing a person's risk of developing a particular cancer from birth compared to the general population. Familial gastric cancer (FGC) is a rare, autosomal dominant cancer susceptibility syndrome. In 2012, guidelines for the clinical diagnosis and management for FGC were proposed (Kluijt et al. 2012). This group sought to establish a standard of care and initiate collaborative studies to improve the quality of care and mortality rates of families with a strong history of GC (Kluijt et al. 2012). A family will be referred to genetic services if they meet one of the following guidelines:

a) GC in one family member before age 40

b) GC in two 1st or 2nd degree relatives with one diagnosis before age 50

c) GC in three or more 1st or 2nd degree relatives independent of age

d) GC and breast cancer in one patient with one diagnosis before age 50

e) GC in one patient and breast cancer in 1st or 2nd degree relative with one diagnosis before age 50

At this point, further evaluation will determine the clinical diagnosis of the family, which is based on histological subtypes of GC within the family and the number of affected individuals. This evaluation determines the diagnostic criteria and subsequent guidelines for the family, including genetic testing for at risk individuals if a cancer susceptibility gene is known. Today, FGC can be categorized into: 1) aggregates of IGC or 2) aggregates of DGC and possibly lobular breast cancer (LBC). It is also coming to light that GC aggregates can also occur in other tumour syndromes with hereditary tendencies such as lynch syndrome, hereditary breast/ovarian cancer, familial adenomatous polyposis, *MUTYH*-associated polyposis (MAP), Peutz-Jeghers syndrome, Cowden-like syndrome, Juvenile polyposis or Li-fraumeni syndrome (Kluijt et al. 2012). Studying such familial syndromes from a clinical perspective has traditionally proven effective to identify a possible genetic cause and assess the risk of cancer for unaffected individuals. Managing FGC correctly requires a multidisciplinary team of professionals, genetic counseling, a detailed family history of disease that preferentially extends into three generations as well as histopathological confirmation of GC (Corso et al. 2013).

1.2.1 Familial Intestinal Gastric Cancer

Familial clustering of IGC is rare, but has been reported in several families worldwide. Guidelines have been outlined for the description of familial intestinal gastric cancer (FIGC) (Kluijt et al. 2012; Caldas, 1999):

a) IGC in 2 or more first/second degree relatives, with at least one diagnosis before age 50

b) IGC in 3 or more first/second degree relatives, independent of age

Families that meet these criteria are not candidates for genetic screening, which is currently only offered to families presenting with DGC. At this time, there are no known susceptibility genes to FIGC and, as a result, no clinically relevant risk-stratification for unaffected relatives. Specific guidelines for possible early detection exist in high-risk families that meet criteria, such as routine gastric surveillance via endoscopy for individuals at 5 years younger than youngest diagnosis in family, *h. pylori* testing and eradication, as well as paying particular attention to dietary habits (Palli et al. 2001; Nam et al. 2012). Aggregates of FIGC are believed to be a combination of both genetic and environmental factors. Identifying these genetic factors would help genetic counselors and clinicians develop the tools necessary to counsel such families and broaden our understanding of this rare hereditary syndrome.

1.2.2 Hereditary Diffuse Gastric Cancer

The first description of FGC was in 1964, when E. Jones described familial aggregates of early-onset GC (later described as DGC) in three Maori families from New Zealand (Jones, EG. 1964). In a family with 98 recorded members, 28 died of GC with no reports of other cancers in the family (Jones, EG. 1964). The average age of onset was 31 for females and 36 for males, significantly younger than the average age of onset for the general population. Though genetic sequencing was unavailable at this time, it was suspected that members of this family were at a higher risk for developing GC.

In 1998, as scientists were only beginning to scratch the surface of sequencing technologies, Guilford et al. embarked on a study to uncover the genetic events attributing to familial pattern of GC in Jones' Maori families (Guilford et al. 1998). Using linkage analysis and traditional

sequencing techniques, three different germline mutations were identified in the individual families in the gene *CDH1* (Guilford et al. 1998). This newfound molecular basis of disease helped coin the term Hereditary Diffuse Gastric Cancer (HDGC) [OMIM #137215], which is widely used today to describe families with clustering of DGC (Caldas et al. 1999). Recently after HDGC was recognized, the International Gastric Cancer Linkage Consortium (IGCLC) was formed. This group of leading the physicians and scientists has met regularly since the discovery of HDGC to present important findings on clinical management, genetics, biology, pathology and treatment of FGC and discuss prospective research in this area. In 1999, the IGCLC outlined the first criteria to recognize HDGC among families in the general population: 1) two or more documented cases of DGC in first/second degree relatives, with at least one diagnosed before the age of 50 or 2) three of more cases of DGC in first/second degree relatives, independent of age of onset (Caldas et al. 1999). Since the relationship between *CDH1* and HDGC was acknowledged, greater than 100 germline *CDH1* mutations have been reported across multiple ethnicities worldwide (Kaurah et al., manuscript in preparation). Segregation analyses among these families has provided further evidence that *CDH1* abnormalities play a direct role in the susceptibility to DGC. Recent penetrance analysis of these mutations outlines a cumulative risk of GC by age 80 of 70% (95% Confidence Interval (CI), 26-100%) for males and 57% (95% CI, 14%-99%) for females as well as a 56% (95% CI, 33%-82%) increased risk of LBC in female carriers (Kaurah et al., manuscript in preparation).

At this time, germline mutations within the *CDH1*-locus are the only known susceptibility to DGC and the only gene screened for HDGC families meeting clinical criteria. However, pathogenic CDH1 mutations are only identified in 46% of HDGC families. The remaining cases

are believed to have mono-allelic expression of *CDH1* through other genetic aberrations at the *CDH1* locus (i.e. epigenetic modification or non-coding variants with functional impact) or hold susceptibility mutations in genes yet to be identified. As mentioned, there are currently no genetic screens available for families meeting criteria for FIGC.

1.3 The Cellular Adhesion Molecule E-cadherin *(CDH1)*

Epithelial-cadherin (e-cadherin) is a member of the transmembrane glycoprotein classical cadherin molecules, which also include vascular endothelial cadherin (VE-cadherin) and neural cadherin (N-cadherin) (Ratheesh & Yap, 2012). Though first described in the chicken, its name was initially used in the 1984 by Yoshida-Noro et al. in mouse teratocarcinoma to describe a calcium-dependent cell-cell adhesion system (Yoshida-Noro et al. 1984). They highlighted that the existence of specific cellular adhesion molecules may be a key feature of cell communication between particular cell types in heterogenous cell populations (Yoshida-Noro et al. 1984). This study opened many doors to research involving cell-adhesion proteins and their encoding genes. *CDH1*, found on the long arm of chromosome 16 (16q22.1), is composed of 16 coding regions (exons) that transcribe into a 4.5-kb mRNA before encoding into 120kDal transmembrane, calcium-dependent protein known as e-cadherin [OMIM #19209]. The 882 amino acid encoded protein is comprised of several key domains: signal (Sig) and precursor peptides initially get cleaved to form a mature protein; a long extracellular domain with five e-cadherin repeats and calcium-binding sites is critical for communicating with cadherins on adjacent cells (Gall et al. 2013); a cytoplasmic domain at the C-terminus end of the protein holds two essential binding motifs: the juxtamembrance domain (JMD) for binding of p120-catenin that contributes to the overall adhesive strength and the beta-catenin binding domain. Once bound to e-cadherin
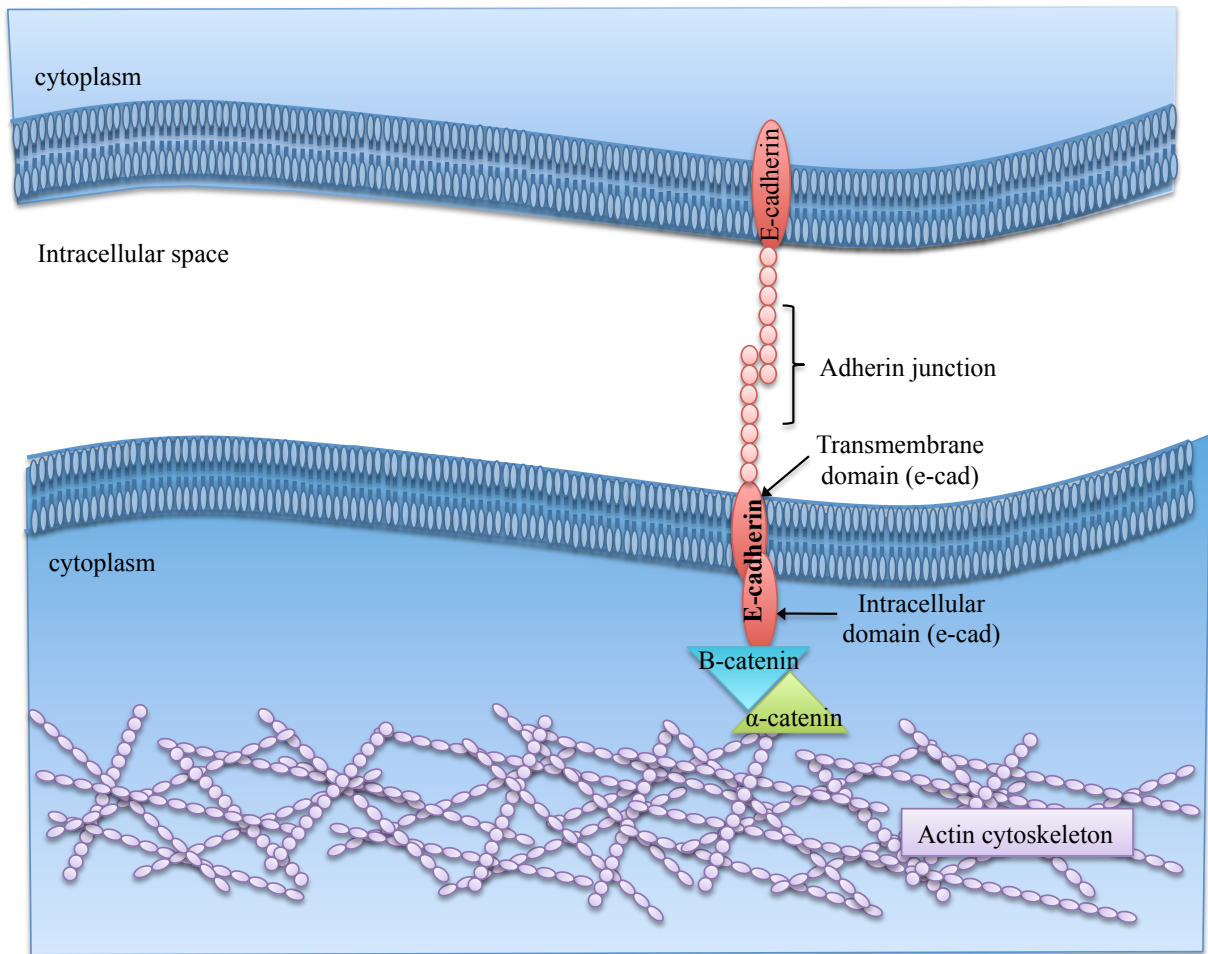
Figure 1.1 Cadherin-catenin protein complex. A schematic view of e-cadherin, beta-catenin and alpha-catenin interaction. Adjacent e-cadherin extracellular domains bind to one another via calcium-dependent dimerization. Intracellular e-cadherin domain binds to beta-catenin which complexes with alpha catenin. Alpha-catenin is able to secure and stabilize to actin cytoskeleton.

intracellular domain, beta-catenin complexes with the actin cytoskeleton via interaction with the protein alpha-catenin (Gall et al. 2013; Oliveira et al. 2013) (**Figure 1.1**). This transmembrane communication between cadherin and catenin proteins and subsequent assembly to intracellular actin-cytoskeleton is essential to maintain cellular integrity and communication between epithelial cells. Along with maintaining cellular adherin junctions, there is evidence that e-cadherin proteins may themselves send signals to regulate cell migration, proliferation, differentiation and apoptosis (Van Roy & Berx, 2008). Loss of e-cadherin expression disrupts these signaling cascades and eliminates adherin junctions, increasing the risk for epithelial malignancies (Gall et al. 2013) (Oliveira et al. 2013; Jeanes et al. 2008). When wild type e-cadherin is reintroduced in cancer cell lines, researchers witness decreased proliferation, motility and malignant tendencies (Conacci-Sorrell, 2002). For these reasons and others, *CDH1* is recognized as a well-established tumour suppressor gene (TSG).

1.3.1 E-cadherin (*CDH1*) and Cancer

The interaction between cytoplasmic domain of e-cadherin and catenins to bind to intracellular actin cytoskeletons is crucial for properly maintaining stable adherin-junctions and acts as a key anti-proliferation technique (Conacci-Sorrell et al. 2002). Genetic aberrations in the *CDH1* gene and downstream loss of e-cadherin have been described in many cancers including gastric, endometrial, oral, and LBC (Oliveira et al. 2013; Yi et al. 2011; Pannone et al. 2013; Masciari et al. 2007). Cancer development and progression through *CDH1* dysfunction is believed to occur through different abnormalities at the *CDH1*-locus, including frameshift insertions and deletions, nonsense

mutations, genomic rearrangements, transcriptional silencing by repressors that target *CDH1* precursor region and/or *CDH1*-promoter hyper-methylation (Oliveira et al. 2013; Oliveira, 2009; Jeanes et al. 2008). Consistent with tumour suppressor activity, a classical two-hit mechanism often occurs, disrupting the second *CDH1* allele for loss of protein expression and subsequent tumour development (Conacci-Sorrell et al. 2002). This loss of heterozygosity (LOH) event can occur via promoter methylation or through somatic second hit. It is correlated with loss of epithelial morphology in tumours as well as increased metastatic potential (Oliveira et al. 2013; Pinheiro, 2010; Corso et al. 2013).

Germline pathogenic *CDH1* mutations have been reported in approximately 45% of families meeting clinical criteria for HDGC and are currently the only known genetic susceptibility to this disease. Despite no detectable germline variants, mono-allelic *CDH1* expression is present in >70% of the remaining HDGC families, suggesting additional abnormalities affecting *CDH1* transcription-regulation are at play (Pinheiro, 2010). It is suspected that unspecific defects at the *CDH1*-locus (such as non-coding variants) may be involved in these *CDH1* mono-allelic cases (Pinheiro et al. 2010).

1.3.2 *CDH1* germline mutations and HDGC

As previously stated, heterozygous germline variants at the *CDH1*-locus remain the only underlying genetic susceptibility event for HDGC (Oliveira et al. 2013). New evidence shows that 155 germline *CDH1* mutations (126 pathogenic and 29 unclassified variants) have been described in 183 HDGC families worldwide, representing approximately 45% of those meeting IGCLC criteria (Kaurah et al., manuscript in preparation). Frequency of

CDH1 germline mutations varies significantly geographically, with increased occurrence

in low incidence regions such as North America and United Kingdom (Kaurah et al.

2007). High incidence countries, such as Japan, China and Korea, tend to have decreased

frequencies of *CDH1* aberrations to explain GC clustering (Carneiro et al. 2008). The

majority of these cases are thought to be strongly influenced by environmental risk

factors, such as diet and *H. pylori* infection, rather than heritable genetic abnormalities

(Carneiro et al. 2008).

Recent penetrance analysis of HDGC families and *CDH1* mutation occurrence have

indicated the relative risk of developing gastric cancer by age 80 in *CDH1* mutation

carriers for men (70%) and women (57%) as well as the risk of lobular breast cancer in

women (56%) compared to the general population (Kaurah et al., manuscript in

preparation). The variable penetrance is not well understood, but it is likely that both

environmental and lifestyle factors act as modifiers of risk in these families. Pathogenic

variants occur across all coding regions of *CDH1* and have not been found to be

restricted to any particular functional domain (Kaurah et al., manuscript preparation)

(**Appendix 1**). Coding insertions/deletions, large multi-exonic deletions, splice-site

junctions as well as 5' and 3' UTR variants have all been described in HDGC families,

although majority of mutations found have been protein truncating from small frameshift

variants (Oliveira et al. 2013; Pinheiro et al. 2010). Three positions (c.1137G, c.1792C,

c.1565) at the *CDH1* locus have been found to contain mutations across multiple

families, suggesting the possibility of hotspot regions (Kaurah et al., manuscript in

preparation) (**Appendix 1**). Further haplotype work is needed to confirm if these recurring

mutations are due to independent events or deviation from common ancestry among families.

CDH1 missense mutations, single nucleotide substitutions resulting in an amino acid change, are detected in 20% of HDGC cases and represent significant burden to genetic counselors (Guilford et al. 2010; Oliveira et al. 2009 (1); Pinheiro et al. 2010). It is difficult to predict the functional implications of such amino acid changes in germline carriers as they result in full-length e-cadherin molecules. Consequently, carrier testing and risk-stratification for these HDGC families is not offered. However, some reports suggest that *CDH1* missense mutations can disrupt e-cadherin protein integrity and function through premature degradation, structural destabilization or by inducing abnormal activation of downstream signaling cascades (Simoes-Correia et al. 2008; Figueiredo et al. 20130; Ferreira et al. 2012; Mateus et al. 2009). In addition to coding missense variant, preliminary work of *CDH1* missense mutations within intron 2 suggests that some non-coding variants may hold functional implications through disruption of transcriptional efficiency (Pinheiro H, Carvalho J & Oliveria C, 2012) (**Chapter 6**). It is plausible that some non-coding variants account for the observed phenotype of allele-specific down regulation found in up to 70% of *CDH1*- negative families. Further work is needed on these missense and non-coding mutations to fully understand the events leading to this allele-specific expression before carrier testing and subsequent risk-reduction strategies can be put in place (**Chapter 6**). If this holds true, and *CDH1* non-coding intronic variants are found to be pathogenic, it suggests that true HDGC should be a genetic, not clinically, defined disease, whereby *CDH1* or *CDH1*-like abnormalities account for its mutational profile (**Chapter 5**). Currently, risk stratification for unaffected

individuals is not offered to HDGC families without germline, coding *CDH1* variants. By genetically classifying FGC through identifying new susceptibility genes or triaging cases into other familial syndromes, a greater number of families may be offered management or prevention strategies.

Genetic Testing for HDGC families: Criteria, Counseling & Cancer Prevention

In 2004, Brooks-Wilson et al. expanded FGC criteria to include full *CDH1* screening should a family be found to carry a pathogenic mutation (Brooks-Wilson et al. 2004) and in 2010, the IGCLC redefined the criteria, which are now used today (Fitzgerald et al. 2010):

a) 2 or more documented cases of GC, one confirmed DGC before the age of 50

b) 3 confirmed cases of DGC in 1st or 2nd degree relative, independent of age of onset

c) DGC diagnosis before the age of 40

d) Personal or family history of DGC and lobular breast cancer, one diagnosed before age 50.

Families found to meet these clinical criteria for HDGC are recommended to undergo genetic screening for *CDH1* mutations. If a *CDH1*-mutation is found, unaffected relatives may be offered genetic counseling services and subsequent testing for that mutation for risk stratification. Healthy individuals found to carry a highly penetrant mutation in *CDH1* are offered to undergo a total prophylactic gastrectomy to prevent the development of GC (Caldas et al. 1999; Fitzgerald et al. 2010). Though a radical procedure with variable post-operative quality of life, it is currently the best cancer-

prevention option, as it is difficult to detect DGC at its earliest stages using endoscopic screening (Caldas et al. 1999; Fitzgerald et al. 2010; Huntsman et al. 2001). In the evaluation of post-operative prophylactic gastrectomy specimens of *CDH1* mutation carriers, cancerous-precursor lesions are found in the majority of cases (Huntsman et al. 2001; Carneiro et al. 2004; Charlton et al. 2004; Norton et al. 2007; Rogers et al. 2008; Barber et al. 2008b; Hebbard et al. 2009; Hackenson et al. 2010; Pandalai et al. 2011; Kluijt et al. 2012). Genetic counseling for mutation carriers must be extensive and tailored to the individual's age, sex and related nutritional issues (Kluijt et al. 2012). The recommended age for prophylactic gastrectomy surgery in *CDH1*-carriers is not fixed and should be guided by ages of onset in the families (Kluijt et al. 2012; Caldas et al. 1999).

CDH1 germline mutations are currently the only known risk factor for increased susceptibility to DGC and there are no further genetic screens for HDGC families without such mutations or families meeting criteria for FIGC. However, GC is been shown in families who meet clinical criteria for other familial syndromes, such as hereditary non-polyposis colorectal carcinoma (HNPCC), Li-Fraumeni, Peutz-Jeghers, Cowden Syndrome and familial breast/ovarian cancers (Lynch et al. 1993; Shinmura et al. 2005; Takahashi et al. 2004; Kluijt et al. 2012; Varley et al. 1995; Jakubowska et al. 2003). This suggests that susceptibility genes for these syndromes may be present in some FGC cases or that GC is overrepresented in some cancer susceptibility syndromes.

Unfortunately, the genetic causes of the remaining families that meet criteria for HDGC, and 100% of FIGC cases, are unknown and there is uncertainty in the management and

risk-stratification of these families. Despite the belief that other genes are at play in FGC susceptibility, these have yet to be directly recognized in FGC. Identifying additional genes that contribute to FGC may increase our understanding of both hereditary and sporadic forms of GC while providing a greater number of families with risk-stratification procedures and possibly cancer-prevention opportunities. With the advances in genetic sequencing technologies, identifying these genetic susceptibilities to hereditary syndromes becomes more efficient when compared to first generation techniques.

1.4 Hypothesis and Summary of Objectives

1.4.1 Hypothesis

A subset of families that meet clinically defined criteria for FGC (HDGC of FIGC) but have previously tested negative for pathogenic *CDH1*-abnormalities, will have identifiable pathogenic mutations in other genes with established or suggested association with upper GI cancers or cancer-susceptibility syndromes.

1.4.2 Rationale and Objectives

Though carrier testing, risk-stratification and preventative options exist for families meeting criteria for HDGC and who are found to carry pathogenic mutations in the *CDH1* gene, the majority of FGC cases (54% of HDGC and all of FIGC) remain with the burden of uncertainty as to their molecular basis of disease. In this study, a multiplexed amplicon-based next generation sequencing assay was used to test FGC cases for germline mutations in a custom panel of 55 cancer-related genes, including many previously associated with upper GI malignancies.

Uncovering such genetic variants is expected to provide insight into the molecular mechanisms of both hereditary and sporadic forms of FGC, as well as improve clinical management and genetic testing in affected families.

The specific **aims & objectives** of this project have been two fold.

Aim 1: Identify new GC susceptibility genes by collecting families with unexplained hereditary patterns of FGC (i.e. no identifiable *CDH1* mutations) and screening them against a custom panel of genes previously related to upper GI cancers or cancer susceptibility syndromes.

Aim 2: Demonstrate the efficacy and efficiency of multiplexed, panel-based sequencing for the identification of hereditary syndrome susceptibilities, specifically FGC.

The targeted next generation sequencing assay chosen for screening germline DNA from FGC families will provide a comprehensive and efficient evaluation of genes previously associated with GC or GC-susceptibility syndromes. By initially assessing a selected panel of genes, variants can be immediately related back to disease and families without identifiable mutations can be triaged for downstream whole genome or whole exome sequencing. **Chapter 2** evaluates the growth of sequencing technologies, compare the usefulness of traditional methods and introduce the latest next generation, targeted multiplexing techniques. The overall improvement of speed, efficiency and simplicity of these new techniques is what separates them from their former counterparts. **Chapter 3** assesses the selection of study design, materials collected and overall methodologies of the conducted research. Of importance is the description of the streamlined, efficient use of Illumina's TruSeq Custom Amplicon assay to multiplex multiple samples at once and sequence targeted areas of interest. **Chapters 4 & 5** will outline the overall results and discuss the impact they have on the field of hereditary cancer.

The intention of **Chapter 6** is to highlight the limitations of this project, focusing on the difficulty of collecting additional samples from affected family members in order to prove pathogenicity of the identified mutations. As well, future directions will be addressed in this **Chapter 6**, emphasizing the newly proposed project to the No Stomach For Cancer organization as well as functional evaluation of mutations within new GC susceptibility genes. Conclusions will be drawn about the overall project, highlighting the efficacy of study design and necessity of panel-based sequencing for prospective FGC cases.

Chapter 2: Next Generation Sequencing: Advances in technology and techniques

If one believes the fundamental pursuit of cancer genetics is to determine the genotypes with respect to unexplained, clinically relevant phenotypes, NGS will have a profound impact on our ability to use genetics to prevent cancer. In the past decade scientists have made remarkable breakthroughs in many aspects of research, but one of the most groundbreaking advances has been genome sequencing and uncovering the relationship between genotypes and unexplained phenotypes. It is with these cutting edge developments, such as the ability to test for the genetic foundations of disease, that science becomes directly translated into patient care.

FGC accounts for a small percentage (1–3 %) of GC cases in North America, with HDGC being the most common clinically defined variant (Kaurah et al. 2007). Its high mortality rate and strong autosomal dominant pattern have made HDGC a focus for hereditary cancer research. Upon the discovery of this hereditary pattern, sequencing technologies were only beginning to scratch the surface of hereditary diseases. In 1998 Parry Guildford used a classic combination of linkage and candidate gene analysis and discovered that germline *CDH1* variants were the cause of HDGC in three Māori families, one of which had an extensive multi-generational pedigree (Guilford et al. 1998). It is now fully supported that *CDH1* mutation carriers have an increased cumulative lifetime risk of developing advanced gastric cancer (Kaurah et al. 2007; Pharoah et al. 2001). Women are also at an increased risk for developing LBC by the age of 80 (Pharoah et al. 2001).

Though *CDH1* mutations have proven to be useful for HDGC management, mutations only account for <50% of cases (Oliveira et al. 2009 (1)) (**Figure 2.1**). The genetic causes of the remaining 50–60% of at-risk families is believed to be some combination of, alternative changes at the *CDH1*-locus or heritable variants in genes that have yet to be identified (Pinheiro et al. 2010) (**Figure 2.1**). Detection of these remaining genetic susceptibilities will greatly enhance the management of HDGC. At this time, there is little to offer clinically for disease prevention to families for which no *CDH1* mutation is found and they are left with the burden of uncertainty as to whether they carry a genetic susceptibility. Now that *CDH1* has been established as a predictive screen for HDGC, researchers are increasing their focus on *CDH1*-negative families and setting forth to uncover the genetic basis of these unexplained patterns of disease using the latest genome sequencing technologies. Today, researchers can harness the power, accuracy and overall efficiency of novel NGS techniques as they improve the overall speed of discovering hereditary susceptibility genes (Coonrod et al. 2012).

2.1 Previous Diagnostic Sequencing Techniques

The field of DNA sequencing has a rich and diverse history and many significant discoveries can be attributed to the first generation sequencing method: Sanger sequencing, which uses basic chemistry and PCR techniques to elongated DNA fragments (Sanger et al. 1977; Sanger, F. 1988). Though remains a reliable resource for genetic and clinical research, the Sanger method has several disadvantages when compared to the latest techniques, such as non-specific primer binding which can create a less accurate read-out and considerable cost when producing large amount of data
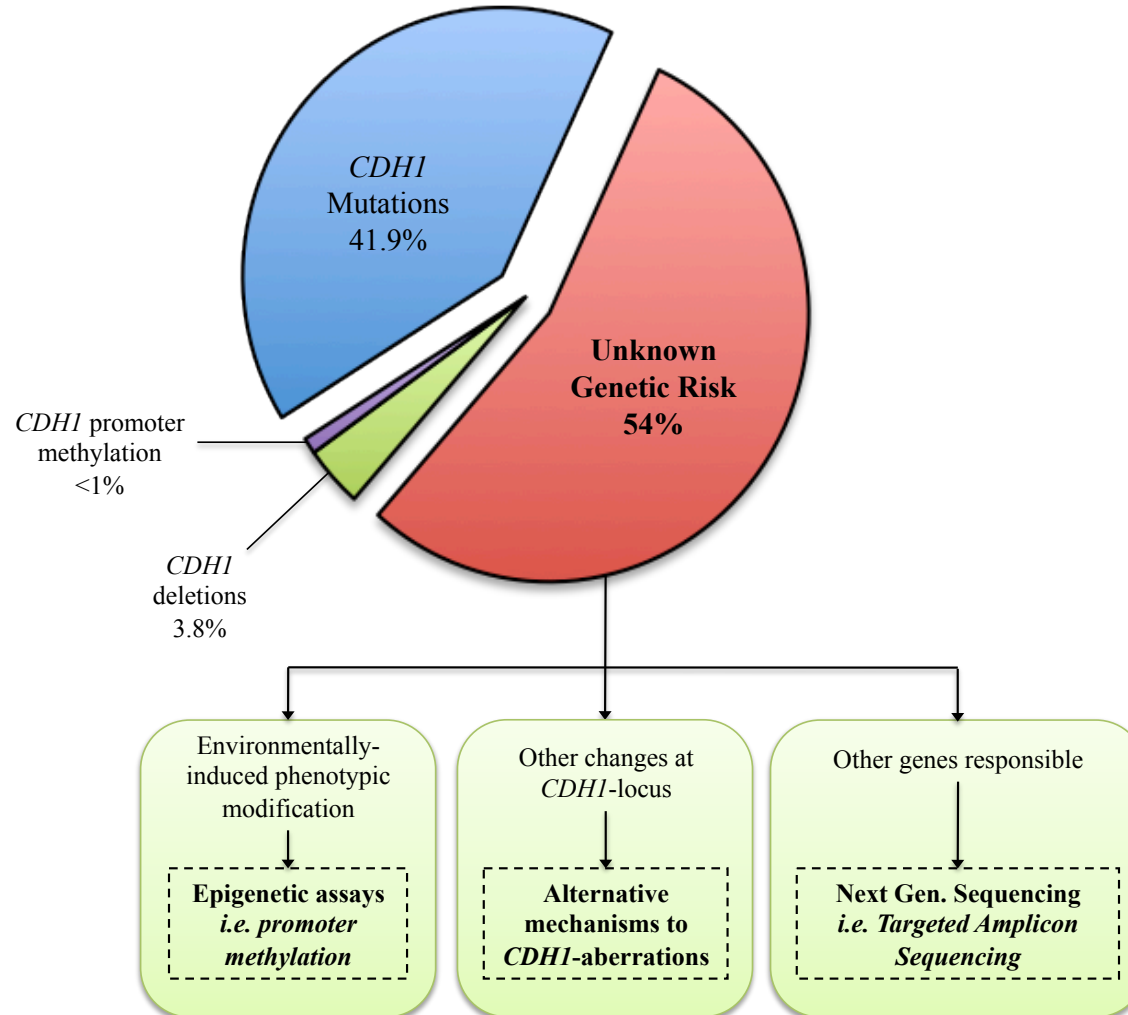
Figure 2.1. Molecular profile and understanding genetic contributions of hereditary diffuse gastric cancer. A distribution of CDH1 positive families (45.8%) by mutation type (truncating, indels, splice-sites and non-synonymous). There are several possible explanations for families that test negative for CDH1 such as environmentally induced phenotypic modifications, other changes at the CDH1 locus and novel genes responsible. The likely most effective way to identify these novel genes is TAS.

(Bybee et al. 2011). First generation techniques are not as useful today when interested in gene discovery and sequencing of a larger cohort of genomic information with respect to cost and efficiency.

In the investigative stage of diagnostic sequencing, scientists were subjected to tedious techniques that required a significant amount of time and funding (Coonrod et al. 2012). Linkage-analysis with subsequent candidate gene selection was the typical method for identifying novel, causal genes in inherited syndromes (**Figure 2.2**). For example, genetic linkage analysis was initially used to demonstrate significant relation to markers that flanked *CDH1* in the initially studied HDGC family (Guilford et al. 1998; Kaurah et al. 2007). Sanger sequencing was then used across the *CDH1* gene to identify a candidate truncating mutation believed to be causative for the familial pattern of disease within the family. When LOH was demonstrated in the tumour samples and segregated with additional family members, *CDH1* was confirmed as the primary candidate gene within the family (Guilford et al. 1998). The role *CDH1* plays in HDGC was later established when recurrent mutations were identified in rover 40 % of high-risk families (Oliveira et al. 2009 (1)). Though highly effective and proven to be a pivotal point in hereditary cancer research, the discovery of *CDH1* would have been much easier today. Despite the similarity to previous techniques, today's NGS methods can be completed on a much greater scale to maximize research efforts and funding.
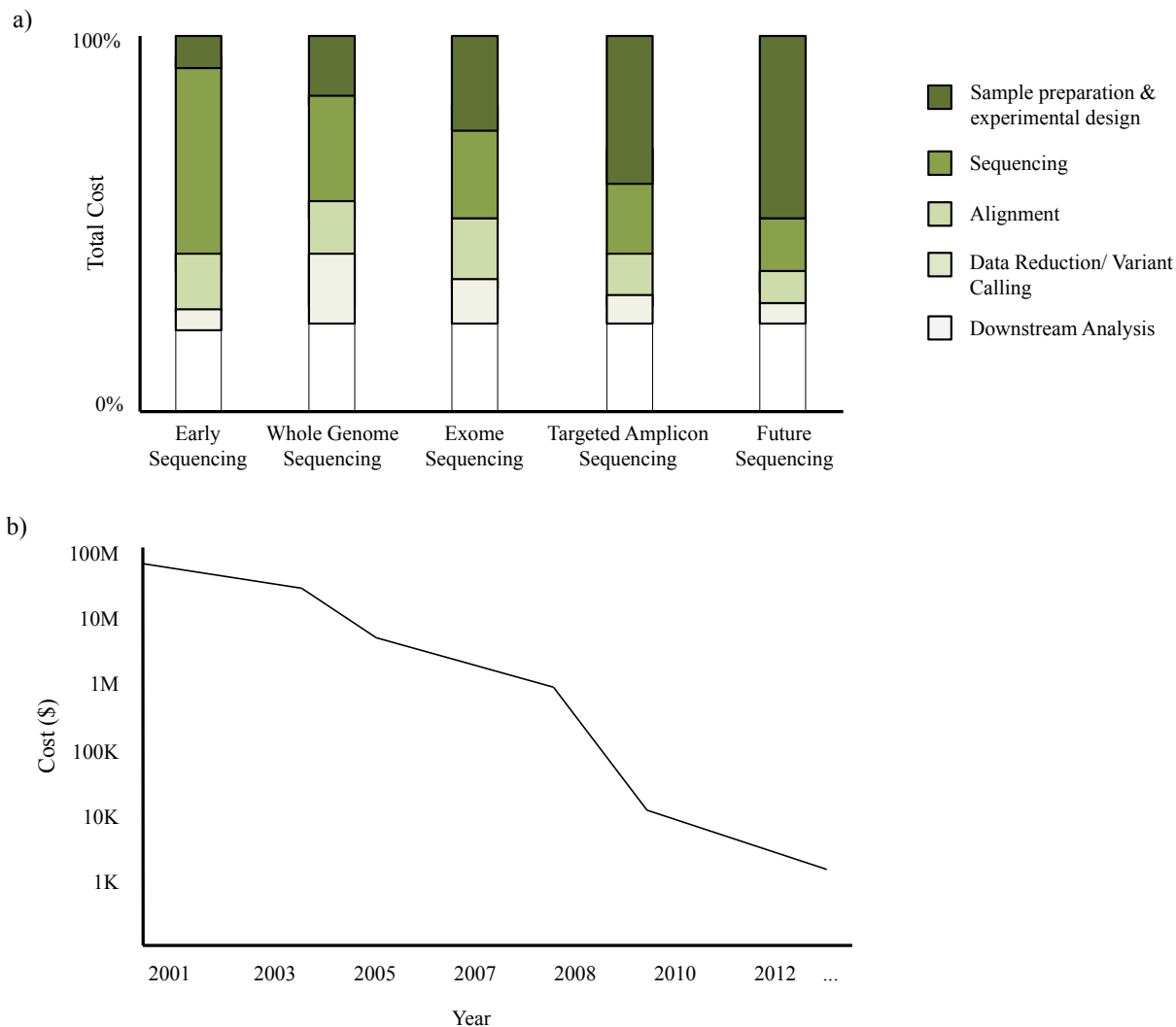
Figure 2.2 Contribution of factors to the overall cost of sequencing over time. (a) In early techniques the majority of cost was put toward tedious and inefficient sequencing protocols. Over time, cost of sequencing has significantly reduced because of advanced high-throughput technologies and funding has become more focused on project design and sample preparation. Costs of alignment, variant calling and downstream contributions have remained steady but (b) the overall cost of sequencing has significantly reduced (Modified from Sboner et al. 2011).

2.2 Advancements in Sequencing Technology: Next Generation Sequencing

With the introduction of relatively cheap, massively parallel DNA sequencing technologies, the overall cost of re-sequencing the human genome has fallen to more affordable rates (Bybee 2011; Hennekam and Biesecker 2012). For example, it is possible today for individual laboratories to re-sequence the human genome in a matter of weeks for tens of thousands of dollars, with the prediction of this cost dropping to 1,000 dollars in the coming years (Aparicio and Huntsman, 2010). Drastic improvements have been largely focused on revolutionizing traditional methods and overcoming rate limiting steps, such as the need of gel electrophoresis to separate DNA polymers (Aparicio and Huntsman, 2010). Today's NGS platforms use an array-based system to exceed these limitations, whereby DNA molecules are physically attached to solid surfaces or on an array of beads and the sequencing is determined *in situ* (Aparicio and Huntsman, 2010). As the DNA strand is elongated, the chemical or enzymatic addition of four colour labeled reversible terminators enables DNA sequencing by measuring which base has been added during the corresponding cycle (Aparicio and Huntsman, 2010). Overcoming the limitations of Sanger-based methods is a key step for improving the overall accuracy and affordability of these new sequencing platforms. Overall, there has been a significant shift in the distribution of cost and time with respect to sample preparation, sequencing methods, data alignment as well as downstream variant calling, which accounts for an overall increase in efficiency (**Figure 2.2**). For example, although bioinformatics challenges are still significant, basic nucleotide chemistry and enzyme engineering have individually improved on a small scale to contribute to the revolution of genome sequencing. Since the birth of sequencing techniques, a timeline of new methods has

paved the road to today's most current NGS models used by research institutes

worldwide. The most significant of these advances have occurred in the past decade and

can be attributed to the increasing desire and ability to uncover the molecular basis of

disease.

2.2.1 Advances in Sequencing Platforms

Sequencing platforms have greatly advanced since the use of first generation techniques.

Amplification methods are a key step in the sequencing process and over the past 10

years, scientists have developed advanced high-throughput systems that take a more

efficient approach in the generation of detectable sequencing features which

(Myllykangas et al. 2012) (**Figure 2.3**). Emulsion-based PCR is an amplification method

whereby the fragmented sequencing library is emulsified with a single enrichment bead

inside an oil-in-water reaction bubble (Myllykangas et al. 2012). A single fragment of

DNA is captured per bead by adaptor sequences, allowing parallel amplification of the

fragments to occur thousands of times within the oil-in- water emulsion mixture (**Figure

2.3-a**) (Myllykangas et al. 2012). The solution is then washed over a picoliter plate

containing wells large enough to hold a single bead and sequencing of the individual,

amplified library fragments can occur. This immobilization method is used by GS FLX

and SOLiD sequencing systems.

Bridge PCR is a second advanced sample preparation technique and is used by Solexa

systems. In this technique, DNA is fragmented and adaptor sequencings are attached to

both ends the immobilized onto a flow cell surface that has been coated with

a) Whole genome and exome sequencing

Exons

Sequence adaptors added

Whole Genome Sequencing

Exome Sequencing

Hybridization of probes to exons and removal of uncaptured fragments

Emulsion PCR **OR** Bridge-PCR

Ready for Sequencing

Sample Preparation

Break DNA into short fragments (~200-350bp)

Immobilization & Amplification

b) Targeted Amplicon Sequencing (Illumina)

Custom oligos target regions of interest *up to 1536 per sample*

Unique indexes target up to 96 samples to allow pooling into a single tube

Barcodes

Custom oligos

Barcoded libraries are mixed and uploaded onto single flow cell for simultaneous amplification and sequencing on the appropriate platform (i.e. Illumina's MiSeq)

Ready for Sequencing

Figure 2.3 Sample preparation work flow for NGS technologies. Downstream of sample fragmentation, methods differ in library preparation and PCR amplification techniques before sequencing. a) Whole genome sequencing and exome sequencing follow similar methodologies; however, ES requires an additional hybridization step to solely capture coding regions of the genome. Uncaptured regions are washed away before amplification via emulsion-based or bridge PCR. b) TAS requires more dedication to project design and library preparation as custom primers and sample-specific barcodes are hybridized to capture regions of interest. Barcodes allow for the pooling of up to 96 samples (each with up to 1536 targeted amplicons) in a single tube for extension, amplification and sequencing to all be run on a single machine at a fraction of the time to complete traditional sequencing methods..

corresponding adaptor sequencing (**Figure 2.3-b**). Template strands then bend and attach to neighboring primer to form a double stranded bridge (Myllykangas et al. 2012). This process continues and millions of dense double stranded clusters are formed in each channel of the flow cell. Images are captured on the flow cell of the cluster sequences and data is aligned with a reference genome for analysis. The benefit of this amplification method is its streamlined-approach and minimal hands on time. The sequencing instruments that utilize bridge-PCR (i.e. MiSeq, HiSeq, etc.) require no user intervention after cluster generation and data is analyzed directly on the sequencing instrument (Meldrum et al. 2011). These advances in sample preparation and amplification techniques allow current NGS platforms, such as Illumina Genome Analyzers, to sequence tens of millions of individual DNA templates in parallel, in comparison to hundreds of thousands of parallel reads in first-generation platforms, such as Roche 454 sequencing (Aparicio and Huntsman 2010). With these high-throughput approaches, researchers can significantly reduce the amount of hands on time and related sequencing bias, amount of sample necessary for evaluation and the overall cost of the sequencing process.

2.2.2 Multiplexing

Along with advances in sample preparation and amplification methods, multiplexing has revolutionized the sequencing of targeted regions and the future of hereditary susceptibility gene identification (Bybee et al. 2011). It allows for the pooling of multiple samples into a single sequencing reaction, further cutting costs and making data

assessment among affected individuals effortless. This subsequently benefits applications

such as targeted enrichment and is becoming a valuable source for research institutes.

Before amplification, a unique sample-specific barcode or index sequence is added to

specific regions of interest, allowing samples to be pooled into a single tube (**Figure 2.3-**

b) (Smith et al. 2010; Bybee et al. 2011). When sequenced, the barcode yields a unique

four base identifier at the beginning of each read that can later be used to separate reads

from the combined sample-pool (Pomraning et al. 2012). For example, the Fluidigm

Access Array is a high- throughput sample preparation system designed to work with

NGS platforms. It has been proven most beneficial for projects wishing to simultaneously

target a small number of regions (such as all exons of a few genes) for up to 48 samples

simultaneously. Similar to massive multiplex Fluidigm system, Agilent has created a new

target enrichment library preparation technique called Haloplex that is compatible with

all major desktop sequencers. After DNA digestion, custom oligonucleotide biotinylated

probes, specific to the targeted regions, are hybridized to each targeted DNA fragment

and the fragments form a circular DNA molecule (Agilent, 2012). Up to 96 sample-

specific barcodes are also incorporated into this hybridization step. Purification, through

bead binding of biotinylated probes, and ligation of targeted regions is then performed,

which ensures amplification of only circular DNA fragments (i.e. regions of interest)

(Agilent, 2012). PCR amplification of targeted areas for all samples is then performed in

parallel and samples are ready for sequencing. This technique removes the need for

significant library preparation, reducing the total cost and hands on time without the need

for robotic automation (Agilent, 2012). These massive multiplex sample preparation

technologies are excellent resources for re-sequencing preprioritized regions of the

genome, but perhaps the most impressive protocol most recently is Illumina's TSCA

technology, where up to 1,536 regions can be targeted for up to 96 samples prior to

amplification, sequencing and data analysis simultaneously on a single platform (i.e.

MiSeq) (**Figure 2.3-b**). Recent data has demonstrated the confidence of using

multiplexed sample preparation in a diagnostic setting through the detection of previously

identified single nucleotide variants (SNV), translocations, insertions and deletions

(Duncavage et al. 2012; Meldrum et al. 2011). When paired with amplicon based

sequencing, this method is proving to be the most efficient way to classify mutations that

attribute to autosomal dominant patterns of disease.

2.2.3 Whole Genome Sequencing

Whole genome sequencing (WGS) is a method that determines the complete genomic

makeup of an organism's genome, including both coding and non-coding regions and it

provides the most comprehensive collection of an organism's genetic variation (Ng and

Kirkness, 2010). As mentioned, the capability to produce massive amounts of data in

parallel for a fraction of the cost of first generation techniques will revolutionize many

aspects of medicine, including our increasing understanding of hereditary diseases. It is

also believed that the cost of re-sequencing will substantially decrease to roughly 1,000

dollars as companies strive to improve performance (Aparicio and Huntsman 2010)

(**Figure 2.2**). WGS is the best sequencing technique for identifying genomic

rearrangements and it is the only sequencing platform capable of picking up

chromosomal abnormalities. That said, the immense amount of data produced during

WGS proves difficult to reassemble and the sequencing of non-coding regions (introns) is unnecessary for certain applications. WGS is also highly subjected to sequencing bias, which must be taken into consideration when analyzing data. Researchers interested in the genetics of hereditary cancers are more concerned of regions that code into proteins (exons) as previous publications have identified mutations within these as targeted regions (Ng and Kirkness, 2010; Calva-Cerqueira et al. 2010; Kaurah et al. 2007; Nozawa et al. 1998). Advancements in NGS methods have been tailored to discovering highly penetrant mutations that attribute to hereditary cancers at a faster, more efficient rate (**Figure 2.4**).

2.2.4 Exome Sequencing

Exome sequencing (ES) is the sequencing of all coding regions (exons) of the genome that translate into protein. It is a more efficient strategy than WGS for uncovering mutations that attribute to rare mendelian disorders, such as hereditary cancers, for several reasons: 1) the majority of hereditary disease with an autosomal dominant inheritance pattern are caused by mutations within these exons, or coding regions, of the genome; 2) many non- synonymous substitutions are predicted to have high functional impact; and 3) the cost in comparison to WGS is substantially less yet provides sufficient data for conclusive reasoning behind most familial trends of disease (Ng and Kirkness, 2010; Ku et al. 2011). In both approaches, a candidate list of variants is created based on prior biological knowledge of the gene and predicted functional impact (**Figure 2.4**). Top candidate variants are validated by Sanger sequencing and segregated amongst additional
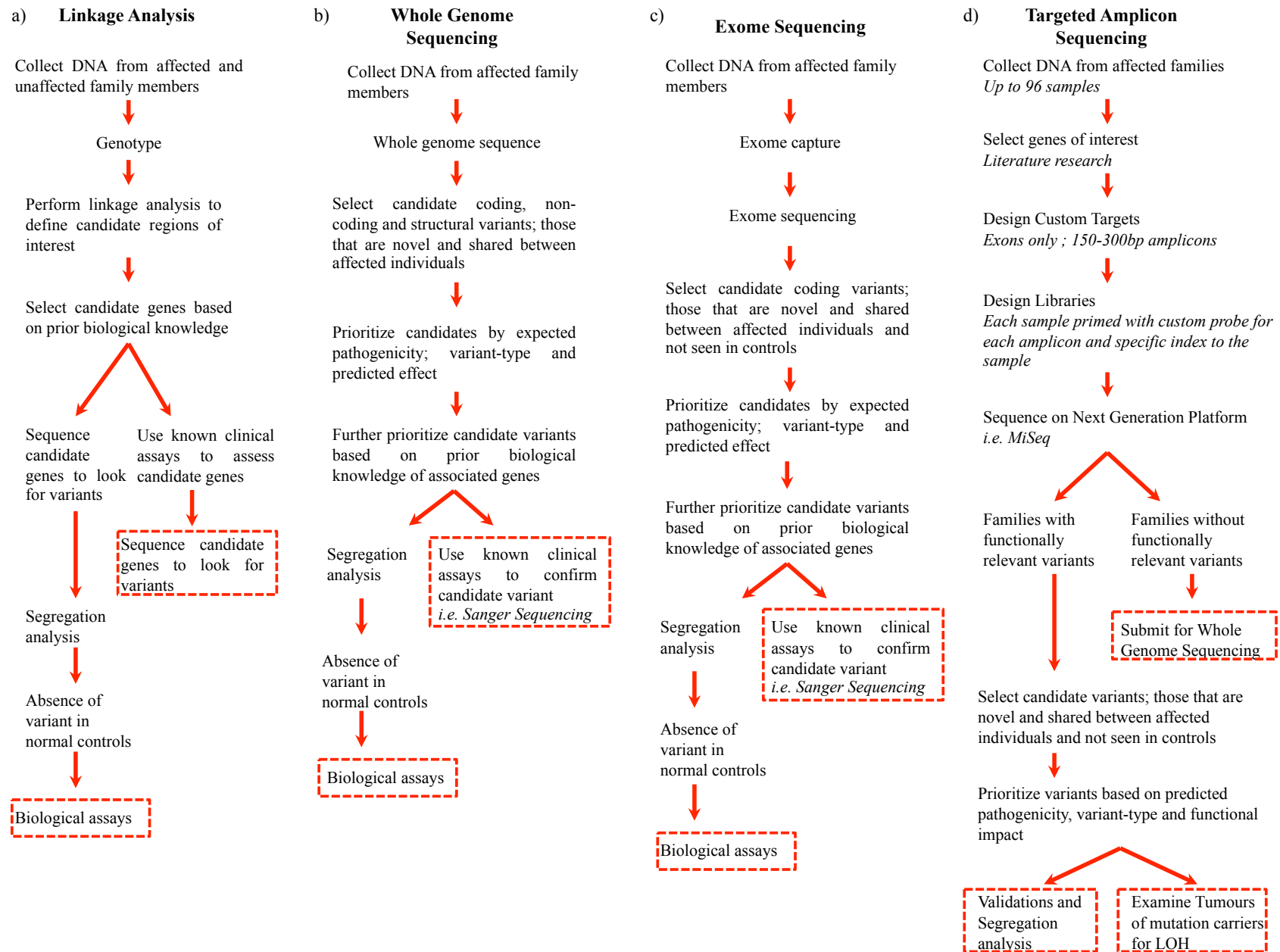
Figure 2.4 Flowchart for the identification of candidate genetic variants using different molecular genetics techniques

family members to solidify pathogenicity of the candidate. ES of families with similar pattern of disease is an effective tool for identifying rare, novel variants that account for hereditary patterns of disease (Wang et al. 2010; Ku et al. 2011). However, these sequencing methods become less efficient when causative variants are believed to be recurrent genes across multiple families. The latest high-throughput methods, capable of screening targeted regions across multiple samples in a single run, are likely more useful for identifying recurrent mutations that account for disease development in numerous families.

2.2.5 Targeted Amplicon-Based Sequencing

Sequencing is becoming increasingly useful in understanding the molecular basis of human health and disease. Targeted amplicon sequencing (TAS) is a new application within the genome sequencing community used to investigate specific genomic regions across multiple samples at a fraction of the cost (Bybee et al. 2011). Traditional Sanger sequencing methods do not compare to this new technology with respect to number of regions and samples sequenced at a given time. It enables the identification and quantification of known and novel sequencing variations that attribute to disease susceptibility within targeted regions (Bybee et al. 2011). TAS also offers the potential to amplify desired gene regions, focusing on short-reads (50–400 base pairs), followed by the use of high-throughput NGS platforms, which can be of great benefit when sequencing DNA from lower quality or preserved specimens that could otherwise not be done using previous methods (Bybee et al. 2011).

As previously mentioned, ES has proven effective in identifying rare, novel variants that attribute to hereditary patterns of disease (Schrader et al. 2011; Wang et al. 2010). However, variants within a single gene across a number of families can account for a high percentage of hereditary cases, as seen with *CDH1* (Kaurah et al. 2007; Schrader et al. 2011). In recent years, researchers have identified a number of genes predicted to cause cancers of the upper GI tract in families with strong inheritance patterns. Some of these studies indicate decreased expression using immunohistochemistry (IHC) (Rocco et al. 2003), copy number alterations (Calva-Cerqueira et al. 2010) and/or germline mutations (Calva-Cerqueira et al. 2010; Kaurah et al. 2007; Nozawa et al. 1998), but all results suggest candidate TSG that lead to the development of the UGI disorder. By expanding these findings and harnessing the power of TAS, it is possible to screen multiple families with unexplained hereditary patterns of DGC. This can be done using previously described Agilent and Illumina platforms that allow for the creation of a custom panel of targeted regions. This highly advanced TAS approach is significantly more cost-efficient, neither time-nor-labour intensive and can be applied to a wide variety of organisms and/or genes (Bybee et al. 2011).

2.3 Ethical Implications of Next Generation Sequencing

As the cost of genome sequencing continues to plummet, there is an increasing interest in personalized genomics. Though data collected from NGS results are extremely beneficial, the massive amount of data produced results in many ethical implications for the researchers and physicians involved. For example, WGS gathers information from the entire genome and while attempting to identify causative variants for a particular disease researchers may uncover additional medically significant information. Not only would this putative information be

important for the index patient, it may be important for relatives at risk of carrying the same variant. When this additional information surfaces, researchers are faced with the difficult decision of whether it is appropriate to disclose this information or not, how to disclose it and what exactly they should reveal to the patient and/or their relatives (Raffan and Semple 2011; Chan et al. 2012). Along with research genomics, the plummeting cost of genome sequencing is making the idea of personalized genomics more realistic. Critics of this growing trend fear the massive amount of information will be misinterpreted, as the majority of the data produced is not completely understood (Chan et al. 2012). The importance of these ethical implications will grow as sequencing technologies continues to advance to increasingly affordable rates.

2.4 Future Direction of Hereditary Diffuse Gastric Cancer: Utilizing Next Generation Sequencing

In utilizing the latest and developing methods of NGS, researchers can harness its power and uncover the genetic basis of HDGC beyond mutations at the *CDH1*-locus. All aspects of today's sequencing techniques will allow these new discoveries to be made at a faster and more affordable rate. WGS is the only NGS technology capable of picking up chromosomal rearrangements and abnormalities, which is highly useful for sequencing of tumour DNA or uncovering the genetic causes of non-hereditary diseases. However, WGS is subject to sequencing bias that may that must be accounted for and corrected during data analysis. The most likely and efficient method today for uncovering pathogneic mutations across responsible for rare familial syndromes is multiplexed TAS whereby targeted regions within the genome are screened across multiple samples (Bybee et al. 2011). Data from this technique is produced and

analyzed at astonishing rates, making comparisons between affected individuals much more rapid than previous methods. It also significantly reduces the cost associated with sequencing unnecessary regions when examining hereditary cases, as done previously with whole exome or whole genome sequencing. This is attributed to the ability of designing a custom panel of regions of interest. Screening *CDH1*- negative families for pre-prioritized candidate regions is an excellent way to quickly and affordably detect disease-causing variants and/or support that mutations causing some familial trends are novel, if no functionally relevant mutations are identified.

The discovery of *CDH1* and its significant contribution to HDGC marks a milestone for hereditary cancer research. However, since preventative measures have been established for *CDH1* mutation carriers, researchers are shifting their focus to families with unexplained inheritance patterns of this lethal disease. It is believed that similar genes harbour pathogenic mutations that account for these families. The latest NGS advances allow for multiple samples to be pooled in a single run and regions of interest selected for targeted sequencing, significantly cutting costs and time allotted for sequencing (Bybee et al. 2011; Duncavage et al. 2012). By identifying these unknown genetic contributions, similar screening programs to *CDH1* may be implemented for high-risk families and HDGC can be reduced to a more manageable disease.

Chapter 3: Study Design, Materials & Methods

3.1   CIHR Funded Grant: A collaboration

This project has been part of a larger collaboration funded in 2012 by the Canadian Institute of Health Research (**Acknowledgements**) entitled 'The Genetics of Hereditary Upper Gastrointestinal Cancers: Beyond *CDH1* Germline Mutations'. The aims of this project have been three-fold, with this particular project falling strictly under **Aim 1:** Identifying causative mutations that predispose to familial upper GI cancers (**Figure 3.1**). From our familial cancer registries and through collaborators, a large number of upper GI cancer families with highly penetrant unexplained cancer susceptibility have been collected. All cases are either consented or available for IRB approved consent for genomic analysis of germline and tumour DNA samples. Patients gave informed consent to take part in mutation identification through NGS. Patients were previously counselled by a certified genetic counsellor before taking part in this study for initial CDH1 screening.

Germline DNA was collected for the purposes of genomic screening against a multiplexed TAS method against known upper GI cancer susceptibility genes. If no mutations are found in these genes, it has been proposed that WGS of the two most distantly related affected individuals available from each kindred will be completed. Interrogation of the genomic data will be performed using informatics tools developed at the B.C. Cancer Agency and elsewhere. Putative mutations were to be validated by Sanger sequencing and then assessed in other families with similar phenotypes and added to the gene panel used in Aim 1. **Aim 2.** Somatic evidence to be obtained supporting pathogenicity of mutations found in Aim 1 and determining if the pathogenesis of hereditary cancers is distinct from their sporadic counterparts. **Aim 3.**

Determining how loss of CDH1 associated with HDGC leads to cancer progression by examining early lesions from prophylactic gastrectomy specimens. Though the focus of this funded project has been on FGC, the custom panel used for targeted sequencing was designed with all upper GI cancers in mind. A total of 304 samples were collected and sequenced against the custom designed amplicon panel. These samples were collected from collaborative agencies and included germline DNA from families with unexplained familial clustering of DGC, IGC, pancreatic cancer, esophageal adenocarcinoma (EAC), Barrett's esophagus (BE), as well as other upper GI cancers (duodenal and gall bladder) (**Table 3.1**).

Aim 2 is a future directive and will be discussed further in **Chapter 6 (Figure 3.1)**. **Aim 3** has been proposed to take place in the third year of funding, therefore has not yet commenced. GI Pathologist, Dr. H Li-Chang, has begun collecting prophylactic gastrectomy specimens and will lead the evaluation of cancerous precursor lesions beyond this project.The rationale for this funded project highlights the urgent need to uncover the genetic basis of hereditary upper GI cancers, to which all have mortality rates substantially greater than 50%. The highly penetrant nature of these conditions makes it clear that heritable mutations are at play and identifying a greater number of genes that play role in these disorders will provide a unique window to better understand both inherited and sporadic forms of disease.

3.1.1   Selection of assay: TruSeq Custom Amplicon Assay by Illumina

TAS was selected for this discovery-based project over extensive whole genome or exome sequencing for several reasons. Though effective, whole genome and exome sequencing are not efficient for sequencing a large number of germline cases simultaneously. Despite a significant
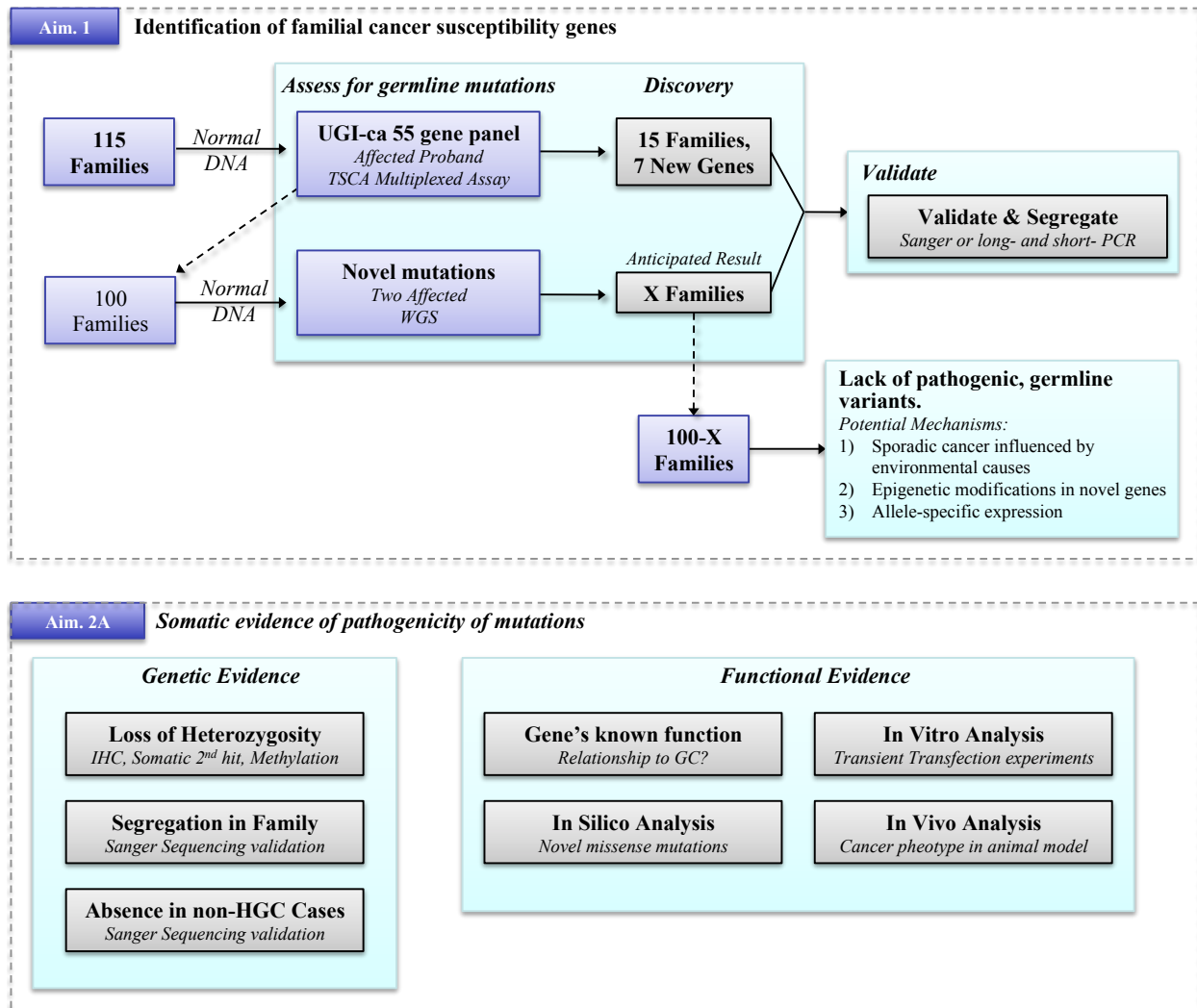
Figure 3.1 Workflow of experimental design proposal and sequence of events during CIHR funded project.

reduction in recent years, the cost of these techniques remain upwards to $7000 per sample at certain institutes with considerable pipelines that translate into long waiting periods before obtaining data (**Figure 3.2**). Raw data must then undergo analysis by a bioinformatician, which also takes a significant amount of time and effort.

A second reason for selecting targeted amplicon-based sequencing was to reduce the amount of data and non-specific variants (**Figure 3.2**). By sequencing targeted regions first, the likelihood of relating a candidate variant back to disease is higher than identifying such a variant in a gene not previously associated with the syndrome of interest. This method also greatly reduces the likelihood of identifying a pathogenic variant previously related to other hereditary syndromes with known risk, thus removing the burden placed on those with access to the data.

A third and significant reason was that it is capable to multiplex up to 96 samples together for simultaneous amplification, sequencing, and data analysis on the MiSeq platform. This significantly reduces both cost per sample and the time to retrieve data (**Figure 3.2**).

In choosing this TAS technique, there is significant heterogeneity among these multiplexed sequencing protocols with respect to cost, number of targeted regions available and number of samples one is capable of multiplexing per sequencing reaction while attaining desirable coverage. Illumina's TSCA Assay (Illumina, San Diego) was selected for its streamlined approach and price. This assay allows for an unprecedented level of multiplexing by integrating indices to support the sequencing of up to 96 samples in a single tube, generating data for up to 1536 custom amplicon targets. Using Illumina's Design Studio software (Illumina, San Diego), a total of 1531 amplicons, each composed of 250 base pairs, were designed to cover the exonic regions of each gene included on the custom panel (Table 3.1).

Whole Genome Sequencing

Single Sample

↓

$7,000 ($4,700)

↓

Illumina HiSeq
Platform

↓

4-6 weeks

↓

> 150Gb

↓

Bioinformatics Pipeline
Filtering and detection of genomic variants

↓

Selection of candidate variants
• functional relevance
• relation of gene to disease
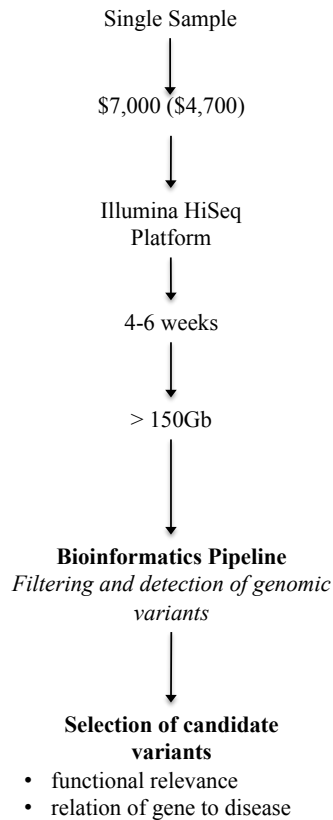
Single Gene Screen
(i.e. *CDH1* for HDGC families)

Single Sample

↓

$2500 ($400)

↓

Sanger Sequencing
or MLPA

↓

2-4 weeks

↓

<1Gb

↓

>40% Families with pathogenic variant   **60%** families with unknown genetic susceptibility

↓

Carrier testing for unaffected family members

↓

PTG for mutation carriers

Repeat with new gene of interest

Multiplex Custom-Panel Sequencing

Up to 96 Samples

↓

$4,000 ($300)

↓

Illumina MiSeq Platform

↓

<1 week

↓

10-15Gb
(dependent on size of panel)

↓

MiSeq Reporter Software
On-machine data analysis

↓

Selection of candidate variants
• Protein truncating variants
• Pathogenic SNPs

↓   ↓

No candidate variants
Submit for WGS   **Validate**
Sanger Sequencing

↓

Segregation & LOH analysis

Figure 3.2 A comparison of current approaches to identify genetic susceptibilities to disease, highlighting the efficiency of multiplex, custom-panel sequencing. Clinical price (manufacturer cost) of sequencing averaged across multiple institutions.

39

Illumina manufacturers then created the custom amplicon oligos for downstream hybridization and amplification. The Design Studio software program allows for control of both quality and coverage per amplicon through adjusting parameters during the digital design process. According to Illumina sources, the algorithm used in the design process considers specificity, GC content, interaction of probes and adequate coverage. During the design, eighty-eight percent of targeted regions were designable and predicted to give adequate coverage. A typical design using this online software has a success rate of ninety percent for desired regions. Rates lower than 90% may occur because of problematic regions of interest such as regions of homology or high G-C content.

3.2 Materials and Methods

3.2.1 Selection of Families and Collection of Germline DNA

Probands from 115 unrelated families that met clinical criteria established by the IGCLC for HDGC (n=106)(Fitzgerald et al. 2010) or FIGC (n=9) (Caldas et al. 1999 were included in this study. All 106 HDGC families previously tested negative for *CDH1* mutations and deletions. Of the 106 HDGC families, 37 had two or more documented cases of DGC in first or second-degree relatives, with at least one being diagnosed before the age of 50. Families consenting to participate in this ethics-approved study were drawn from three centers: British Columbia Cancer Agency (BCCA) (Vancouver, Canada), University of Siena (Siena, Italy), and the Institute of Molecular Pathology and Immunology of the University of Porto (IPATIMUP) (Porto, Portugal). Germline DNA was extracted at each institution to screen for *CDH1* point mutations and large rearrangements, and material was subsequently frozen upon obtaining negative test results.

Table 3.1. Breakdown of DNA samples collected and sequenced against panel for CIHR-funded upper GI Study

Institution	Diffuse Gastric Cancer		Intestinal Gastric Cancer		Pancreatic Cancer	Esophageal Cancer / Barrett's Esophagous	Other UGI Cancers	Positive Controls (CDH1 +)
DNA Type:	Normal	Tumour	Normal	Tumour	Normal	Normal	Normal	Normal
BC Cancer Agency	65					16		17
UK						32		
McGill					13		4	
Mount Sinai					29			
Porto, Portugal	29		2					
Sienna, Italy	17	19	8	53				
Total: 304	111	19	10	53	42	48	4	17

Matched-tumor DNA was available for 25 cases (18 HDGC and 7 FIGC) from the Italian cohort. Proband tissue sections were requested for downstream analysis when germline truncating mutations were detected. Germline DNA from probands of 94 families with history of other upper GI syndromes were also included for sequencing analysis (**Table 3.1**).

3.2.2 Selection of Genes for Custom Panel

Through literature research and concurrent projects, a panel of 55 genes with confirmed or suggested involvement in upper GI diseases, such as gastric, esophageal, pancreatic and gall bladder cancer; lynch syndrome; cowden syndrome; carney-stratakis syndrome; gastric polyps and Peutz-Jeghers disease, was developed (**Table 3.2**). Genes were selected based on particular criteria: 1) well established TSG in the literature with strong genetic susceptibility to upper GI cancer or cancer-susceptibility syndromes; 2) established TSG in literature with suggested pathogenic germline mutation(s) with a family history of upper GI cancers or cancer-susceptibility syndrome; 3) suggested TSG in the literature, with germline mutation(s) in hereditary upper GI cancers or cancer-susceptibility syndromes; 4) suggested TSG in the literature, with loss of protein expression in upper GI tumours to suggest loss of heterozygosity; 5) candidate genes from data based on WGS or ES of upper GI families from collaborators. Of these 55 genes, 14 were included for specific relation to GC and were suspected to hold pathogenic variants in the FGC cases screened. By screening families against this gene panel, we anticipate discovery of pathogenic mutations in families with unexplained heritable forms of FGC, while avoiding the costly and unspecific nature of broader sequencing technologies. This project will then triage families in which no candidate variants are identified for more intensive, WGS efforts.

Table 3.2 Genes selected for custom panel based on association with upper GI syndromes

Syndrome	Gene	Mutation Status	Penetrance
Carney-Stratakis	*SDHB*	Heterozygous	High
	SDHC	Heterozygous	High
	SDHD	Heterozygous	High
Colorectal Carcinoma & Polyposis	*MUTYH*	Homozygous	High
Esophgeal Adenocarcinoma / Barrett's Esophagous	*AKAP12*	Heterozygous	High
	CTHRC1	Heterozygous	Intermediate
	FOXF1	Heterozygous	Intermediate
	MSR1	Heterozygous	Intermediate
Familial Adenomatous Polyposis	*APC*	Heterozygous	High
Gastric Cancer	***ARID1A***	**Heterozygous**	**High**
	BCL2L10	**Heterozygous**	**High**
	BRCA1	**Heterozygous**	**High**
	BRCA2	**Heterozygous**	**High**
	CASP10	**Heterozygous**	**High**
	CDH1	**Heterozygous**	**High**
	CTNNA1	**Heterozygous**	**High**
	FAT4	**Heterozygous**	**Unknown**
	FHIT	**Heterozygous**	**High**
	HSPA5	**Heterozygous**	**High**
	IDH1	**Heterozygous**	**Unknown**
	IDH2	**Heterozygous**	**Unknown**
	PSCA	**Heterozygous**	**High**
	PTEN	**Heterozygous**	**High**
Cowden Syndrome	*PTEN*	Heterozygous	High
	SDHB	Heterozygous	High
	SDHD	Heterozygous	High
Gastrointestinal-type Polposis	*MSH3*	Heterozygous	High

Table 3.2 (continued) Genes selected for custom panel based on association with upper GI syndromes

Syndrome	Gene	Mutation Status	Penetrance
Hereditary Mixed Polyposis Syndrome	GREM1	Heterozygous	High
	SCG5	Heterozygous	High
	TGFR2	Heterozygous	High
Juvenile Polyposis & Pancreatic Cancer	BMPR1A	Heterozygous	High
	SMAD4	Heterozygous	High
Lynch Syndrome	EPCAM	Heterozygous	High
	MLH1	Heterozygous	High
	MSH2	Heterozygous	High
	MSH3	Heterozygous	High
	MSH6	Heterozygous	High
	PMS1	Heterozygous	High
	PMS2	Heterozygous	High
Pancreatic Cancer	ATM	Heterozygous	Intermediate
	CDKN2A	Heterozygous	High
	CFTR	Heterozygous	High
	CHEK2	Heterozygous	Intermediate
	PALB2	Heterozygous	High
	PRSS1	Heterozygous	High
	SPINK1	Heterozygous	High
	TP53	Heterozygous	High
Peutz-Jeghers Syndrome	STK11	Heterozygous	High
Collaborative Projects	AKR7A3	Heterozygous	Unknown
	GAB2	Heterozygous	Unknown
	ITIH2	Heterozygous	Unknown
	MAP3K6	Heterozygous	Unknown
	MCCC1	Heterozygous	Unknown
	PRR5	Heterozygous	Unknown
	PXN	Heterozygous	Unknown
	SCARF2	Heterozygous	Unknown
	SLC22A4	Heterozygous	Unknown

3.2.3 Description of TruSeq Custom Amplicon Protocol

Genomic DNA Extraction/Quantification and Procedure

Before the sequencing process can begin, samples were extracted and quantified to ensure genomic DNA met assay input requirements. Genomic DNA (gDNA) from probands of each family of interest was previously extracted at each initiating institution for initial *CDH1* sequencing. Upon negative test result, gDNA was stored at -20 or -80 degrees Celsius to maintain quality of DNA.

Input requirements for the TSCA protocol are dependent on both quality of gDNA and designed amplicon size. For this particular experiment, amplicons were designed to be 250bp in length and genomic DNA (gDNA) input requirements for these parameters are recommended to be 250ng (but may go as low as 150ng). gDNA from FGC families were quantified using Qubit fluorometer dsDNA Broad Range Kit according to manufacturers instructions (Invitrogen, Carlsbad, CA). One sample was not of sufficient quantity and was removed from cohort; all others exceeded minimum input requirements and were diluted (when necessary) to 250ng in 5ul. It should be noted that the protocol used is subjected to change and has been updated since its use in 2013. The most recent version (TSCA v1.5) reflects the improved reagents to account for GC-rich regions and increased library yield, uniformity and stability. Additionally, the new kit has optimized its amplification steps to enable lower sample input to a minimum of 50ng. Samples were subjected to the previously 2013 version, TSCA v2.0 protocol, as per manufacturers instructions.

Day 1

The first day of the streamlined TSCA protocol includes hybridization and extension of custom

primers to individual samples, addition of unique index primers and a PCR-based amplification.

Custom amplicon oligonucleotides (oligos) were added to each individual 250ng sample in a 96-

well plate. This plate containing samples and custom oligos is heated to 95 degrees Celsius for

one minute and gradually reduced to 40 degrees Celsius over the course of 80-90 minutes to

hybridize custom oligos to targeted regions. The gradually decrease in temperature is critical for

proper hybridization to specific regions. Samples are transferred to a filter-plate unit that contains

a 96-well filter plate securely placed on top of a 96-well 1.5mL MIDI plate. Filter plate is

centrifuged to filter through unbound oligos and capture larger gDNA on the filter. To ensure the

removal of unbound oligos, two cycles of stringent wash reagent is used a filtered through each

sample followed by a universal buffer to keep pH at a constant value and ensure hybridized

oligos remain bound to gDNA.

Extension and ligation of hybridized oligos is then completed to attain the targeted region for

amplification by adding the Extension-Ligation Mix 3 to each individual sample. At each custom

oligo pair, DNA polymerase extends from the upstream 3' oligo across the targeted region where

a DNA ligase then ligates this extension to the 5' end during a 45-minute incubation period at 37

degrees Celsius. This results in the formation of targeted products that are ready for PCR-

amplification. Fresh NaOH (50mM) is then added to each individual sample to ensure

denaturation of double stranded DNA and to neutralize inhibitors of Taq polymerase during PCR

reaction. The PCR-base amplification follows standard preparation and reagents but require the

addition of index sequences for downstream sample multiplexing. These index barcodes are

small 8bp sequences that individualize each sample for downstream recognition post-sequencing. Prior to addition of these indices, a sample sheet must be created to ensure no two samples have the same indices and that the indices added will be diverse enough for unique recognition. PCR master mix (provided by Illumina) and polymerase reagents are pooled together according to the number of samples then 22ul of this master mix is added to each individual sample. Samples are then loaded onto the thermocycler using the following program: Step 1: 95°C for 3 minutes; Step 2: 95°C for 30 seconds; Step 3: 66°C for 30 seconds; Step 4: 72°C for 60 seconds; Step 5: Return to Step 2 and repeat for 22 more cycles; Step 6: 72°C for 5 minutes. Samples are held at 10 degrees Celsius overnight before verification of amplification using a 2% gel electrophoresis.

Day 2

PCR-amplified samples were validated using a 2% gel electrophoresis, including Illumina positive controls. Samples that were adequately amplified are carried forward to the library preparation and normalization protocol.

PCR-cleanup stage using AMPure XP beads was completed on all individualized samples post PCR-amplification to purify products from other reaction components. AMPure XP magnetic beads have a high affinity to bind small PCR-amplified amplicons of single stranded DNA. After addition of magnetic beads to PCR mixtures, amplified targeted regions become bound to them over a ten-minute waiting period at room temperature. The plate is then placed on a magnetic stand, long enough for beads to attract to the side of the well. Supernatant containing PCR components is removed and a 70% ethanol wash is performed twice to remove any unwanted elements. Illumina then provides an elusion buffer to release PCR amplicons from the AMPure

XP beads. A micro-plate shaker is used to mix beads and reagents when necessary for control purposes according to manufacturers instructions. Once the magnetic stand is applied again, the supernatant now contains the desired amplified PCR-products for transfer to a fresh 96-well MIDI plate for the library normalization.

Library normalization is a critical step for multiplex NGS protocols. It ensures that an equal representation of all desired amplicons is obtained before proceeding with downstream sequencing phase. Though different normalization techniques exist (i.e. spectroscopy and size-restricted spectroscopy), the technique use in Illumina TSCA assay, known as quantitative binding, is proven superior and represents the most efficient process for construction large, multiplexed amplicon pools for sequencing (Harris et al. 2010). During this process, Illumina Library Normalization Beads (LNB), which have a maximum binding capacity for single stranded DNA amplicons, are first mixed with Library Normalization Additive (LNA) reagents (provided by Illumina) to create a master mix (8.18ul LNB and 36.82ul LNA per sample). When placed on the magnetic stand, amplicons are inadvertently bound to normalization beads to their maximum capacity. Beads are washed with the appropriate Library Normalization Wash then DNA is eluted from the beads used freshly prepared 0.1N NaOH. Any unbound amplicons will be removed with the supernatant in this downstream washing phase. The final library pool now consists of amplified single stranded DNA-regions of interest. It is suggested that quantitative PCR can be performed at this time if DNA is of poor quality (i.e. from FFPE tumours). This procedure was performed for the first run with 24 germline samples, where sufficient quantity and quality was demonstrated.

Five ul of each normalized library was placed in an eppendorf tube and 6ul total was transferred into a fresh tube along with 594ul of Hybridization Buffer supplied by Illumina. The remainder of the pooled amplicon tube can be stored in -20 degrees for dilution adjustments if inadequate cluster density is found during the sequencing reaction. If cluster density is poor, it is recommended that a larger ratio of barcoded, pooled amplicons to hybridization buffer be used. However, cluster density was adequate for all sequencing runs so this ratio was not changed. The diluted amplicon library is then placed on the heat block and on ice for two and five minutes, respectively, for denaturation of dsDNA.

To prepare the MiSeq platform for loading, the appropriate MiSeq cartridge (v500-cycles) must be completely thawed and mixed by inverting several times. The supplied flow-cell must be appropriately washed with 70% ethanol and distilled water then delicately dried for proper cluster generation and sequencing. Once the flow-cell is loaded onto the MiSeq platform, the entire mixture of pooled sample library and hybridization buffer (600ul total) is added to the appropriate position on the MiSeq cartridge. The touch screen on the MiSeq instrument takes the user through the steps to start the sequencing process to be completed over a 48-hour period. During this time, the MiSeq instrument simultaneously completes cluster generation, reverse-terminator- sequencing and data analysis.

3.2.4 Organizing Samples to be Multiplexed

Samples were organized and multiplexed to attain desired cluster generation and coverage across regions of interest. As previously mentioned, the FGC germline samples included in this project were sequenced alongside germline samples from other familial upper GI cases as part of a

larger funded grant (**Table 1.1**). A total of 295 samples were separated into 7 sequencing runs over the course of the project (**Table 3.3**). All sequencing runs followed the same protocol and use of reagents supplied by Illumina. Positive controls supplied by Illumina manufacturers were also sequenced to ensure quality control for each individual run.

A total of 111 germline HDGC samples and 9 FIGC germline samples were sequenced against the custom amplicon panel. Three families (5 samples total) were later removed from the study post sequencing, as they were found to not meet criteria for hereditary DGC (Fitzgerald et al. 2010). Data from 106 HDGC and 9 FIGC families were considered for downstream analysis. As well, tumour DNA from 7 FIGC and 18 HDGC cases (all with matched normal DNA) were also sequenced against this custom panel of UGI-associated genes.

3.2.5 Data Analysis

Data can be extracted directly off the MiSeq in three file formats: .vcf, FastQ and BAM. FastQ files can be used for secondary analysis by a bioinformatician whereas .vcf and BAM files can be extracted directly for examination of variants per sample on excel or on Integrated Genome Viewer to visualize the coverage depth across all regions of interest.

On average, 7.54 Gb of aligned sequence was produced per run with coverage varying significantly across individual samples and amplicons in each run, with a mean of 838x and median of 534x. Sequencing coverage in next generation sequencing assays is highly dependent on the number of amplicons in the custom design and the number of samples multiplexed in the particular sequencing run (**Table 3.2**). The maximum number of samples multiplexed in a single

Table 3.3 A break down of the number of multiplexed samples per TSCA sequencing run

Run No.	Diffuse Gastric Cancer		Intestinal Gastric Cancer		Pancreatic Cancer	Esophageal Cancer /Barrett's Esophagous	Other UGI Cancers	Positive Controls (CDH1 +)	Total No. Samples Multiplexed
DNA Type:	Normal	Tumour	Normal	Tumour	Normal	Normal	Normal	Normal	
1	17							7	24
2					13	45	4	1	63
3	29		9		29	3		1	71
4	8	18		7				2	35
5	40			1					41
6				38					38
7	17		1					5	23
Total	111*	18	10*	46	42	48	4	16	295^

*Number is indicative of number of samples sequenced, not number of families included for these disorders;
^ 9 samples were rerun due to amplification failure or sequencing errors (total of 304 attempted samples).

run was 71, with adequate but variable coverage and some regions of low (<50x) mean coverage. For this reason, the number of multiplexed samples for the remaining runs was lowered to increase the probability of high coverage across all regions of interest.

Selection of Candidate Variants

Genome Analyzer software (Illumina, San Diego) was used to call variants on the MiSeq instrument. Files (.vcf) were extracted directly off the MiSeq and analyzed for candidate variants in the form of insertions/deletions, nonsense splicing, novel missense and dbSNP variants. Single nucleotide polymorphism database (dbSNP) is an online public archive for the frequency of individual genetic variants across the population. This domain is a useful tool to use as a control when attempting to identify rare or novel variants leading to increased predisposition for a familial syndrome. Each separate file extracted off the MiSeq instrument contained all variants called across the 1531 amplicons for each individual sample. After converting each file to an excel spreadsheet, candidate variants were selected and sorted based on likelihood of pathogenicity. Firstly, each sample file was searched for frameshift and nonsense mutations, as heritable protein truncating mutations are the most likely pathogenic mutations in germline cases. All truncating mutations were highlighted and considered top candidates for validation. This included *CTNNA1, BRCA2, ATM, PALB2,* and *MSR1* truncating mutations to be discussed in **Chapters 4-5**. Protein truncating mutations were also identified in familial pancreatic cases. This data was passed onto collaborators who submitted the samples for validation and downstream analysis. Novel missense mutations were selected for each individual sample and sorted based on functional impact of amino acid change using *in silico* methods (PROVEAN, SIFT, PolyPhen), which is to be discussed further in **Chapters 4-5**.

Genome Analyzer software (Illumina, San Diego) was able to call single nucleotide

polymorphisms (SNP) that lied within the regions of interest. These were distinguished between

novel missense mutations by a dbSNP identification number. Using this identification number,

SNP can be sorted based on frequency in the general population (ensemble.org). SNP were

considered rare and included in downstream analysis if they were seen in <1% of the general

population. Interestingly, this database also includes information on the pathogenicity of such

SNP and if the variant has been associated with a particular disease. All pathogenic or possibly

pathogenic rare missense mutations were noted and researched further to determine their

relationship to upper GI syndromes. Using this technique, rare missense mutations in *SDHB* and

STK11 were identified that had previously been related to hereditary cancer-risk syndromes

Cowden-like syndrome and Peutz-jeghers disease (**Chapters 5**).

Secondary Analysis by Bioinformatician

To gauge accuracy of the on-instrument analysis, raw data from the first 25-germline cases and

18 tumor-normal matched samples were analyzed through a custom variant calling analytical

pipeline by a bioinformatician. All sequences in the raw data files from 25 cases were first

trimmed by 225 base pairs to account for sequence masking done by the MiSeq software in the

event that amplicon regions were shorter than 250 base pairs. Samples were then aligned

(Burrows-Wheeler Alignment tool-0.5.9) (Li & Durbin, 2009) to a custom genome then

repositioned for variant calling on the full set of loci comprised by the amplicon coordinates.

Reads with greater than 5 mismatches were filtered out, and those remaining were called for

SNV using SNVmix2 (Shah et al. 2009; Goya et al. 2010), after passing thresholds for base

quality (>10) and mapping quality (>20). Each SNV position was tested using binomial exact

test, with the background distribution defined by the reference allele frequency from the entire amplicon coordinate list. The resulting p-value was corrected for the number of total tests using the Benjamini Hochberg FDR (Benjamini et al. 1995) procedure and any resulting q-value <0.00001 were considered true mutations. Indels were called using SAMtools (Li et al. 2009), separating lines containing short indels and filtered on a minimum SNP quality score (>300), number of variant reads (>3) and variant "allele" frequency (≥0.1). Coding indels were annotated using information in Ensembl.

Resequencing Candidate Variants for Validation

PCR was performed on candidate pathogenic and likely pathogenic mutations. Variant-specific primers were designed using Integrated DNA Technologies (Coralbille, Iowa) online software: (*CTNNA1* (N71*) F5'-TTGTAAGGTTTACTGGGTCTTCA-3' and R5'-GGTTAACTAAACCCATGCATCAA-3'; *CTNNA1* (R129X) F5'-TGAAAACTCTTAAACTA AATTTGTGC-3' and R5'-AAAACATCTCTGGTCCATT GAGA-3'; *BRCA2* (N1287*) F5'-GTGGGTTTGCAATTTATAAAGCAG-3' and R5'-TAACTTACCAGAAGCTTGTTTCC-3'; *BRCA2* (K936*) F5'-GGAACTTCATGAAACAGACTTGAC-3' and R5'-CTATATTCAAGG AGATGTCCGATT-3'; *SDHB* (S163P) F5'-GCTGAGGTGATGATGGAATCT-3' and R5'-ACCACACTCCTGGCAATCATCT-3'; *STK11* (F354L) F5'-GAGGAGCTGGGTCGGAAA-3' and R5'-TGGCCGAGTCAGCAGAG-3'; *PALB2* (V398*) F5'-GAAAGTGAGATTCTAAGTC AACCTAAG-3' and R5'-TTCTTGACATCCAAATGACTCTG-3'; MSR1 (R293X) F5'-AGTA CCTTGACAGATGACTAACC-3' and R5'-CCCTACACATGTACCTGG ATG-3'; ATM E1267* F5'-TGTAAAACGACGGCCAGTGCTACTGAACAAGGTCCCATTT-3' and R5'-CAGGAAACAGCTATGACCCAGTCCTCTTGAATCTGATTAGC-3; ATM R521* F5'-GAG

GTCAAACCTAGAAAGCTCA-3' and R5'-GTGTGTGTCTGTGTGTGTTTATC-3'; ATM

Y2791* F5' – GCTGAATGATCATCAAATGCTCT- 3' and R5' –ATGGCTTATTAAAGCTG

ACAGC- 3'). High Fidelity Taq Polymerase (Life Technologies, Carlsbad, CA) was used

during PCR according to manufacturer's instructions. Post-sequencing cleanup was performed

using ExoSAP (Affymetrix, Santa Clara, CA) prior to tagging with M13 primer sequences and

Big-dye fluorescence (Life Technologies) and sequencing on a 3130xl Genetic Analyzer

(Applied Biosystems, Carlsbad, CA). M13 primers were not used in the case of *CTNNA1*

validations as primers were borrowed directly from a collaborating institution and were not

initially tagged with the sequences (Majewski et al. 2013).

In Silico Methods Predict Pathogenicity of Variants of Unknown Significance

Missense variants with dbSNP identification numbers were sorted using ensembl.org based on

frequency in the general population. Rare variants (<1% frequency) and novel missense

mutations were then subjected to three *in silico* analyses to predict functional implication of

amino acid change: Protein Variation Effect Analyzer (PROVEAN) (Choi et al. 2012; Choi,

2012), SIFT (Kumar, Henikoff & Ng, 2009; Ng & Henikoff, 2006 & 2003) and PolyPhen-2

(Adzhubei et al. 2010). Those variants predicted as damaging to pathogenic in at least two of the

three software programs were considered possibly pathogenic. No further downstream analysis

was done on these variants of unknown significance as it is difficult associate missense

mutations with hereditary syndromes without extensive *in vitro/in vivo* analyses, as previously

mentioned with respect to *CDH1* missense mutations. These mutations pose a burden on genetic

counselors, as it is difficult to communicated relative risk to families. The genes containing these

variants of unknown significance (VUS) may hold pathogenic or likely pathogenic mutations in

other families with GC and for this reason, including them in further studies may prove beneficial in establishing them as susceptibility genes. This is further addressed in the future directions section of **Chapter 6** with respect to screening a larger cohort of FGC cases against a panel of newly described, possible susceptibility genes.

3.2.6 Downstream Validation of Pathogenic and Likely Pathogenic Variants

To gauge pathogenicity of candidate variants, tumour materials from the proband with detected germline mutation are studied to identify mechanisms that may disrupt protein expression. Immunohistochemical (IHC) staining using antibodies against the protein of interest can be conducted on Formalin-fixed paraffin-embedded (FFPE) tumour materials with germline truncating mutations to identify a loss of protein expression compared to normal material. This then confirms if the truncating mutation in question is disrupting the production of a functional protein. If tumour DNA is available from the candidate variant carrier, a somatic second-hit analysis can be completed to identify a source of loss of heterozygosity for the gene in question. Tumour suppressor genes that have a germline truncating mutation will often have such somatic second-hit mutations to knock out the remaining functional allele that can be seen by sequencing DNA from the tumour. When available, additional affected and unaffected family members can be genotyped for the same mutation through Sanger sequencing to identify penetrance and confirm pathogenicity of the heritable variant. If the mutation is heritable in affected individuals, it is considered highly pathogenic and has likely significantly increased the carrier's risk of developing the disease.

Immunohistochemistry of Candidate Truncating Variant Carriers

FFPE tissue blocks containing whole sections of tumor were obtained from a surgical resection

specimen corresponding to the proband of families 25 and 80, found to carry truncating *CTNNA1*

mutations. Tissues were assessed for IHC expression of E-cadherin, alpha-catenin, cytokeratin

and CD31 and reviewed by two pathologists. Sections (4μm in thickness) were deparaffinized,

rehydrated, and stained using the semi-automated Ventana Discovery® XT System (Ventana

Medical Systems, Tucson, AZ, USA). Antigen retrieval was performed using Cell Conditioning

Solution 1 (Ventana). The primary antibodies used for staining were EP1793Y (Rabbit

monoclonal for alpha-catenin; 1:200 dilution; Abcam, Cambridge, UK), JC/70A (Mouse

monoclonal for CD31; 1:100 dilution; Dako, Carpinteria, CA, USA), CAM5.2 (Mouse

monoclonal for cytokeratin; 1:5 dilution; BD Biosciences, Franklin Lakes, NJ, USA) and NCH-

38 (Mouse monoclonal for E-cadherin; 1:25 dilution; Dako). Pre-diluted UltraMap® (Ventana)

anti-Rabbit and anti-Mouse horseradish peroxidase secondary antibodies were used, and signal

detection was performed using the UltraMap® DAB Detection Kit (Ventana). The cytokeratin

and CD31 stains were used to delineate epithelial cells and endothelial cells, respectively. Two

registered pathologists, Dr. D F. Schaeffer and Dr. H Li-Chang, examined the specimens for

tumour content and presence or absence of E-cadherin and alpha-catenin staining. Histologic

images were obtained using the Olympus DP21® digital camera (Olympus, Shinkuju, Japan).

Tumour material in the form of FFPE block was also available for a single HDGC case with a

germline truncating *ATM* variant. The anti-ATM antibody used for IHC was Y170 (ab32420)

(Rabbit monocloncal for ATM; Abcam, Cambridge, UK). Upon review by pathologist Dr. D

Huntsman, it was concluded that the tumour material sent was not from stomach and was likely

from a breast metastasis. Staining was of poor quality and it could not be concluded by comparison to normal material if there was loss of protein expression.

Somatic Second Hit Analysis

Due to both the rarity of this disease and high mortality rate, tumour DNA of candidate mutation carriers was only available from Family 80 with truncating mutation in *CTNNA1*. Somatic mutation analysis did not reveal a second hit in the tumour material. However, loss of heterozygosity is suggested as loss of protein expression in the tumour was described during IHC analysis (**Chapter 5**). Several mechanisms may be at play to account for this loss of protein expression, such as epigenetic modification or functionally relevant non-coding mutations within the *CTNNA1* locus. Tumour material was also available from two affected individuals carrying the *ATM* truncating mutation in Family 104 (E1267*). Both mutant and wild-type *ATM* alleles were present in both cases, suggesting no loss of heterozygosity at the *ATM*-locus.

Inheritance Pattern of Variant Across Additional Family Members

Germline DNA was available from the affected mother of the proband from Family 25 who was found to have a *CTNNA1* truncating mutation. Sanger sequencing of the region of interest showed that the affected mother (diagnosed with GC at age 59) is a carrier of the truncating mutation, further supporting its highly pathogenic nature (**Figure 4.2**). Additional family members from Family 104 with a truncating mutation in *ATM* (E1267*) were available for genotyping to gauge inheritance pattern. Two affected and two unaffected individuals were genotyped for the mutation in question using Sanger sequencing techniques.

Chapter 4: Results

Of the 115 probands, pathogenic or likely pathogenic variants were identified in 15 cases (13.0%), of which 11/106 were HDGC families (**Table 4.1; Figure 4.1**) and 4/9 were FIGC families (**Table 4.1**). Pathogenic mutations were defined as protein-truncating mutations that exist in a gene previously related to an upper GI disorder with highly penetrant mutations. Rare, missense mutations in genes previously related to GC susceptibility syndromes with supporting *in vitro* data were also considered pathogenic. Protein truncating mutations were labeled as 'likely pathogenic' if they were found in a gene with previous data suggesting lower-moderate penetrance. All pathogenic and likely pathogenic variants (**Table 4.1**) were validated via Sanger sequencing. Novel missense variants predicted to be damaging were found in another 27 cases (23.5%) (**Table 4.2**). The TSCA assay on the MiSeq platform provided adequate but variable coverage across amplicons, ample cluster density in all runs and greater than 90 percent of clusters passing quality score on average.

4.1 Pathogenic Variants in Familial Gastric Cancer

Pathogenic truncating mutations in *CTNNA1* (N71fs and R129X) (**Figure 4.2**) and *BRCA2* (N1287fs) (**Figure 4.3**) were found in three unrelated HDGC families using the TSCA assay (**Table 4.1**). An additional truncating variant in *BRCA2* (K936fs) (**Figure 4.3**) was found in a FIGC family. These genes all have suggested involvement in the development of GC, which further supports their roles as susceptibility genes in familial subtypes (Goldgar et al. 2011; Renwick et al. 2006; Erkko et al. 2008; Byrnes et al. 2008; Maier et al. 2006; Orloff et al. 2011). Previous studies suggest pathogenic mutations within these genes have high penetrance with respect to familial upper GI syndromes, further supporting their causality in these FGC cases

(Kluijt et al. 2012; Majewski et al. 2013; Moran et al. 2012). Rare pathogenic missense variants in *STK11* (F354L) and *SDHB* (S163P) (**Figure 4.4; Figure 4.5**), previously associated with Peutz-Jegher's disease and Cowden-like syndrome, respectively, were also identified (Forcet et al. 2005; Ni et al. 2008).

Germline DNA from another affected family member was available from Family 25 who carried the N71fs truncating *CTNNA1* (alpha-catenin) mutation. Sanger sequencing was completed for the targeted region on the affected mother of the proband who also was diagnosed with DGC at age 59. The protein truncating mutation was shown to be inherited from the proband's affected mother, strongly supporting it's likelihood for pathogenicity (**Figure 4.2-b**). IHC staining of the tumor from both families showed loss of alpha-catenin expression, suggesting loss of heterozygosity at the *CTNNA1* locus (**Figure 4.6, Panels C and F**), while E-cadherin expression was preserved (**Figure 4.6, Panels B and E**). Alpha-catenin staining of tumour material from proband of Family 80 also revealed loss of protein expression in the signet-ring cells (**Figure 4.7**). Somatic mutation analysis was performed on tumor DNA from Family 80, but did not reveal a somatic second-hit. Tumor DNA was available for the proband in FIGC Family 109, who is a carrier of a novel truncating *BRCA2* variant (K936fs), but a somatic second-hit mutation was not identified in the *BRCA2* coding region. As tumor sections were not available, IHC analysis could not be completed on families carrying *BRCA2* mutations (**Figure 4.3**). With a strong history of highly penetrant mutations at the *BRCA2* locus in relation to hereditary diseases, these mutations are strongly suggested as pathogenic and, thus, the first *BRCA2* protein truncating mutations described in a clinically defined GC families.

Table 4.1 Candidate germline variants from TSCA panel sequencing runs

ID	Initial Institution	IGCLC Criteria Met	Proband (AOD)	Relatives with GC or BC (AOD)	Family Hx Other Cancers (AOD)	Additional Family Hx	Gene (Chromosome)	Position	Consequence	Coverage Depth	Mutation Type
					PATHOGENIC						
25	BCCA	HDGC, c	DGC (22)	GC (59), BC (70)	Brain (70), GEJ (82)	-	CTNNA1 (chr.5)	c.211A>AT	N71fs	245	Frameshift
80	Italy	HDGC *	DGC (72)	DGC (52)	-	-	CTNNA1 (chr.5)	c.385C>T	R129X	442	Nonsense
2	BCCA	HDGC, a	DGC (64)	DGC (21), BC (50)	-	-	BRCA2 (chr.13)	c.3862TAATA>T	N1287fs	205	Frameshift
109	Italy	FIGC	FIGC (unknown)	-	-	-	BRCA2 (chr.13)		K936fs	154	Frameshift
13	BCCA	HDGC, d	LBC (39)	GC (53), GC (44), BC & Uterine (34), BC (unknown)	Brain (unknown)	Developmental delay on paternal side	SDHB (chr.1)	c.487T>C	S163P[1]	248	Missense
110	Italy	FIGC	FIGC (unknown)	-	-	Poor prognosis for FIGC	SDHB (chr.1)	c.487T>C	S163P[1]	74	Missense
44	BCCA	HDGC, c	DGC (37)	GC (70), LBC (45), BC (33), BC (56)	GEJ (unknown), CRC (70), Prostate (unknown)		STK11 (chr.19)	c.1062C>G	F354L[2]	681	Missense
46	BCCA	HDGC, c	DGC (22)	-	Father Dx n/a (Lung)	-	STK11 (chr.19)	c.1062C>G	F354L[2]	681	Missense

Table 4.1 (Continued) Candidate germline variants from TSCA panel sequencing runs

ID	Initial Institution	IGCLC Criteria Met	Proband (AOD)	Relatives with GC or BC (AOD)	Family Hx Other Cancers (AOD)	Additional Family Hx	Gene	Position	Consequence	Coverage Depth	Mutation Type
					LIKELY PATHOGENIC						
96	Portugal	FIGC	IGC (45)	IGC (60), GC (45), GC (73), GC (50), 2 x BC (unknown)	CRC (unknown), Ov (56), Leukemia (84)	-	PALB2 (chr.16)	c.1193AC>A	V398fs	44	Frameshift
16	BCCA	HDGC, b	GC (58)	GC (72), GC mda (51), GC (52), DGC (unknown),	Uterine (unknown), Cervical (61), Lung (71), Bladder (69), endometrium adeno. (unknown), Thyroid (51), Prostate (unknown),P (62), Bone (9), CRC (61)	-	ATM (chr.11)	c.3800AG>A	E1267fs	244	Frameshift
104	Portugal	FIGC	IGC (72)	IGC (57), GC (72)	-	-	ATM (chr.11)	c.3800AG>A	E1267fs	105	Frameshift
42	BCCA	HDGC, d	BC (59)	GC (71), GC (82), GC (53), GC (38, GC (42), GC (59), GC (unknown), GC (73)	Ov (49), Ov (74), Head & Neck (78), CRC (42), CRC (39), Leukemia (unknown), Leukemia (55), CRC with GC mets (57), P (70)	-	ATM (chr.11)	c.8369GATAC>G	Y2791fs	258	Frameshift
58	BCCA	HDGC, d	LBC (56)	GC (40s), GC (unknown), BC & Brain (70), Bilateral BC (40s), BC (55), BC (69), LBC 49), BC (unknown), BC (unknown)	CRC (50s)	Blood clots, stroke on paternal side	ATM (chr.11)	c.1560CAG>C	R521fs	2866	Frameshift
90	Portugal	HDGC, c	DGC (22)	-	-	-	MSR1 (chr.8)	c.877C>T	R293X[3]	93	Nonsense
61	BCCA	HDGC, a		GC (50s), GC & BC (78), GC (50), GC & Liver (77), GC (87), GC (62), GC (47), BC (42)	Prostate (74), Skin (unknown), Prostate (82)	-	MSR1 (chr.8)	c.877C>T	R293X[3]	94	Nonsense
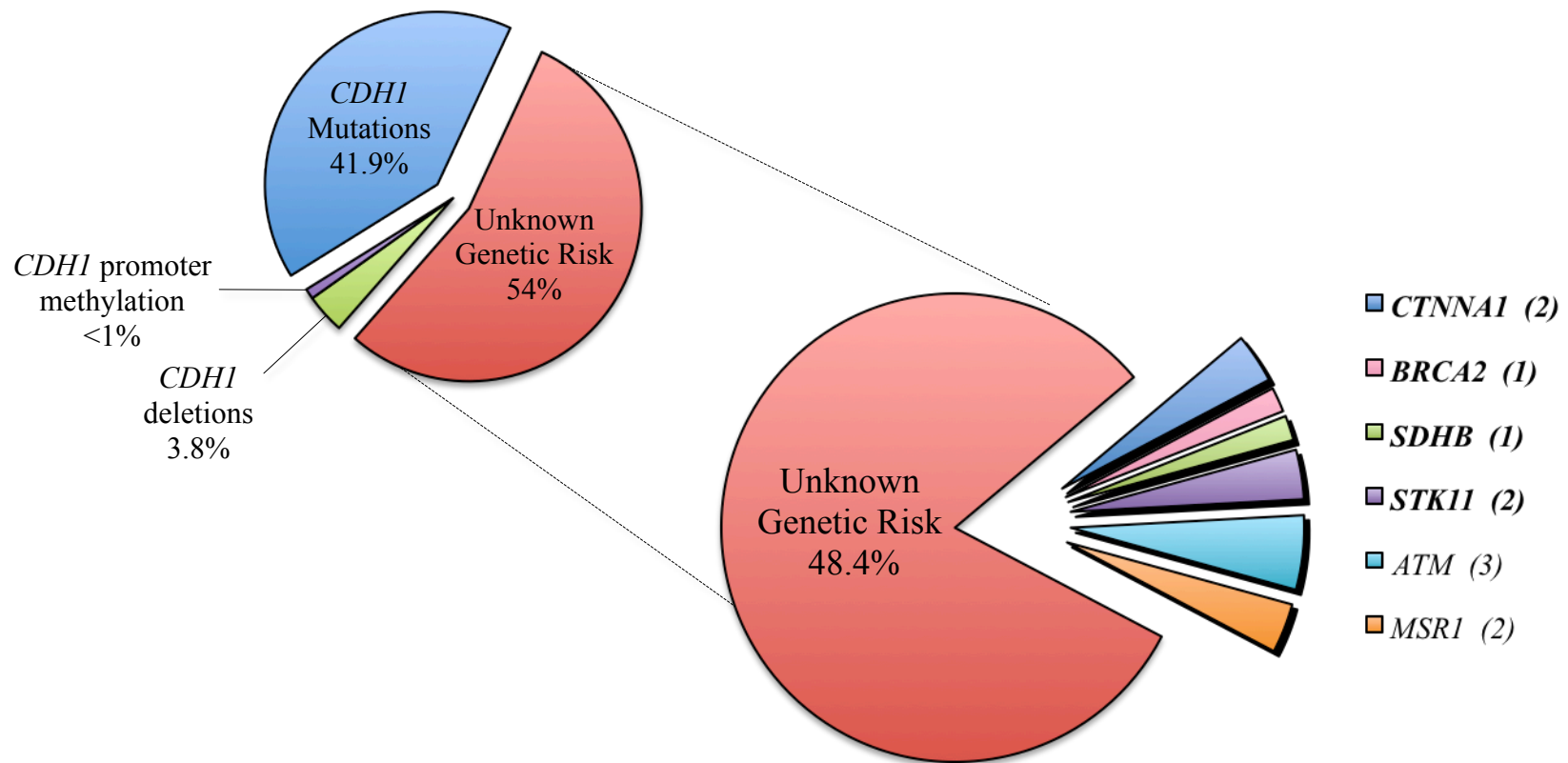
Figure 4.1 An updated mutational profile of hereditary diffuse gastric cancer. Newly identified pathogenic variants in *CTNNA1, BRCA2, SDHB* and *STK11* and likely-pathogenic variants *ATM, PALB2* and *MSR1* (number in bracket indicates the number of HDGC families with variant in the indicated gene) were identified using multiplexed, panel-based next generation sequencing in HDGC cases (n=106).

Table 4.2. Novel and rare missense mutations detected in HDGC families using a custom upper GI gene panel and predicted impact using *in silico* methods

ID	Gene (Chromosome)	Coordinate	dbSNP ID	Mutation Frequency	Ref.	Var.	Consequence	Coverage Depth	PolyPhen	PROVEAN	SIFT Score
20	AKAP12 (chr.6)	151673820	-	-	G	A	G1432R	2388	Probably Damaging	Neutral	Damaging
37	AKAP12 (chr.6)	151674257	-	-	G	C	Q1577H	2233	Probably Damaging	Neutral	Damaging
70	AKAP12 (chr.6)	151670490	-	-	G	A	E322K	1679	Probably Damaging	Deleterious	Damaging
19	AKR7A3 (chr.1)	19615041	rs145878095	0%	C	T	E55K	2397	Probably Damaging	Deleterious	Damaging
3	ARID1A (chr.1)	27089742	rs138418893	0%	G	A	A900T	246	Probably Damaging	Neutral	Damaging
68	ATM (chr.11)	108141988	-	-	T	C	S978P	734	Probably Damaging	Deleterious	Damaging
38	BMPR1A (chr.10)	88679061	-	-	T	C	L334S	4012	Probably Damaging	Deleterious	Damaging
86	BMPR1A (chr.10)	88676988	-	-	T	C	V258A	1194	Probably Damaging	Deleterious	Damaging
101	BMPR1A (chr.10)	88676994	-	-	T	C	V260A	1146	Probably Damaging	Deleterious	Damaging
56	BRCA2 (chr.13)	32915190	rs41293501	0%	C	A	A2233D	915	Probably Damaging	Damaging	Damaging
68	BRCA2 (chr.13)	32945102	-	-	A	G	K2833E	20	Probably Damaging	Neutral	Damaging
34	CASP10 (chr2)	202074096	-	-	C	G	S366C	2663	Probably Damaging	Deleterious	Tolerated
8	CFTR (chr.7)	117149143	rs115545701	0%	C	T	R74W	211	Probably Damaging	Neutral	Tolerated

Table 4.2 (continued) Novel and rare missense mutations detected in HDGC families using a custom upper GI gene panel and predicted impact using *in silico* methods

ID	Gene (Chromosome)	Coordinate	dbSNP ID	Mutation Frequency	Ref.	Var.	Consequence	Coverage Depth	PolyPhen	PROVEAN	SIFT Score
59	CFTR (chr.7)	117251657	-	-	T	A	H1054Q	2692	Probably Damaging	Deleterious	Damaging
89	CFTR (chr.7)	117250625	rs149279509	0%	A	G	Y1014C	100	Probably Damaging	Deleterious	Damaging
94	CFTR (chr.7)	117250575	rs1800111	0%	G	C	L997F	220	Probably Damaging	Tolerated	Tolerated
100	CFTR (chr.7)	117176630	rs191456345	0%	A	G	R258G	133	Probably Damaging	Deleterious	Damaging
24	EPCAM (chr.2)	47601029	rs146480420	0%	G	C	Q89H	15	Probably Damaging	Neutral	Damaging
43	ITIH2 (chr.10)	7759650	rs138977046	0%	G	A	V177M	2008	Probably Damaging	Deleterious	Damaging
3	MLH1 (chr.3)	37035088	-	-	A	C	N17T	185	Probably Damaging	Deleterious	Damaging
24	MSH2 (chr.2)	47641548	-	-	C	A	N311K	425	Probably Damaging	Deleterious	Damaging
24	MSH2 (chr.2)	47698172	rs149796110	0%	T	C	I577T	250	Probably Damaging	Deleterious	Damaging
112	PMS1 (chr.2)	190660586	rs61756360	0%	C	T	T75I	166	Probably Damaging	Deleterious	Damaging
84	PRSS3 (chr.9)	33796794	-	-	A	G	Y79C	184	Probably Damaging	Deleterious	Damaging
36	PTEN (chr.10)	89624228	-	-	T	G	M1R	1552	Probably Damaging	Deleterious	Damaging
55	PTEN (chr.10)	89624228	-	-	T	G	M1R	1372	Probably Damaging	Deleterious	Damaging
33	SLC22A4 (chr.5)	131676347	-	-	G	A	G512R	248	Probably Damaging	Deleterious	Tolerated
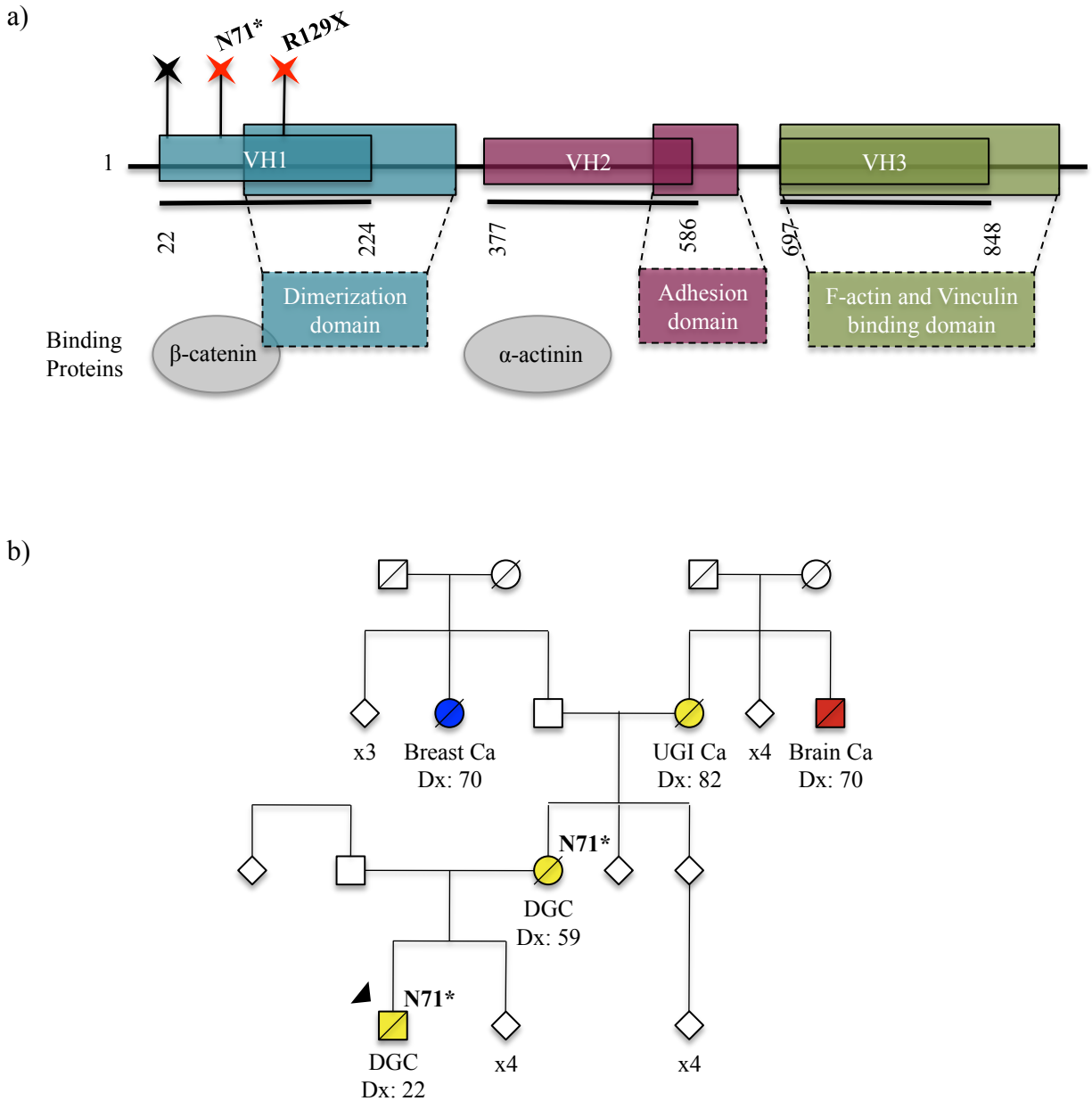
Figure 4.2 Evidence of CTNNA1 germline mutation pathogenicity and relationship to familial gastric cancer. a) Mutation distribution of 2 novel truncating (red stars) two unrelated HDGC families. Black star shows truncating mutation from Majewski et al. 2013 in close proximity to our novel mutations. B) Pedigree of CDH1-negative HDGC Family 25. CTNNA1 frameshift mutation N71fs was discovered in proband (indicated with black triangle) and segregated with the affected mother.
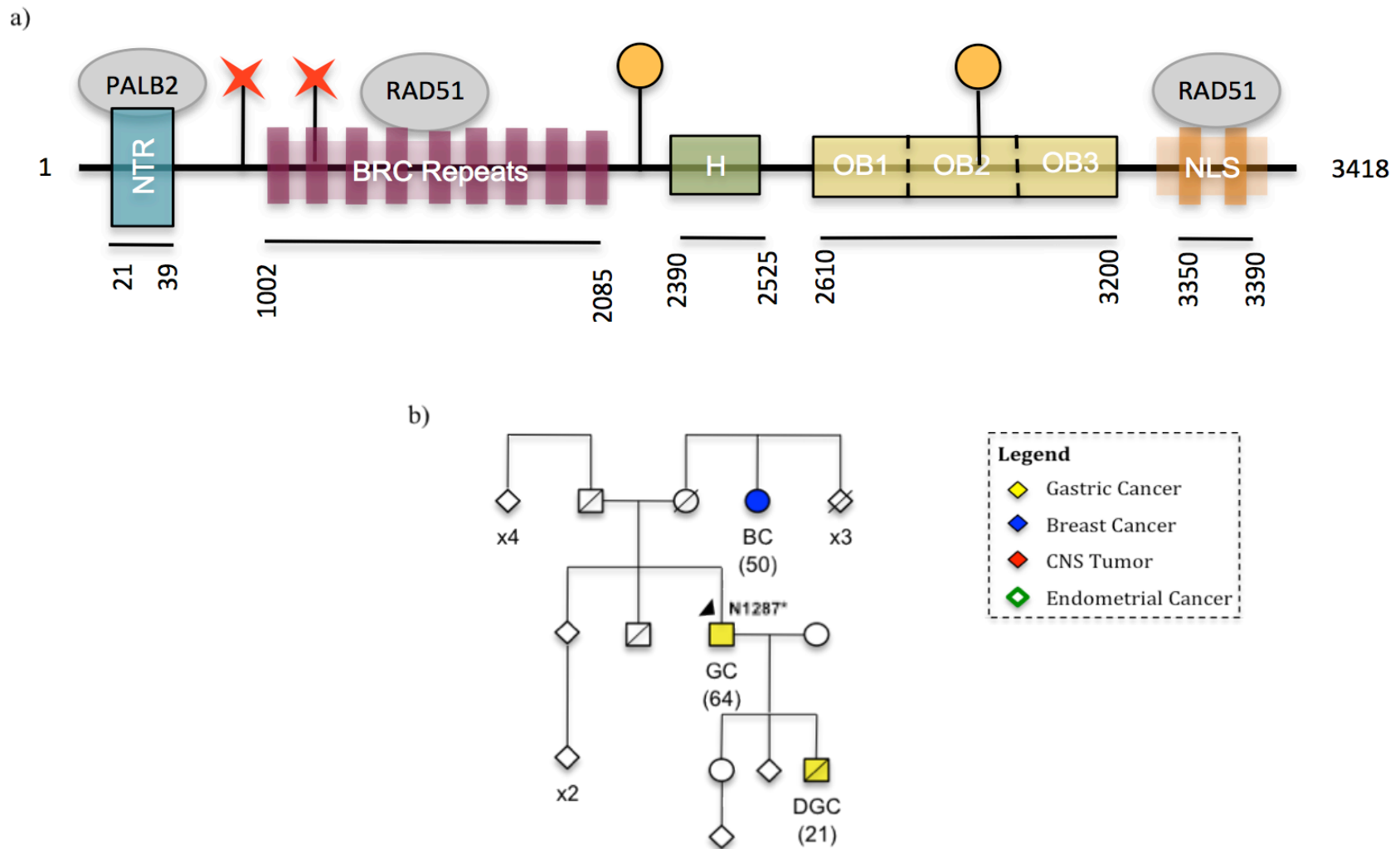
Figure 4.3 Evidence of *BRCA2* connection to ko kdcrli cuvtke'ecpegt. Novel *BRCA2* truncating mutations (red stars) and missense mutations (orange circles) identified in familial gastric cancer families using TSCA Assay. Missense mutations are considered variants of unknown significance but have been predicted as damaging in at least 2/3 in silico methods
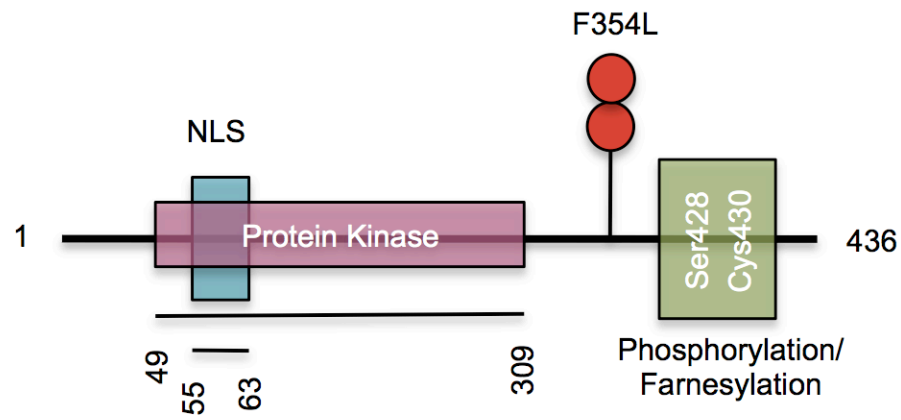
Figure 4.4 Rare, pathogenic missense mutation (F354L) in *STK11* that has previously been associated with Peutz-Jeghers disease was identified in two, unrelated familial gastric cancer families using TSCA Assay.
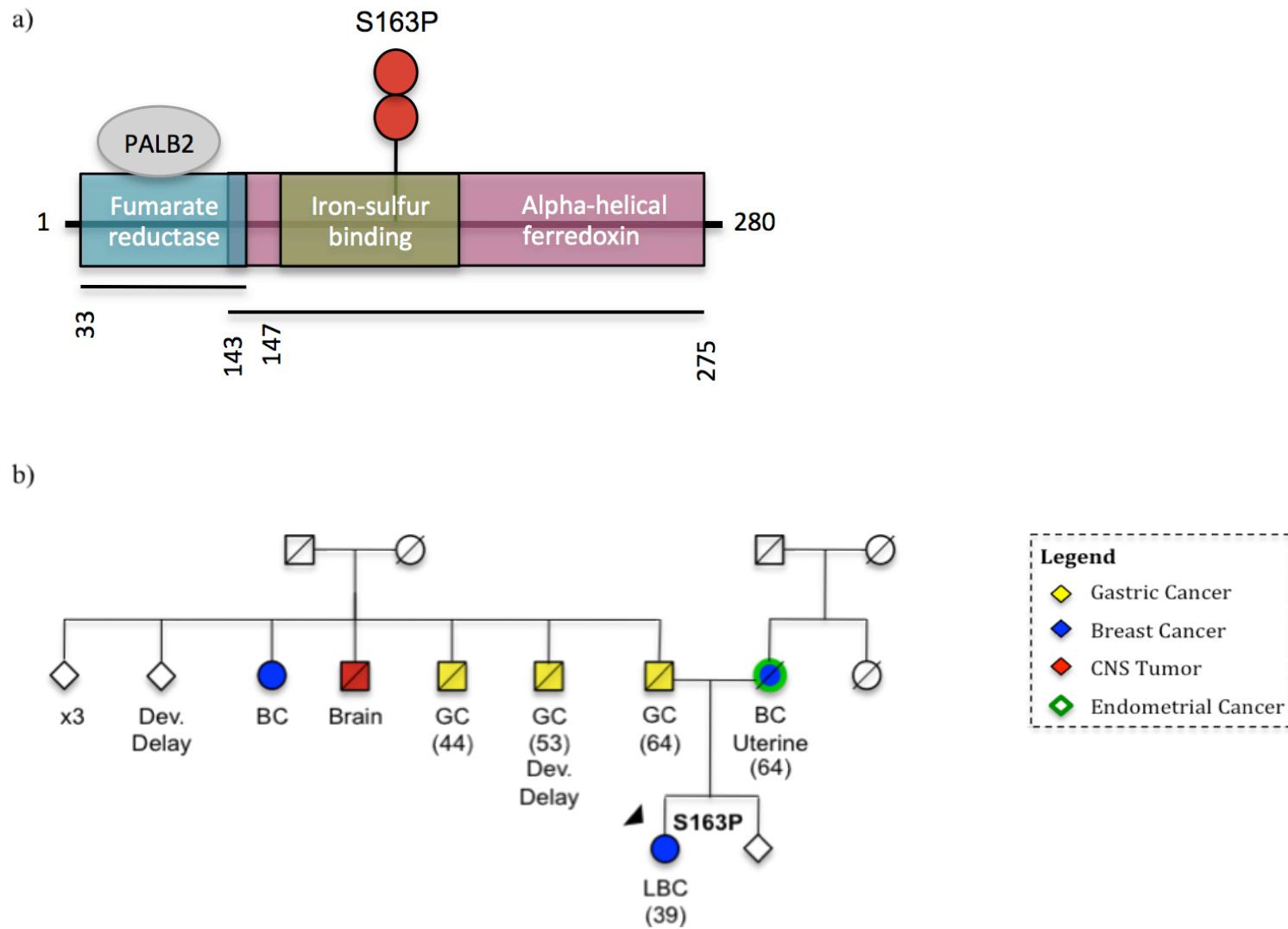
Figure 4.5 Rare, pathogenic missense mutation (S163P) in *SDHB* that has previously been associated with Cowden-like syndrome was identified in two, unrelated familial gastric cancer families using TuSeq Custom Amplicon Assay.
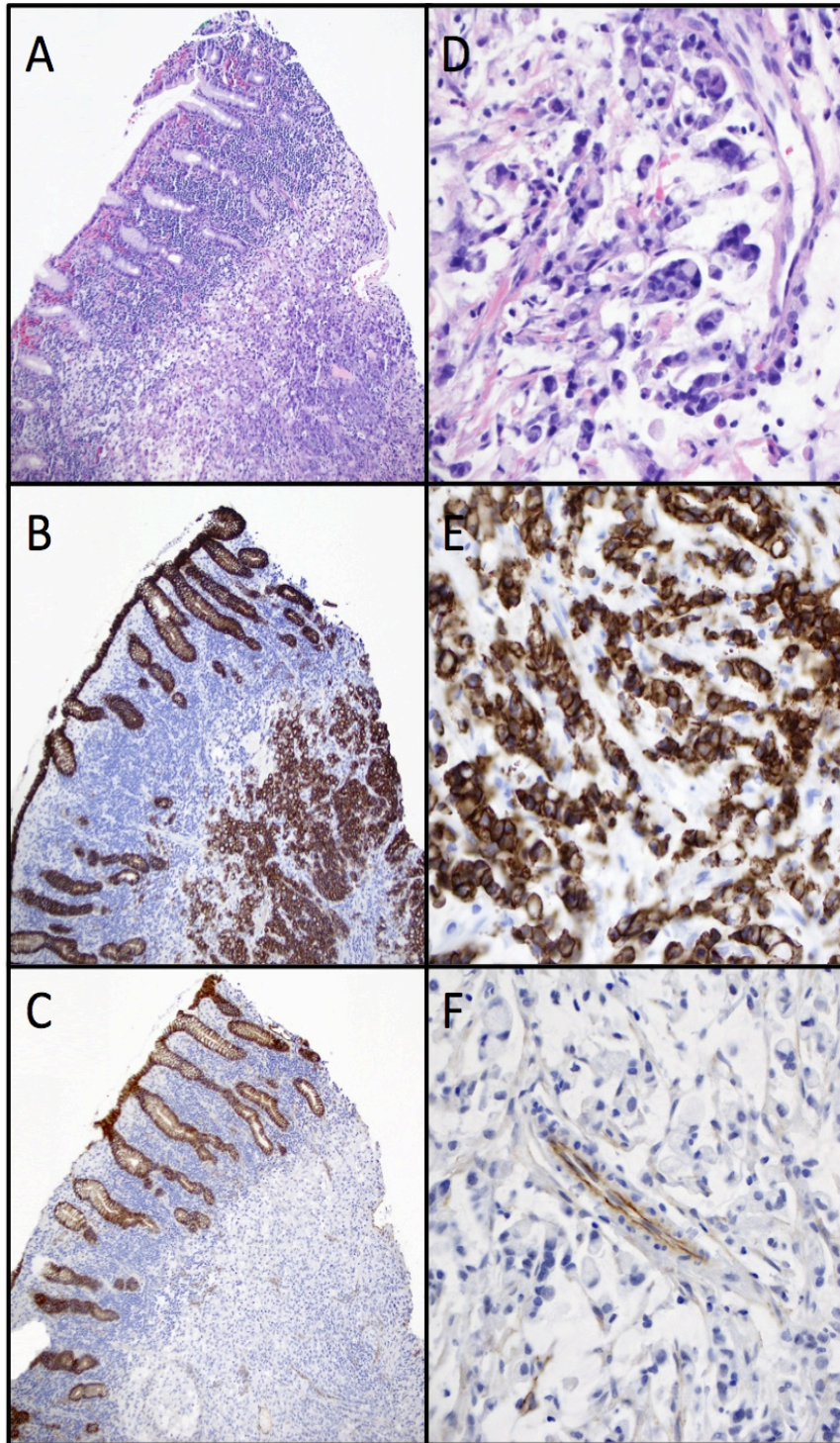
Figure 4.6 Immunohistochemical analysis of tumour material of germline *CTNNA1* truncating variant carrier (Family 25). Panels A&D show H&E staining to decipher tumour and normal cells. Panels B&E show retention of e-cadherin protein expression in both normal and tumour material. Panels C&F demonstrate loss of alpha-catennin expression in tumour material only, suggesting loss of heterozygosity at the *CTNNA1* locus.
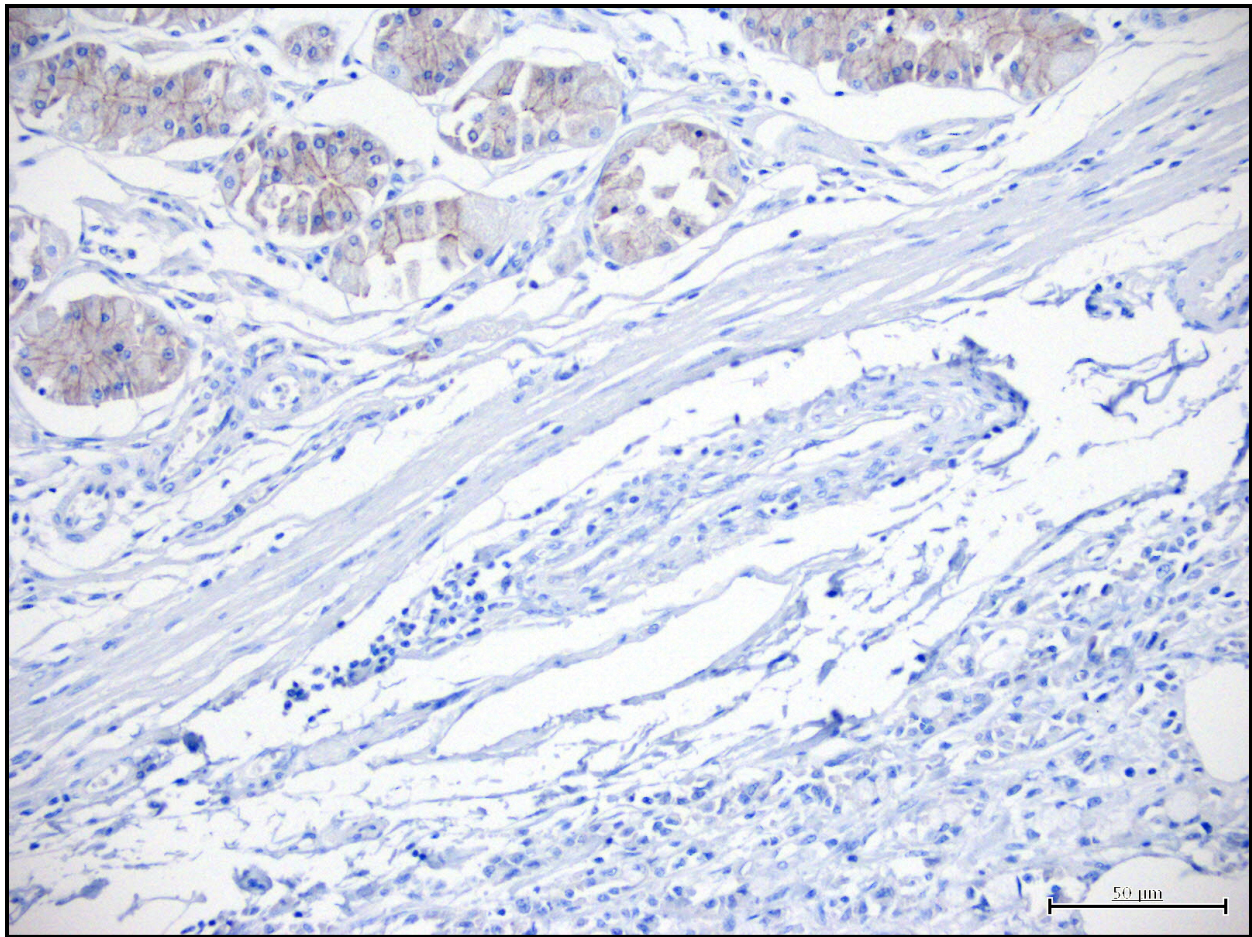
Figure 4.7 I**mmunohistochemical staining of tumour material of germline *CTNNA1* truncating variant carrier (Family 80).** This HDGC family does not have a germline *CDH1* variant and was found to have a germline, truncating variant in *CTNNA1* by custom-panel sequencing in this study. IHC demonstrate loss of alpha-catenin protein expression in tumour cells

4.2 Likely Pathogenic Variants in Moderately-Penetrant Genes

Truncating variants within low to moderately-penetrant genes were identified in 5 HDGC and 2 FIGC families (**Table 4.1**). Heterozygous frameshift variants were identified in the DNA-mismatch repair gene *ATM* (E1267fs (1 HDGC and 1 FIGC), Y2791fs and R521fs) (**Figure 4.8**), *MSR1* (R293X (n=2)) and *PALB2* (V398fs, 1 FIGC) (**Figure 4.9**, **Table 1**). Each of these genes have been implicated in the development of upper GI diseases (including GC) but carry mutations with moderate penetrance (Erkko et al. 2008; Byrnes et al. 2008; Maier et al. 2006; Orloff et al. 2011). *ATM* variant E1267fs found in Family 104 was confirmed in one relative with IGC diagnosed at age 57 and two unaffected offspring of the proband (**Figure 4.8-b**), supporting the likelihood that this truncating variant is of moderate penetrance. Somatic analyses of tumor DNA of both proband and affected relative revealed both mutant and wild-type alleles, suggesting that there is no LOH in the cases. IHC of tumour material was not available.

Additional family members were not available for other families with pathogenic or likely pathogenic mutations at this time as communicating the causality of these germline mutations to genetic counselors and the families in question proved difficult (**Chapter 6: Limitations**). It is likely that these mutations are associated with familial aggregation of GC, but including these genes in further screening of additional FGC cases with available tumour materials is necessary to prove their direct relationship to FGC (**Chapter 6: Future Directions**). Overall, limitation of sample availability prevented confirmation of the identified variant among additional family members for other likely pathogenic truncating variants.
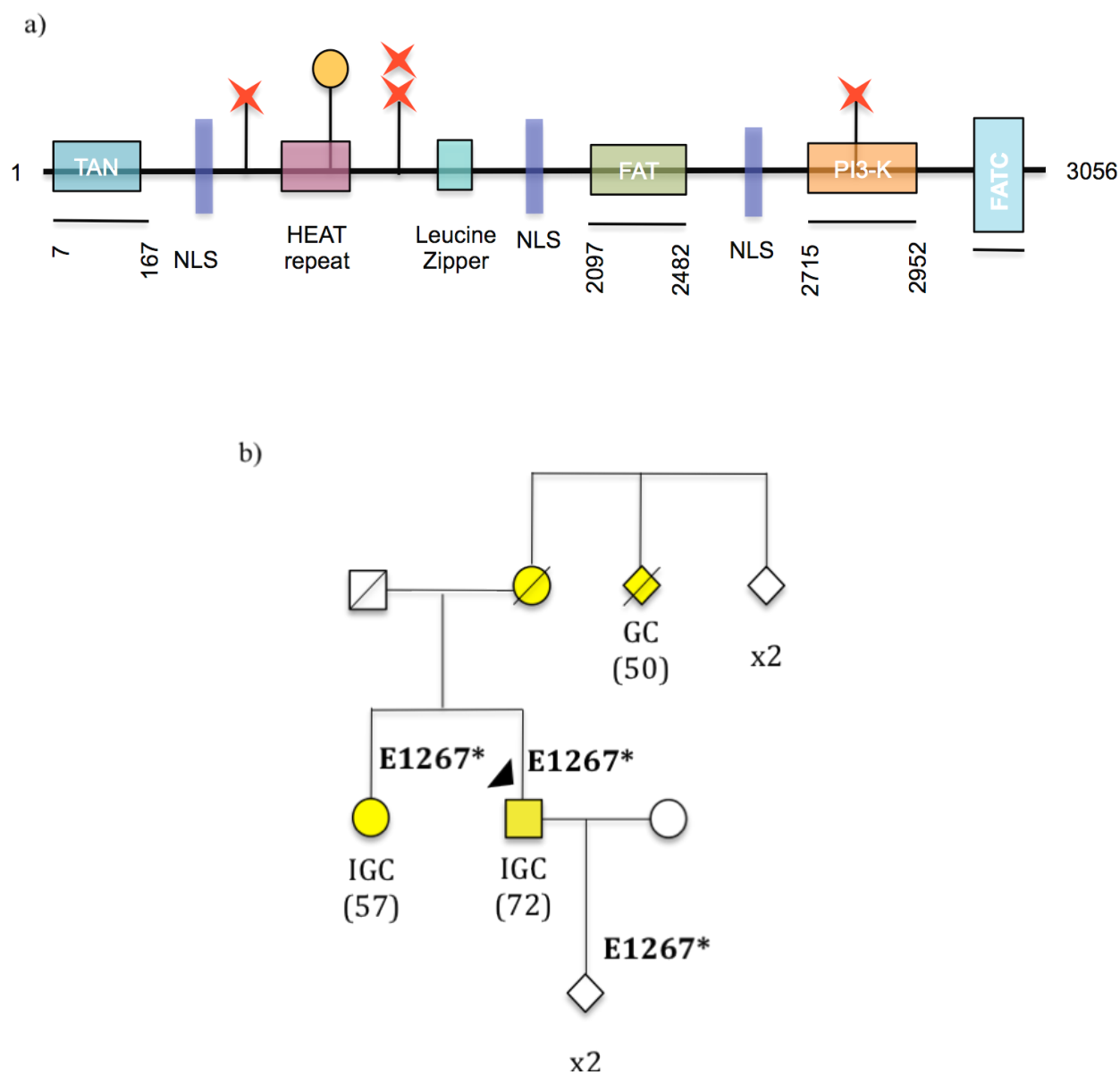
Figure 4.8 Novel, likely pathogenic mutations identified in *ATM*. a) A schematic representation of the *ATM* gene, including key domains. Red stars indicate novel truncating mutations aidentified in this study. A novel missense mutation predicted to be damaging via *in silico* methods is indicated as orange circle.
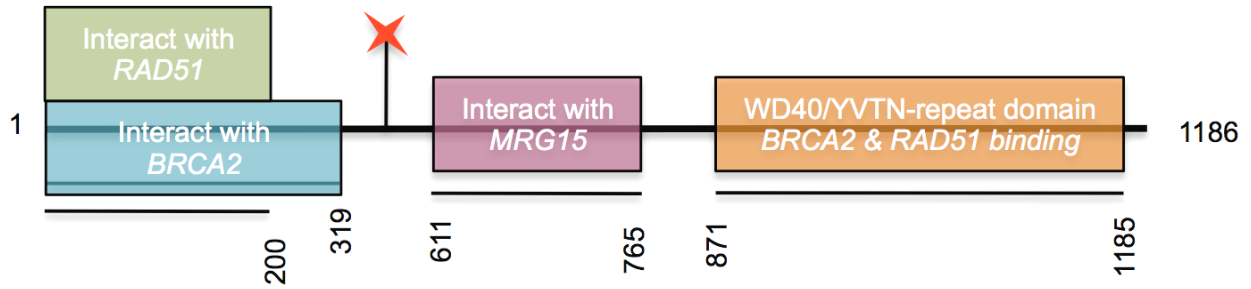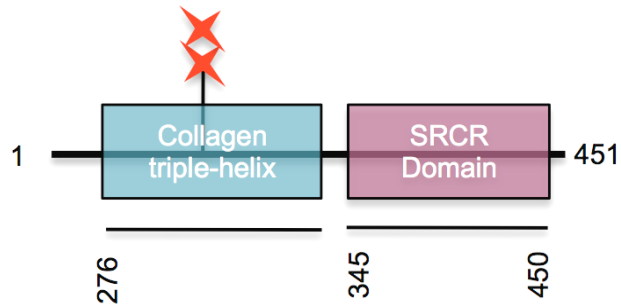
Figure 4.9 Schematic view of genes (a) *PALB2* and (b) *MSR1* and variants identified in this study. Likely pathogenic truncating mutations (red stars) identified in unrelated FGC cases using panel-based sequencing.

4.3 Variants of Unknown Significance

It is often difficult to prove pathogenicity of germline missense variants with hereditary syndromes, as often a single amino acid change is not sufficient enough to cause disruption of protein function. However, *in silico* methods are able to predict the damaging effect an amino acid substitution can have on the overall function of a protein.

Rare (<1% general population) or novel missense VUS were identified in 47 probands out of the remaining germline cases without candidate pathogenic or likely pathogenic variants. *In silico* methods PROVEAN, SIFT and PolyPhen (described previously in **Chapter 3**) were used to predict the functional impact of amino acid change in each of these missense mutations. Overall, 27 variants in individual germline cases (23.5%) were predicted as damaging in at least two *in silico* methods (**Table 4.2**). Some of the genes with these predicted-damaging missense mutations occurred in genes with pathogenic or likely pathogenic mutations in other GC families (*BRCA2* and *ATM*). This gives further support for the pathogenicity these potential GC susceptibility genes and further reason to include them in screening of prospective families.

4.4 Sequencing of Tumours from Familial Gastric Cancer Cases

Tumour DNA was available from 46 intestinal GC and 18 DGC cases from our Italian cohort. Of the DGC cases, data revealed two novel *ARID1A* somatic truncating mutations in two unrelated families, one of which also contained a somatic *MSH3* truncating variant. Sequencing of FIGC tumours revealed a greater number of truncating variants, including those in the genes *ARID1A* (8 mutations in 7 cases), *APC* (4 mutations in 3 cases), *CTNNA1* (2 mutations in 2 cases), *PTEN* (3 mutations in 3 cases), *TP53* (3 mutations in 3 cases), *CFTR* (2 mutations in 2 cases), *BRCA2*

(1 mutation), *PMS1* (1 mutation) and *MSH2* (1 mutation). All variants presented with good coverage across the corresponding targeted regions.

4.5 Candidate Variants Detected in Other Familial Upper GI Cancers

Our screen of FGC germline cases was part of a larger collaboration involving the identification of new susceptibility genes to familial upper GI cancers, including esophageal, pancreatic, and G-E junction cancers (**Chapter 3**). A total of 304 families meeting criteria for familial upper GI cancers were screened against a custom panel of genes with previous association to familial upper GI cancers or cancer-risk syndromes (**Table 3.2**). Of the 304-germline samples, 48 familial EAC or BE cases from 17 families were screened and no identifiable candidate variants were detected. Some novel missense mutations common among affected family members were identified and predicted as pathogenic via *in silico* methods; however, functional impact via *in vitro* and *in vivo* methods is necessary before such SNV can be declared pathogenic.

Forty-two familial pancreatic cases with unexplained hereditary patterns of disease were also included in the custom panel based sequencing (n=29 and n=13 from collaborators in Mount Sinai and McGill collaborating institutes, respectively) (**Table 3.2**). Novel, heterozygous truncating mutations in the form of nonsense variants were identified in two cases within the hereditary pancreatitis susceptibility gene *CFTR* and a new candidate susceptibility gene *MAP3K6*. Mutations validated via Sanger sequencing and raw data was sent to corresponding collaborative institutes for secondary validation and downstream analysis.

Chapter 5: Discussion & Conclusions

This data demonstrates an improved ability to detect pathogenic variants in families with unexplained FGC by using a custom-designed, multiplexed amplicon-based NGS assay. We identified novel and rare genetic risk factors associated with GC or GC-risk syndromes in 13.0% of the probands screened, all of whom who have been diagnosed with and have a family history of GC. Of particular interest are the two families with protein truncating mutations in the adhesion-complex molecule alpha-E-catenin, the first novel *BRCA2* mutations described in clinically defined FGC families and the novel findings of germline variants that may potentially cause FIGC in 4 out of the 9 families analyzed. Additional variants of interest in low-moderately penetrant genes (*ATM, PALB2, MSR1*) were also discovered in both HDGC and FIGC cases. This data suggests the refining of FGC criteria and proposes that true HDGC is a genetically, not clinically, defined disease. By targeting genes with known diagnostic value, multiplexed panel sequencing limits total test output compared to whole-genome or exome approaches, while simultaneously providing efficiency advantages over traditional, low-throughput single-gene testing methods (**Figure 4**) (Bybee et al. 2011).

5.1 Alpha-E-catenin (*CTNNA1*) and Hereditary Diffuse Gastric Cancer

Of the most interesting findings in this data set were those solidifying the involvement of additional genes in the same cellular-adhesion complex as e-cadherin in some unexplained cases of HDGC. Alpha-E-catenin, encoded by the 906 amino acid gene *CTNNA1,* is a 102kDa protein involved in the cell-adhesion complex to help facilitate adhesion and communication between neighboring epithelial cells. As previously mentioned, alpha-E-catenin molecules directly

regulate actin-filament assembly and organization with cell-adhesion complex (Gall et al. 2013) (Oliveira et al. 2013) (**Figure 1.1**). Specifically, beta-catenin protein binds directly to the cytoplasmic domain of e-cadherin while alpha-catenin acts as an intermediate to secure this complex to the actin cytoskeleton. This relationship between cadherin-catenin molecules and the adhesion complex as a whole is crucial for the development and maintenance of cellular linkage and tumour suppression in epithelial tissues (Conacci-Sorrell et al. 2002). Loss of intercellular adhesion is one of the earliest steps in epithelial tumour invasion and metastasis so it is no surprise that loss of e-cadherin expression is found in the majority of tumours from GC families with germline variants at the *CDH1*-locus. However, some families without detectable germline *CDH1*-mutations retain e-cadherin protein expression in tumours. It has therefore reasonably been suggested that other genes in the cadherin-catenin protein complex may play a pivotal role in the development of epithelial-type cancers, such as GC and lobular breast cancers.

Abnormal or lost expression of other proteins involved in cellular adhesion have long been reported in cancers of the stomach, breast, colon, pancreas, bladder, and prostate and correlate with advanced stage and high grade (Oka et al. 1993; Shimoyama et al. 1991; Matsuura et al. 1991; Nigam et al. 1993; Li et al. 2003; Bringuier et al. 1993; Umbas et al 1993). With respect to gastric adenocarcinomas, studies show that loss of e-cadherin through mutations or epigenetic alterations occurs more frequently in diffuse rather than IGC (Joo et al. 2002). Though germline or somatic mutations in alpha-catenin are not yet convincingly demonstrated in GC tumours, loss of protein expression in both diffuse and intestinal GC is frequent (Oka et al 1993; Joo et al. 2002). Interestingly, in sporadic GC, *CDH1* and *CTNNA1* genes are somatically mutated or altered in 14.6% of cases and mutations are mutually exclusive, though not significant (p-value=

0.296694) (*cbioportal*). When specifying histotype to sporadic DGC, this number increases to 22% of cases (*cbioportal*).

Recently, ES was used to identify the first novel, germline truncating variant in *CTNNA1* a HDGC family without a *CDH1* mutation (Majewski et al. 2013). Familial segregation and loss of protein expression in tumour material supported this heritable variant as pathogenic. Despite the lack of an identifiable somatic variant on the second allele, it was suggested that epigenetic modifications may be at play to lead to loss of heterozygosity in the tumour tissues of this family. Epigenetic modifications, such as promoter hypermethylation, occur in approximately 32% of DGC tumours of germline *CDH1* mutation carriers to completely disrupt protein development and function (Oliveira et al. 2009 (2)), so it is likely that similar events may accompany germline mutations at the *CTNNA1*-locus. This article officially marked *CTNNA1* as a probable susceptibility gene to HDGC (Majewski et al. 2013). However, it was concluded that further studies on *CTNNA1* with respect to FGC, specifically the identification of additional families with germline modifications, are needed to solidify the inclusion of *CTNNA1* in future screening of GC families. This was the first report of a pathogenic variant outside of the *CDH1*-locus in a HDGC family and provided evidence that additional genes involved in the cellular adhesion pathway may increase susceptibility to GC.

Two additional HDGC families have been identified using this multiplex, custom-gene panel approach with truncating mutations at the *CTNNA1*-locus (**Figure 4.2**) (Hansford et al. 2014, manuscript in preparation). Family 25 was found to carry a novel frameshift mutation in *CTNNA1* (N71fs), leading to a truncated protein and subsequent disruption of alpha-E-catenin properties

(**Figure 4.2-a**). We showed that this variant segregates with the affected mother of the proband (**Figure 4.2-b**), further supporting it as a highly penetrant, pathogenic mutation and one that likely increases the susceptibility to GC in this family. Family 80 was found to carry a heterozygous, germline nonsense mutation in *CTNNA1* (R129X). Nonsense mutations lead to a premature stop codon and, ultimately, protein truncation and have been highlighted as pathogenic in many genes, including *CDH1* with respect to HDGC. Additional family members were not available for genotyping as the only other known affected individual was deceased. An extensive pedigree was not available from the initiating institution to confirm if additional diagnoses within the family exist.

IHC analysis of tumor tissues from both probands with germline *CTNNA1* truncating variants revealed loss of alpha-catenin protein expression and retention of E-cadherin, suggesting loss of the *CTNNA1* wild-type allele by a second-hit mechanism (**Figure 4.6 & 4.7**). Abnormal, cytoplasmic expression of e-cadherin can also be seen (**Figure 4.6, Panel E**), suggesting truncating *CTNNA1* mutation may disrupt normal cellular distribution of other adhesion molecules, such as interfering with the intermolecular binding with cytoplasmic anchors to e-cadherin. Tumor DNA was available from the proband in Family 80 who carry the R129X variant, but second-hit analysis did not reveal a somatic mutation at the *CTNNA1* locus. It is proposed that epigenetic factors may be disrupting the functionality of the second *CTNNA1* allele, suggested similarly in a previous study (Majewski et al. 2013).

Our findings of 2 additional FGC cases with germline truncating *CTNNA1* variants, coupled with the previous report of a HDGC family with *CTNNA1* variant, as well as the role of alpha-catenin in the e-cadherin cellular adhesion pathway, strongly implicate *CTNNA1* as a GC susceptibility gene. Differential expression of cadherin-catenin complex components is also observed in LBC, which is prominent in families with HDGC and reinforces the role of *CTNNA1* in HDGC (Morrogh et al. 2011; Nakopoulou et al. 2002; Park et al. 2007; Koslov et al. 1997). This evidence strongly suggests the necessity to include *CTNNA1* in the screening of prospective HDGC families.

5.2 *BRCA2* and Hereditary Diffuse Gastric Cancer

Germline mutations in the DNA-mismatch repair gene *BRCA2* are responsible for approximately 20-25% and 15% of hereditary breast and ovarian cancers (Ferla, R et al. 2007; Janavicius, R. 2010; Tulinus, H et al. 2002). This gene plays an important role in DNA-repair process through transcriptional regulation of genes involved in the cell cycle, apoptosis and those in response to DNA damage (Yoshida et al. 2004). Some regulatory proteins, such as *ATM, CHK2, ATR, RAD51*, directly interact with specific regions of the *BRCA2* locus and regulate DNA-repair (Yoshida et al. 2004). Disruption of the interactions amongst *BRCA2* and these proteins lead to increased sensitivity to ionizing radiation and accumulation of chromosomal breaks *in vitro* (Yoshia et al. 2004). Loss of *BRCA2* has been shown to strongly contribute to cancer initiation and progression through the accumulation of such DNA-damage mechanisms. Specifically, *BRCA2* is responsible for regulation the activity of *RAD51*, which in turn controls homologous recombination (Marmorstein, Ouchi & Aaronson, 1998). Germline mutations within this gene are now widely recognized contributors to cancer susceptibility; however, the overall penetrance

of mutations varies significantly. Depending on location of the mutation, female carriers are found to be at a 25-80% increased risk for breast cancer and 10-40% lifetime risk for ovarian cancer (Ferla, R et al. 2007; Janavicius, R. 2010; Tulinus, H et al. 2002). As well, Men found to carry *BRCA1/2* mutations have been shown to be at a 30-65% risk for developing prostate cancer (Ferla, R et al. 2007; Janavicius, R. 2010; Tulinus, H et al. 2002; Kote-Jarai, Z et al. 2011). Familial ethnicity and position of mutation have been found to contribute to the phenotypic variation observed in *BRCA2*-postive families. For example, higher penetrant mutations for cancers other than breast have been found in the ovarian cancer cluster region (OCCR) in exon 11 of *BRCA2* (Jakubowska et al. 2002; Risch et al. 2001; Moran et al. 2012).

Families that show strong hereditary patterns of breast/ovarian cancers are offered genetic testing to screen for germline mutations within *BRCA2* and find out if they are at an increased risk for disease. For mutation positive families, management guidelines have been set for unaffected mutation carriers, such as routine screening, mastectomy or hysterectomy surgeries (Wainberg & Husted, 2004; Narod & Offit, 2005; McKinnon et al. 2007; Ingham et al. 2013).

With the strong correlation of *BRCA2* mutations to hereditary breast and ovarian cancers, it has long been suspected that mutations may lead to an increased risk for other cancers as well. In 1999, it was demonstrated that cancer in areas other than the breast and ovaries are over-represented in *BRCA2* mutation-positive families (Johannsson et al. 1999). For example, Jewish descendents that carry the *BRCA2* founder mutation 6174delT are at a 5.7% increased risk for the development of GC (five times greater than the general population) (Figer et al. 2001). Another study sought to uncover the frequency of *BRCA2* germline mutations in families that meet criteria for *BRCA1/2* testing but have a clear aggregation of both breast and GC (unspecific

histology) (Jakubowska et al. 2002). In addition, 20.7% of families with at least one diagnosis of GC before the age of 50 were found to carry *BRCA2* variants, including three protein truncating and three potentially pathogenic missense mutations. Though lack of available specimens prevented confirmation of these *BRCA2* mutation within GC patients, they concluded that aggregations of both stomach and breast cancer can be used as a phenotypic indicator for the pre-selection of families for *BRCA2* testing (Jakubowska et al. 2002). Familial aggregates of LBC in HDGC families with *CDH1* mutations further supports the relationship between gastric and breast cancer and the likelihood that BRCA2 mutation may be found in some clinically defined GC families.

Our data provides further evidence that germline-mutations at the *BRCA2*-locus can be a genetic risk factor for GC, as we identified two novel truncating variants (N1287fs and K936fs) in 2 families who met IGCLC criteria for HDGC and FIGC. Family 2 (N1287fs) has an autosomal dominant inheritance pattern of both DGC and breast cancer (**Table 1**; **Figure 4.3-b**). Complete pedigree for Family 109 was not available, but does meet IGCLC criteria for FIGC (Caldas et al. 1999) as confirmed by the initiating institution. No *CDH1* germline variant was identified in either case, but somatic *CDH1*-LOH was reported for Family 109 through IHC analysis (Corso et al. 2013). Tumor DNA of Family 109 did not reveal a somatic second-hit in the *BRCA2*-locus but did reveal a somatic nonsense mutation in the gene *APC*. Neither tumor material nor additional family members were available for downstream analyses, but the pathogenic nature of these mutations is supported by the fact that both truncating variants (N1287fs and K936fs) occur in the ovarian cancer cluster region of the *BRCA2* locus. Higher penetrance mutations are more associated with this region of *BRCA2* and increases risk for cancers other than breast, such

as ovarian, colorectal, prostate, pancreatic and stomach (Jakubowska et al. 2002; Risch et al. 2001; Moran et al. 2012). Despite this continued support of the familial relationship between breast and GC, this is the first report of *BRCA2* mutations being described in FGC cases and *BRCA2* is not included in genetic screening for such families. It is likely that *BRCA2* mutations may explain a rare portion of FGC cases and that current management strategies for *BRCA2* carriers should implemented in these families.

5.3 Previously Reported Pathogenic Missense Mutations Detected in Familial Gastric Cancers

STK11 (*LKB1*) encodes a serine-threonine kinase involved in maintaining cell-cycle function, metabolism and cellular polarity (Forcet et al. 2005). Heterozygous germline mutations in *STK11* are associated with Peutz-Jegher's syndrome (PJS), an autosomal dominant syndrome characterized by oral pigmentation and GI hamartomas (Forcet et al. 2005; Yoon et al 2000). Clinically defined PJS patients are at a significantly increased risk (47% by age 65) for the development of GI malignancies and GC is found to be the third most common malignancy among these patients (Chun et al. 2012; Volikos et al. 2006). The relative risk for upper GI cancers in PJS patients ranges from 85, 132, 213 and over 500 for colon, pancreatic stomach and small intestine cancers, respectively (Chun et al. 2012; Giardiello et al. 2000) . *STK11/LKB1* is currently the only known susceptibility gene to PJS and mutations have been described in 30-80% of cases, of which >50% have a personal family history of PJS (Hearle et al. 2006 (1)). Previous studies describe the highly-penetrant nature of germline *STK11* mutations and concluded that though there was not a significant increased risk for cancer among germline

mutation carriers, there is for PJS, which carries a predisposition for various cancers of up to 75% by age 70 (Hearle et al. 2006 (2)).

The missense variant F354L was first identified in a PJS patient and showed inheritance by the asymptomatic mother, suggesting moderate penetrance (Forcet et al. 2005). Though mutations within the kinase domain disrupt enzymatic activity and lead to neoplastic progression, *in vitro* analyses on mutations in the C-terminus end, such as F354L, show impaired *STK11*-mediated activation of AMPK pathways and disruption cellular polarity of intestinal epithelial cells (Forcet et al 2005). We have identified the F354L missense mutation (**Figure 4.4-a**) in two HDGC families that lack *CDH1* mutations. Family 44 presented with 2 cases of GC (diffuse histology confirmed in one) as well as breast, colorectal and esophageal cancers (**Table 1**). Family 46 had a single case of DGC at age 22. It is unknown if members of either family had other features of PJS. Additional family members or tumour materials were not available for segregation analysis amongst these families as the pathogenic nature of these variants is not yet definitive. Communicating such data to genetic counselors (and ultimately collecting additional samples from the families in question) proves difficult, as there are not yet risk stratification procedures in place. If additional FGC cases are identified with predicted pathogenic mutations in the *STK11* locus, it is likely that support of penetrance analysis would be an effective step in identifying the GC risk for mutation carriers (**Chapter 6**).

Cowden-like syndrome (CLS) is a phenotypically mild variant of Cowden syndrome (CS), which often presents with GI hamartomas as well as cancers of the breast, thyroid, colon, genitourinatry. Typically, germline mutations within the phosphotase tensin homolog (*PTEN)*

locus are described in >80% of CS families (Orloff et al, 2013; Marsh et al. 1998). However, homozygous germline mutations in succinate dehydrogenase-B subunit gene (*SDHB*) have been shown to cause CLS along with severe neurological dysfunction, heterozygous mutations can lead to hereditary leiomyomatosis and renal cell carcinoma (Ni et al. 2008). Succinate dehydrogenase participates in both the Krebs cycle and electron transport chain of cellular metabolism (Ni et al. 2008). Recently, a heterozygous germline polymorphism (S163P) in the gene *SDHB*, reported at a 2% frequency among African American populations and 0% in other ethnicities, was detected in two CLS patients (but not 700 controls) that had a history of thyroid carcinoma, benign uterine pathologies and family histories of breast cancer and papillary thyroid carcinoma (Ni et al. 2008). The variant was supported as pathogenic through *in vitro* assays resulting in mimicking *PTEN* abnormalities through downstream activation of AFT and MAPK pathways and characterized as a susceptibility polymorphism for CLS (Ni et al. 2008*)*. Our panel has identified 2 FGC cases with this rare *SDHB* variant, S163P (**Figure 4.5**). Broader analysis of Family 13 history revealed patterns of CLS, including incidences of breast, uterine, and GI cancers, as well as developmental/neurological delay observed in some family members (**Table 1**; **Figure 4.5-b**). This rare, pathogenic variant was also seen in germline DNA from a FIGC family, but it is unknown if this family had an additional history suggestive of other cancers or cancer-predisposition syndromes. Similar to the rare, pathogenic missense variant in *STK11* identified in FGC cases, as there is no risk-stratification for *SDHB* missense mutations and CLS families, it was difficult to attain additional family members for analysis. The identification of *SDHB* mutations in additional families with a history of both GC and CLS-phenotypes would further support its relationship to GC-risk and the necessity broaden the genetic screening of prospective GC families outside the *CDH1*-locus.

5.4 Likely Pathogenic Variants Identified in Moderately-Penetrant Genes

Truncating mutations in low-to-moderate penetrant genes *ATM, PALB2 and MSR1* (**Table 1**; **Figure 4.8; Figure 4.9)** were identified through TSCA gene panel. Additional germline samples for other family members and tumor material were difficult to obtain for the aforementioned families, as rationalizing interventions for mutation carriers was difficult to communicate to genetic counselors due to the low observed penetrance in previous studies. Further analyses of the relation to GC should be completed to fully support cancer susceptibility risk of these genes.

Ataxia telangiectasia-mutated (*ATM*) gene encodes a 350kDA kinase that is a crucial signaling molecule involved in DNA double-stranded break-repair, activation of cell-cycle checkpoints and induction of apoptosis (Shen et al. 2012; Zhang et al. 2004). This large gene composed of 66 exons and contains several key regulatory and binding domains that are critical for its function. Of importance is the ~400 amino acid C-terminal end that is highly similar to the catalytic subunit of phosphatidylinositol 3-kinase (PI-3-Kinase), which is important for progression through the cell-cycle and genomic stability (Rotman & Shiloh, 1998).

ATM has been found to be responsible for Ataxia Telangiectasia (A-T) syndrome, a rare autosomal recessive inherited syndrome characterized by immune deficiency, progressive dysfunction of the cerebellum that causes loss of coordination and increased risk for cancers (Savitsky et al. 1995). Cancers associated with this recessive syndrome are caused by lack of DNA double-stranded break repair and cell cycle arrest, to which *ATM* plays a significant role (Shiloh et al. 2003; Zhang et al. 2004). The inheritance of A-T is autosomal recessive, whereby each parent must be a carrier of a defective/mutated allele on the *ATM* locus for the offspring to be homozygous, and therefore affected. Heterozygous carriers of pathogenic *ATM* mutations,

though virtually not at risk of developing A-T, have been shown to be susceptible to certain

cancers, such as breast, pancreatic, glioma, lymphoma and lung cancers (Shen et al. 2012).

Interestingly, both gastric and colon cancer cell lines have also shown a high frequency of

variants within coding regions of *ATM* and GC tumours of unspecific histology have also shown

to have reduced or lost expression of *ATM* protein in tumour cells only (Zhang et al. 2004;

Ejima, Yang & Sasaki. 2000). *In vitro* assays of *ATM*-deficient cells show they are sensitive to

ionizing radiation and lack cell-cycle regularity after exposure to such radiation (Gilad et al.

1998). Though the evidence to suggest *ATM* holds significant tumour suppressor roles is

mounting, the penetrance data of germline heterozygous variants and increased susceptibility to

cancers varies significantly. Some reports suggest that carriers of germline, truncating mutations

at the *ATM* locus are at high susceptibility to breast cancers (Bernstein et al. 2006), yet other

describe pathogenicity of *ATM* variants through dominant-negative mutations, whereby

truncating variants act in an antagonist fashion to the normal gene activity (Chenevix-Trench et

al. 2002). The estimated average penetrance of truncating *ATM* mutations identified in three

hereditary breast cancer families was 60% by age 70 (Chenevix-Trench et al. 2002). One of the

families indicated in this study also had a family history of GC, without confirmed histology, but

analysis was not completed upon lack of available genomic DNA. Large epidemiological

studies examining the relationship between germline *ATM* variants and increased risk for breast

cancer among female carriers show statistical significance, with increased risk of approximately

2 to 5-fold compared to the general population (Renwick et al, 2006; Thompson et al, 2005).

However, other studies describe low penetrance of heterozygous germline variants, at

approximately 15% (Apostolou & Fostira, 2013). Though there is evidence of significant

increased risk amongst heterozygous carriers in some studies, there is lack of definitive and

consistent penetrance data across the wide spectrum of mutations across the gene. This demonstrates the difficulty of assessing the possible clinical utility of *ATM* genetic screening for prospective hereditary cancer families.

A recent study has shown that loss of *ATM* expression coupled with low microsatellite instability may be a prognostic marker for gastric adenocarcinoma without restriction to histotype, though family history of participants was not known or reported (Kim et al. 2014). Low *ATM* expression in GC has also been shown to be associated with higher stage disease and lymph node metastasis (Kim et al. 2014; Kang et al. 2008). Despite definitive proof that *ATM* expression is lost in some gastric adenocarcinomas, no germline mutations have been identified to explain this phenomenon. We have identified four novel truncating mutations at the *ATM* locus in families that meet clinical criteria for HDGC or FIGC (**Table 4.1**; **Figure 4.8**). Interestingly, the same *ATM* truncating mutation (E1267fs) was found in two cases: a Canadian HDGC family (unknown ethnicity) and a Portuguese FIGC family (**Table 1**). It is not known is these truncating variants occur in different kindred from separate events of if they families have deviated from common ancestry. Screening among one family with the truncating variant (E1267fs) supports the moderately penetrant nature of *ATM* as it was identified in two affected and two unaffected individuals. As previously mentioned, IHC analysis of tumour material from a germline *ATM* truncating mutation carrier was inconclusive due to poor quality of antibody and the lack of available gastric tumour and normal material (**Chapter 3**). It is suspected that, like the relationship with hereditary breast cancer, these truncating mutations at the *ATM* locus are of moderate penetrance and likely increase the risk for GC in the mutation carriers.

Partner and localizer of *BRCA2* (*PALB2*) is a DNA repair protein involved in the maintenance of homologous recombination and double-stranded breaks. It was originally identified as a breast cancer susceptibility gene when a group was in search of novel components of the *BRCA2*, DNA-repair complex (Rahman et al. 2007). *PALB2* consists of 13 coding regions and functionally acts as a bridge between *BRCA1/2* proteins for double-stranded-break repair and homologous recombination (Fernandes et al. 2014). Homologous mutations across the *PALB2*-locus have been implicated in hereditary Fanconi anemia in a similar matter to *BRCA2* relationship to the same disease (Fernandes et al. 2014). Mono-allelic or heterozygous germline mutations in *PALB2* have been implicated in a rare number of hereditary breast and pancreatic cancers, though mutations confer with a moderate risk of disease (Fernandes et al. 2014). In one study, 12 highly pathogenic (truncating or splice site) *PALB2* mutations were identified in 1479 patients with breast cancer (0.8%) (Fernandes et al. 2014). The mutation prevalence for high-risk individuals (those with a strong family history of breast cancer) was 1.05% (95% CI = 0.5-1.92). Another study identified a 2.3 fold-increased risk for breast cancer among mono-allelic germline PALB2 mutation carriers (Rahman et al. 2014). In hereditary pancreatic families *PALB2* variants have been found in 3-4%, again with variable penetrance (Blanco et al. 2013; Slater et al. 2010; Jones et al. 2009). Somatic mutations in *PALB2* have only been reported in 1% of gastric adenocarcinomas (cbioportal.org) and until this study, *PALB2* mutations have not been implicated in clinically defined FGC. It is unknown how many hereditary breast cancer families with *PALB2* mutations had a history of GC, but in one study a *PALB2* truncating variant was identified in a breast cancer family with a reported history of pancreatic cancer (three counts) as well as infiltrating ductal carcinoma, skin, stomach and CNS cancers (Blanco et al. 2013). According to the pedigree, the GC occurred in the paternal uncle of the proband at age 84 with

unspecific histology but samples from these affected relatives were not available for screening (Blanco et al. 2013). In one other study focusing on specific hot spot regions across the *PALB2* locus, carriers were predicted to have a significantly higher risk for breast cancer (Southey et al. 2010). Of particular interest was the significant increased risk for breast cancer (up to 49%) in four families with the same heritable germline mutation (c.3113 G>A). Three of the four families carrying the variant had history of both breast and GC (unspecific histology). With these reports and others, it is likely that *PALB2* is a gene of moderate-penetrance with respect to increased susceptibility to breast and upper GI-related cancers.

We have identified the first clinically defined FGC case with a germline truncating *PALB2* variant (V398fs) (**Figure 4.9-a**). The proband of this family was diagnosed with IGC at the age of 45 and has an extended family history of GC with unspecific histology across first/second degree relatives at the ages of 60, 45, 73 and 50. There are also two reports of breast cancer in the family, unknown age of diagnosis as well as a history of colorectal and ovarian carcinomas. With the variable penetrance data reported across *PALB2* mutation carriers, it is likely that this truncating mutation is a disease-causing allele in this family. However, further work is needed before *PALB2* can be supported as a pathogenic susceptibility gene to GC and management strategies put in place. This further highlights the need to broaden the screening of FGC cases to genes outside of the *CDH1*-locus.

Macrophage scavenger receptor-1 (*MSR1*) is an integral membrane glycoprotein important for many macrophage and hormonal-associated processes such as inflammation, immunity (innate and adaptive), oxidative stress and apoptosis (Orloff et al. 2011; Xu et al. 2002; Xu et al. 2006).

Though risk-association and penetrance is controversial, a specific truncating mutation within the *MSR1* locus (R293X) has been previously related to prostate cancer and EAC in specific ancestries (Xu et al. 2002; Wang et al. 2003; Orloff et al. 2011). One study suggested that *MSR1*, *ASCC1* and *CTHRC1* are candidate familial esophageal cancer genes, but further work is needed to determine their contribution (Orloff et al. 2011). This particular variant (R293X) occurs in a conserves coiled-coil domain within the *MSR1* locus and near a critical glycosylation site at amino acid 249 (Orloff et al. 2011) (**Figure 4.9-b**). Truncation of *MSR1* at this location results in altered topology of both transmembrane and collagen-like motifs of the protein (Orloff et al. 2011).

This particular mutation in *MSR1* (R293X) was identified in two families of this study meeting clinical criteria for HDGC (**Table 4.1**). One family has an extensive family history of GC with 7 diagnoses among first and second-degree relatives between the ages of 47-87. This family also had two relatives diagnosed with BC at the ages of 42 and 78 with unspecific histology and a history of prostate cancer in two cases. The second family had a single known case of DGC at the age of 22. Though additional DNA from family members or tumour material were not available, the correlation between complex inflammatory processes and increased risk for GC further supports the possibility the *MSR1* may be a GC and other inherited neoplasia (such as prostate and EAC) susceptibility gene.

At this time, risk stratification and management is not available for carriers of likely pathogenic variants in low-moderately penetrant genes (such as *ATM*, *PALB2* or *MSR1*). However, carriers of these moderately penetrant alleles may benefit from advanced screening to detect GC at its

most early stage, should it develop. Identifying germline mutations in genes that likely predispose families to GC may have significant implications with respect to risk assessment, genetic counseling and targeted surveillance/screening and increase our understanding of all forms of GC.

5.5 Using Genetics to Change the Taxonomy of Familial Gastric Cancer

The findings from this project suggest that true hereditary DGC classification may be a separate entity from its FGC counterparts and should be a genetically, not clinically, defined disease (**Figure 5.1**). Currently, the clinical classification of FGC (HDGC, FIGC and familial syndromes that present with GC or phenocopies) is problematic. Currently, if a family that meets IGCLC criteria for HDGC has a *CDH1* mutation or not, they are still diagnosed as such (Fitzgerald et al. 2010). However, the only relevant data on the management of HDGC is for *CDH1* positive families. Secondly, we have identified genetic mutations within HDGC families that place such kindred into other well-known cancer susceptibility syndromes (**Figure 5.1-b**). Using this data, a more rational characterization system for FGC is plausible and can be based on genetic classification rather than clinical.

According to this proposed genetic classification of FGC, pathogenic mutations, large deletions, intronic-variants leading to reduced or eliminated *CDH1* expression and allelic imbalance, as well as (in rare cases) germline variants in other cellular-adhesion molecules (i.e. *CTNNA1*) should account for the mutational profile of true HDGC families (**Figure 5.1-c**). Currently, management strategies are available for families with pathogenic *CDH1* mutations (Fitzgerald et al. 2010) and research is underway to better understand the pathogenicity of *CDH1* non-coding mutations (**Chapter 6**). Risk-stratification methods and specific management strategies for
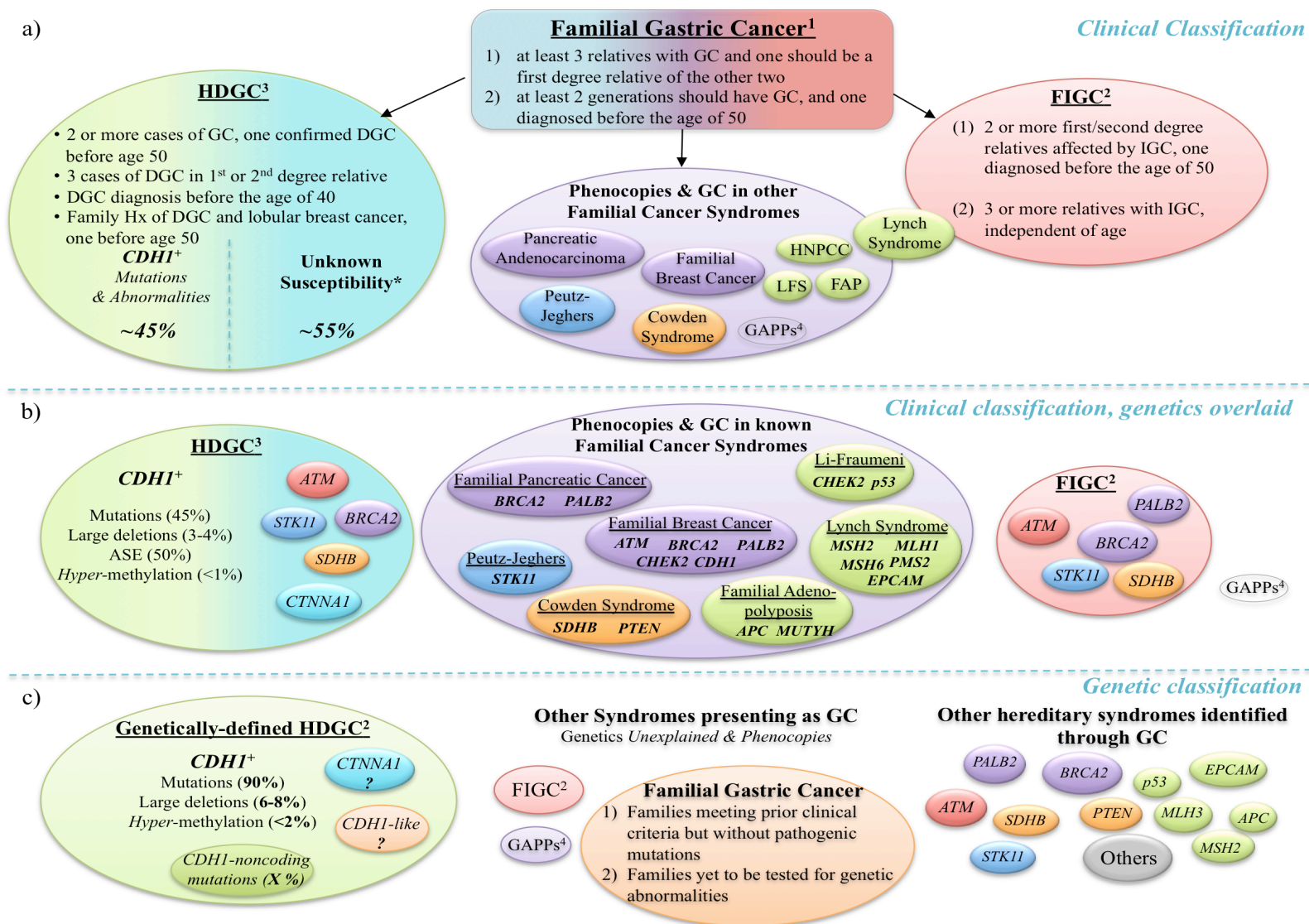
a) Clinical Classification

Familial Gastric Cancer[1]
1) at least 3 relatives with GC and one should be a first degree relative of the other two
2) at least 2 generations should have GC, and one diagnosed before the age of 50

HDGC[3]

- 2 or more cases of GC, one confirmed DGC before age 50
- 3 cases of DGC in 1st or 2nd degree relative
- DGC diagnosis before the age of 40
- Family Hx of DGC and lobular breast cancer, one before age 50

CDH1+
Mutations & Abnormalities
~45%

Unknown Susceptibility*
~55%

Phenocopies & GC in other Familial Cancer Syndromes
Pancreatic Andenocarcinoma
Familial Breast Cancer
HNPCC
Lynch Syndrome
LFS FAP
Peutz-Jeghers
Cowden Syndrome
GAPPs[4]

FIGC[2]
(1) 2 or more first/second degree relatives affected by IGC, one diagnosed before the age of 50
(2) 3 or more relatives with IGC, independent of age

b) Clinical classification, genetics overlaid

HDGC[3]

CDH1+
Mutations (45%)
Large deletions (3-4%)
ASE (50%)
Hyper-methylation (<1%)

ATM
STK11 BRCA2
SDHB
CTNNA1

Phenocopies & GC in known Familial Cancer Syndromes

Familial Pancreatic Cancer
BRCA2 PALB2

Li-Fraumeni
CHEK2 p53

Familial Breast Cancer
ATM BRCA2 PALB2
CHEK2 CDH1

Lynch Syndrome
MSH2 MLH1
MSH6 PMS2
EPCAM

Peutz-Jeghers
STK11

Familial Adeno-polyposis
APC MUTYH

Cowden Syndrome
SDHB PTEN

FIGC[2]
PALB2
ATM BRCA2
STK11 SDHB
GAPPs[4]

c) Genetic classification

Genetically-defined HDGC[2]

CDH1+
Mutations (**90%**)
Large deletions (**6-8%**)
Hyper-methylation (**<2%**)
CDH1-noncoding mutations (X %)

CTNNA1 ?
CDH1-like ?

Other Syndromes presenting as GC
Genetics *Unexplained & Phenocopies*

FIGC[2]
GAPPs[4]

Familial Gastric Cancer
1) Families meeting prior clinical criteria but without pathogenic mutations
2) Families yet to be tested for genetic abnormalities

Other hereditary syndromes identified through GC

PALB2
BRCA2 p53 EPCAM
ATM
SDHB PTEN MLH3 APC
MSH2
STK11 Others

Fig 5.1 Proposed new genetic classification of FGC and associated syndromes. a) Traditional clinical classification of FGC into HDGC, phenocopies or GC in known familial cancer syndromes and FIGC. b) Clinical classification of FGC with genetic overlay, including newly identified susceptibility genes to GC from this study. c) proposed genetic classification of FGC, suggestive that HDGC be a genetically defined disease through *CDH1* or *CDH1*-like (i.e. *CTNNA1*) mutations. [1]Shinmura et al. 1999; [2]Caldas et al. 1999; [3]Fitzgerald et al. 2010; Worthley et al. 2012[4]; *Allele Specific Expression presents in >70% of *CDH1*-negative cases

families *CDH1* or *CDH1*-like variants provide the best comprehensive understanding of HDGC as a whole. Hereditary cancer susceptibility syndromes with history of GC such as CS, PJS, familial adenoma-polyposis, GAPPS, Lynch syndrome and others may increase the occurrence of GC across relatives, leading to the family meeting clinically defined criteria for HDGC or FIGC (**Figure 5.1-a**). It is likely that mutations in genes outside of the cell-adhesion pathway, such as *BRCA2*, *ATM*, *MSR1*, *PALB2*, *SDHB* and *STK11* may increase susceptibility to GC as well as the familial syndrome associated with the gene of interest. Families found to carry pathogenic or likely pathogenic mutations within these genes may benefit from management strategies associated with such familial syndromes, as opposed to those in place for *CDH1*-mutation positive HDGC families.

It should be proposed that families with a strong history of diffuse or intestinal-type GC be genetically screened against a custom panel of known upper GI cancer susceptibility genes through panel sequencing as well as abnormalities at the *CDH1*-locus. This would provide more efficient and comprehensive genetic testing for such families, increasing the chances of detecting a causative genetic variant. Should a *CDH1*-abnormality (or possibly *CTNNA1*-germline variant) be detected, only then should a family be clinically defined as HDGC and unaffected carriers be provided with the appropriate risk stratification guidelines.

5.6 *In Silico* Methods Predict Pathogenicity of Variants of Unknown Significance Detected Using Panel Sequencing

Upon screening multiple FGC cases against our genes of interest, a number of missense mutations were identified in both moderate and highly penetrant genes. Rare (<1% general population) or novel missense VUS were identified in 47 probands without candidate pathogenic or likely pathogenic variants. It is often difficult to assess the functional implication of such missense variants, as they often do not result in a pathogenic phenotype. To predict functional impact of the amino acid change, variants were input into three separate *in silico* methods. Overall, 27 out of the 115 cases (23.5%) contained missense variants that were predicted as damaging in at least two *in silico* methods (**Table 4.2**). Mutations were found in genes *CFTR, EPCAM, BMPR1A, CASP10, MLH1, MSH2*, which have all been directly related to diseases of the upper GI tract such as pancreatic cancer, Lynch syndrome, Juvenile Polyposis and GC (Sharer et al. 1998; Lynch et al. 2011; Kempers et al. 2011; Park et al. 2001). Some genes are highly penetrant with respect to their associated disease; however, without additional *in vitro* and *in vivo* evaluation of the functional impact of these missense mutations, pathogenicity with respect to GC cannot be confirmed.

5.7 Candidate Variants Detected in Other Hereditary Upper GI Cancers

Additional familial upper GI cancers were sequenced against this panel of genes as art of a collaborative and ongoing project (**Figure 3.1; Table 3.1**). Familial EAC (n=17) familial pancreatic adenocarcinoma cases (n= 42) were collected from collaborating institutes in hopes of identifying genetic susceptibilities to the diseases using multiplexed TAS. Sequencing did not reveal any candidate truncating mutations in more than one family member with EAC/BE for the

genes screened. This concludes that either familial patterns of EAC/BE are influenced by heritable mutations in genes yet to be identified or there is little genetic basis of EAC/BE and familial clustering is more strongly influenced by environmental/epigenetic factors or phenocopies. Broader, WGS techniques would be a useful tool to conduct on such families to decipher the major genetic contributors (if any) to familial clustering of this rare disease.

TSCA assay was completed on 42 familial pancreatic cases with unexplained hereditary patterns of disease (**Table 3.1**). Protein truncating mutations in the form of nonsense variants were identified in 2/42 germline cases within the hereditary pancreatitis susceptibility gene *CFTR* and a novel candidate gene *MAP3K6*. These mutations were validated and the information relayed to collaborating institutes for downstream analysis. Germline mutations within *CFTR* have been long been described in families with a strong history of pancreatitis and pancreatic adenocarcinoma (Sharer et al. 1998). This supports the idea that a novel, truncating variant in the *CFTR* locus plays a pathogenic role in the familial aggregates of PC in this family. *MAP3K6* (*ASK2*) is a newly identified, possible upper GI cancer susceptibility gene, though likely with low to moderate penetrance. It has been implicated in epithelial tumour formation through *in vivo* methods by the creation of an *ASK2* knockout mouse (Takeda et al. 2007; Iriyama et al 2009). Along with the data from familial pancreatic cases, novel and rare truncating mutations were also identified and validated in 3 HDGC families within this study. As this gene of interest was part of collaboration outside of this funded project, this data was used for use in a manuscript that is currently under review (Gaston & Hansford et al., manuscript under review). Further work is ongoing and needed to establish *MAP3K6* as an upper GI cancer susceptibility gene.

Data collected from families meeting criteria for other familial upper GI cancers were analyzed and returned to the initiating institution for downstream analysis and validation of pathogenicity. This data is currently undergoing functional validation and will liely be included in future publications.

5.8 Conclusions

To this point, *CDH1* germline dysfunction is the only genetic risk factor for HDGC, and this is the only gene routinely tested in these families. As *CDH1* mutations have been identified in a large number of HDGC families (approximately 45%) for penetrance analysis, management strategies are put into place for families that meet clinical criteria for HDGC and are found to harbour such pathogenic variants. Extensive genetic counseling and possibly a radical prophylactic gastrectomy for *CDH1*-positive family members virtually eliminate the risk for GC in these individuals. However, all of FIGC cases and HDGC families found to not harbour *CDH1* germline mutations cannot be offered such risk stratification opportunities and there are currently no other genetic screens available. This demonstrates the power and importance of identifying additional genetic risk factors for FGC.

Using an efficient, multiplexed, next-generation sequencing assay, additional pathogenic mutations in FGC have been found in genes previously related to other upper GI syndromes, including *BRCA1*, *CTNNA1*, *SDHB*, *STK11*, *ATM, PALB2* and *MSR1*. The possibility of new GC susceptibility genes may increase the number of families who can benefit from targeted risk-reduction procedures. Of note was the identification of pathogenic or likely pathogenic mutations within the same gene across multiple families, of which clinical criteria for HDGC or FIGC. This suggests that though families meet clinical criteria specifically for HDGC of FIGC, they may

There is no doubt that as genetic sequencing approaches the realm of personalized medicine and the cost of such assays becomes increasingly more affordable, there will be a growing necessity to uncover specific guidelines that researches, genetic counselors and clinicians must follow with respect to reporting the discovery of incidental findings. Though the likelihood of uncovering such incidental variants is relatively lowered when using a targeted sequencing approach compared to the large output of whole genome or exome sequencing, it is likely that pathogenic mutations related to syndromes outside of the disease of interest may be identified. Certain questions such as how, what and when to report such variants are debated among professionals but it is agreed across professionals that actionable variants identified during the sequencing process (i.e. variants with a known associated risk with existing preventative measures) should be reported. It will become increasingly important to identify standardized guidelines to coincide with genetic sequencing assays. These guidelines will have to address issues such as the level consent by the patient, whether participants have an opportunity to opt-out of reporting once sequencing is conducted and is there a threshold specific variants or variants identified in a particular gene must meet to be considered 'actionable'.

Uncovering such genes that predispose familial to GC may provide prospective patients with therapeutic and management options as well as possibly detecting tumours at their earliest stage. Though it has been difficult to outline the risk for possible unaffected carriers in these families and the definitive pathogenicity of mutations within these genes, including such genes in future genetic screening of FGC cases may help identify additional families, which can be used for penetrance analysis. The genetic basis of unexplained cases of FGC is likely some combination of mutations in as of yet undetermined genes, phenocopies among families or, in the case of

HDGC families, other abnormalities at the *CDH1* locus. Further work is needed to uncover these genetic predispositions or sporadic events and genome-wide screening in families with unexplained susceptibilities is a logical next step for further discoveries.

Chapter 6: Limitations & Future Directions of Study

6.1 Limitations

In order to assess the ability of TSCA to detect large deletions and associated copy number alterations, a positive control sample with a large deletion of *CDH1* exons 1-2 was included in a sequencing run. However, neither the TSCA mutation analyzer nor secondary analysis by a bioinformatician could identify this deletion. Thus, the assessment of copy number alterations within targeted genes may be incomplete and such variants cannot be ruled out as possible causes of FGC.

It has been suggested that cryptic abnormalities within the *CDH1* locus could account for many HDGC families without easily detectable coding *CDH1* germline variants (Oliveira et al. 2009 (2)). However, the mechanism to which this mono-allelic expression ratio leads to pathogenic phenotype is not well understood. Further research is ongoing to evaluate the functional implications of *CDH1* non-coding variants as well as additional methods that may leads to mono-allelic *CDH1* expression (**Section 6.2 Future Directions**). RNA was not collected upon initial screening for pathogenic mutations within the *CDH1* locus for the families and is needed to complete allele-Specific Expression (ASE) analysis. Without this completed analysis and downstream functional validation of such transcriptional expression ASE at the *CDH1* locus cannot be excluded as possible causation. With new evidence coming to light of the possible pathogenicity of germline *CDH1*-ASE-carrying individuals, this highlights the necessity of RNA collection from prospective HDGC families for downstream screening.

When a HDGC family meets criteria for *CDH1* testing, the affected proband is submitted for screening and if no pathogenic mutation is identified within the *CDH1*-locus, no additional family members are collected for screening purposes. An evident issue with respect to validating the pathogenicity of our detected variants was the limited availability of samples. It is critical to segregate these novel genetic abnormalities within additional family members to validate their pathogenicity and downstream penetrance analyses. When applicable, genetic counselors or families were contacted directly to inform them of the results and to retrieve additional samples from affected/unaffected individuals for carrier analysis. Communicating the impact of the identified variants to genetic counselors also proved difficult, as many resisted in contacting families unless assurance could be given that the mutation is causative. As the genomics era moves forward and families have increased availability to such unclassified genetic information, developing a multidisciplinary foundation of genetic counselors, geneticists, researchers, clinicians and more becomes essential to communicate with families and provide them with valuable information.

Another challenging aspect of sample collection is the vulnerable state affected family members are in when screening is conducted. With the late-stage detection and high mortality rate of this cancer, the health of the proband often rapidly deteriorates through the screening process and they often succumb to disease. This makes it impossible to collect further genomic specimens for additional analyses. For incoming HDGC families submitted for screening and who have shown interest in being included in future research studies, it is recommended that RNA and additional affected family members be collected for downstream ASE and detailed segregation analyses, respectively.

A notable issue was raised when unexpected, challenging truncating and splice site mutations were detected in low-moderately-penetrant genes such as *ATM, MSR1, PALB2* and *SLC22A4*. Though pathogenicity is likely, communicating these likely moderate-penetrance variants to respective genetic counselors in order to collect samples from additional family members has proven difficult. Uncovering the genotype-phenotype relationship of such mutations without functional validation is challenging and the rarity of FGC further complicates this process.

It is also difficult to rationalize a radical cancer prevention tool for such families when penetrance is not well supported. Identifying additional families through screening of a larger cohort woud help uncover the true molecular contribution of these genes to this rare phenotype.

A further limitation, to reconcile the information herein collected about genetic basis of inherited GC, is the potential ambiguity related to classification of families in different syndromes. In fact, the information collected from pathology reports often establishes whether a family is classified as HDGC or FIGC. Any inaccuracy in this classification may have as a consequence the description of the same causative or potentially causative gene in two distinct clinical entities. This may be the case of families 13 and 110 (carrying the same *SDHB* germline mutation, one classified as HDGC and another as FIGC) and of families 16 and 105 (carrying the same *ATM* germline mutation, one classified as HDGC and another as FIGC). We did not have access to the appropriate tissue blocks to revise the histopathology of these cases, but it would be advisable to do so in the future, in order to determine whether a given germline defect is more likely to occur in HDGC or FIGC-related contexts.

6.2 Future Directions

The current study has brought to light new gastric cancer susceptibility genes, all of which have been previously identified in upper GI-related syndromes and some of which (*CTNNA1*) are directly involved in the same pathway as the current susceptibility gene *CDH1*. Although limitations in sample collection and downstream segregation precluded the confirmation that all identified variants are highly penetrant and pathogenic, this multiplexed, next generation sequencing method has proven both useful and efficient in identifying novel and rare germline mutations. Further work will be needed to identify the overall contribution of these genes to familial gastric cancer as well as the penetrance of heritable mutations within these genes (**Figure 6.1**).

6.2.1 Functional Analysis of *CTNNA1* Genetic Abnormalities

Further look into the functional implications of mutations within cell-adhesion molecules (*CDH1* and *CTNNA1*) would provide a great deal of insight into the molecular pathogenesis of both hereditary and sporadic forms of diffuse gastric cancer. Now that *CTNNA1* is being established as a gastric cancer susceptibility gene, studying the functional consequences of *CTNNA1* protein-truncating mutations and the downstream affected pathways could leads to further understanding of germline *CDH1* mutations. Upon the discovery of two additional families with germline, truncating *CTNNA1* mutations, a gastric cancer cell line mimicking the lost expression of *CTNNA1* via immuno-flourescence was identified as a candidate model for exploring molecular pathogenesis. Dr. H Li-Chang has been assigned the project to investigate specific pathways that may be involved in the development and progression of gastric cancer in patients with germline cell-adhesion molecule mutations, such as *CTNNA1*. He will first identify the genetic
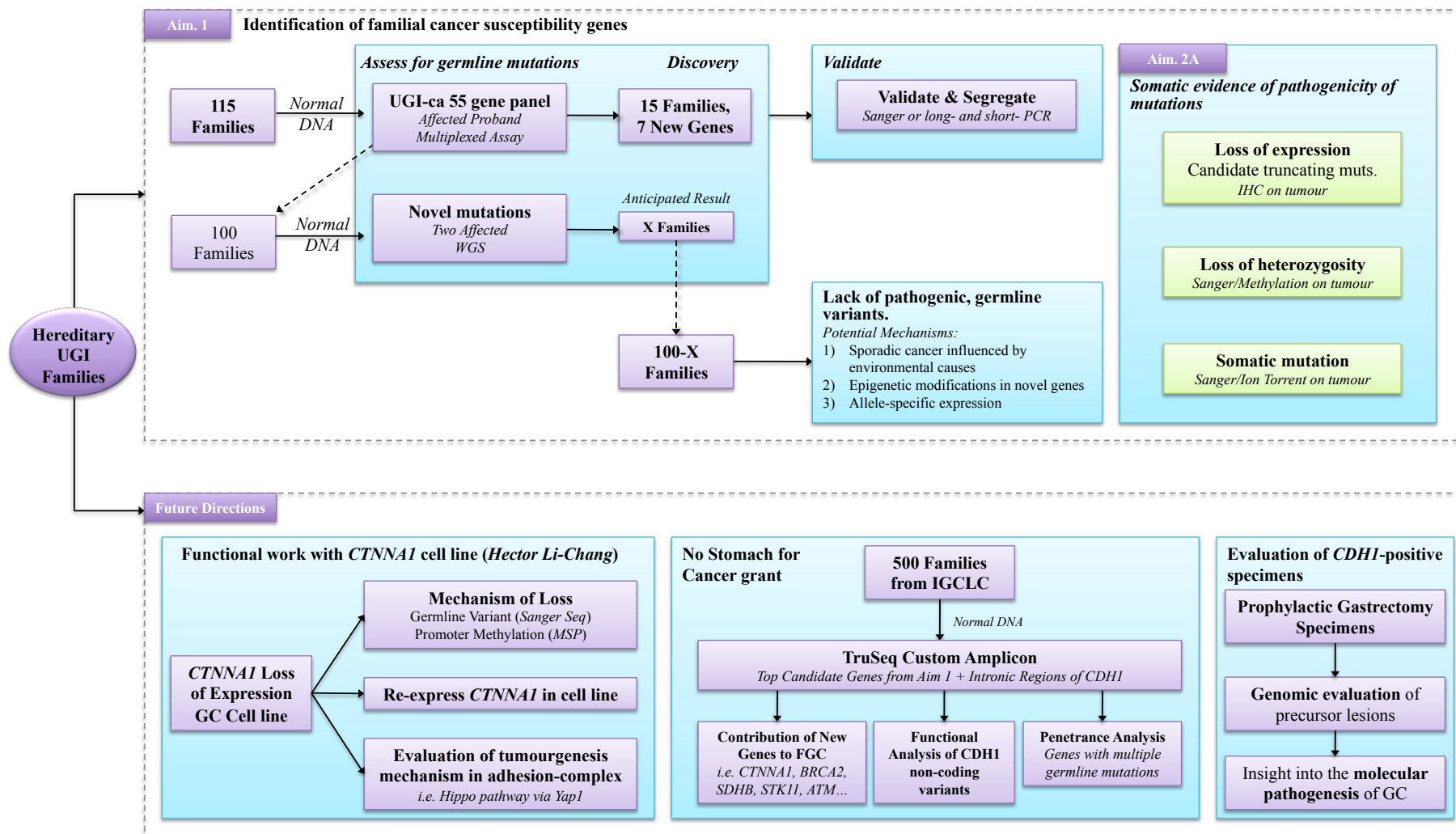
Figure 6.1 Workflow of current project and future directions for continued research

abnormality leading to loss of alpha-E-catenin expression in the cell line through Sanger sequencing and epigenetic analyses. Next, re-expressing *CTNNA1* and subsequent functional analyses is crucial to understanding the important role *CTNNA1* plays in cell-adhesion and if loss of *CTNNA1* is truly the key genetic defect to pathogenicity. The role of *CTNNA1* as a tumour suppressor with respect to its functional relationship to the transcription co-activator Yap1, has previously been studied in the context of epithelial cancers (Silvis et al. 2011). I was shown that loss of *CTNNA1* increases nuclear localization of *Yap1* and subsequent activation of the contact inhibition and cell proliferation Hippo pathway (Silvis et al. 2011). Genetic deletion of alpha-E-catenin in the stem cell hair follicle leads to squamous cell carcinoma, in vivo. Immunohistochemistry of the tumour reveals complete loss of alpha-E-catenin and correlation to nuclear localization of *Yap1* (Silvis et al. 2011).

6.2.1 No Stomach For Cancer Grant

Dr. M Woo, Dr. H Li-Chang, Dr. D Huntsman, Dr. C Oliveira and S. Hansford equally contributed the written work of this grant. The proposed project to carry the work herein forward was submitted to the No Stomach For Cancer Organization on March, 3rd 2014 for competition in the form of a \$50,000 funded grant. This grant proposes identifying the contribution of the newly identified susceptibility genes from this study to FGC worldwide as well as the functional analysis of *CDH1* non-coding mutations. The overall objective will be to collect and screen >500 *CDH1* mutation negative families that have been identified worldwide to determine the genetic contribution of new susceptibility genes. With this proposed study, we may be able to conduct penetrance analysis on those genes in which multiple families carry mutations, stratify risk in

these families through carrier screening and possibly (in genes found to carry highly penetrant mutations) propose risk-reduction surgeries, such as those offered to *CDH1*-carriers.

In addition to identifying the relative contribution of new GC susceptibility genes, this grant will be in collaboration with Dr. C Oliviera's group in Porto, Portugal. This portion of the grant proposes to evaluate the functional implications of non-coding mutations within the *CDH1*-locus, specifically those found in intron 2. Preliminary data shows these non-coding single nucleotide variants may underlie germline *CDH1* mono-allelic down regulation through disrupting *CDH1* transcription factor binding sites. It is conceivable that mutations in these non-coding regulatory sequences could reduce transcription efficiency, resulting in the observed phenotype of allele-specific down regulation found in ~70% *CDH1*-negative HDGC families. The overarching goal of this proposed study is to improve patient management by identifying disease-causing variants in *CDH1* sequences conventionally excluded from current genetic testing, and assess the contribution of mutations in other genes on the development of familial gastric cancers.

Background

Although pathogenic *CDH1* mutations are present in 45% of families with hereditary diffuse gastric cancer (HDGC) and represent a powerful tool for cancer prevention, the genetic basis for the remaining cases of HDGC and all familial intestinal gastric cancers is unknown. We recently implicated other gastric cancer (GC) susceptibility genes in *CDH1*-negative families. In addition, we have data suggesting that mutations within non-coding regions of *CDH1* may contribute to HDGC. Our goal is to identify and characterize disease-causing variants in *CDH1* non-coding sequences and better characterize the impact of other genes on the development of FGC.

Hypotheses. 1. FGC that are not attributable to *CDH1* coding mutations will have mutations in *CDH1* non-coding regions or other genes that increase risk for disease. **2.** Non-coding *CDH1*-variants may cause monoallelic down-regulation, reduce protein expression and increase GC susceptibility. **Aim 1A.** Sequence germline DNA from >500 *CDH1*-negative FGC cases against a custom panel of newly suggested GC susceptibility genes using a multiplexed next generation sequencing assay. **1B.** Quantify the impact of genes mutated in multiple families by conducting penetrance analyses. **Aim 2A.** Determine if single non-coding *CDH1* variants impair protein expression in HDGC patients using allele specific expression and segregation analyses. **2B.** Determine the functional impact of candidate non-coding variants selected in Aim2. A using a knock-in strategy (CRISPR), RNAi and electrophoretic mobility shift assays.

Impact. Clinical care for families that test negative for pathogenic CDH1-variants is limited by the absence of adequate testing and early cancer detection. Identifying new genetic abnormalities would improve risk-stratification in these families, offer more comprehensive screening and empower decision making in at-risk patients considering prophylactic surgery. This study will also help the global community of HDGC researchers determine which families are most suited for full genome sequencing and other more intensive gene discovery exercises.

Preliminary Data

Screening of 115 *CDH1* mutation negative familial gastric cancer families using targeted, multiplexed next generation sequencing. Our group recently identified pathogenic germline mutations in non-*CDH1* genes in FGC cases using a multiplex panel sequencing approach (Hansford et al. Manuscript in preparation. Though literature search and concurrent projects,

we developed a custom panel of 55 genes for which germline or somatic aberrations have been implicated in familial upper gastrointestinal cancers or cancer susceptibility syndromes. The list included well-established tumor suppressor genes and those associated with loss of expression in upper gastrointestinal tumour samples. We screened germline DNA from 115 probands from families who met International Gastric Cancer Linkage Consortium clinical criteria for HDGC (Fitzgerald et al. 2010) (n=106, all tested negative for *CDH1* mutations) or FIGC (n=9) (Caldas et al. 1999). Samples were collected from the BC Cancer Agency (Vancouver, Canada), Institute of Molecular Pathology and Immunology of the University of Porto (Porto, Portugal) and the University of Siena (Siena, Italy). Using a multiplexed next generation sequencing assay, we identified pathogenic or likely pathogenic variants in 13% of cases (n=15) (11 HDGC and 4 FIGC families). The variants included clearly pathogenic, novel truncating mutations in 4 unrelated families (two different mutations in *CTNNA1* (alpha-catenin) and two different mutations in *BRCA2)*. Mutations in genes *SDHB* (2 families) and *STK11* (2 families) were also identified, of which have been implicated in the GC susceptibility syndromes Cowden-like syndrome and Peutz-Jeghers disease, respectively. Additional truncating mutations were identified in genes of moderate-penetrance (risk-association): *ATM* (4 families), *MSR1* (2 families) and *PALB2* (1 family). Novel missense variants predicted to be damaging by *in silico* methods were found in 27 cases (23.5%) in different genes within the panel. This data demonstrates the utility of a targeted sequencing approach and our improved ability to detect pathogenic mutations in families with unexplained FGC. By targeting genes with known diagnostic value and multiplexing up to 96 samples simultaneously, this approach limits total test output compared to WGS or ES while providing cost-advantages over traditional, low-throughput, single-gene testing methods (Figure 3.2).

Germline variants in the non-coding region of the *CDH1* locus. Dr. C Oliveira's group (who is a collaborator on this proposed project) have completed allelic imbalance studies showing that several germline abnormalities in *CDH1* could not be accounted for by coding mutations or copy number changes. This led to hypothesize that families without *CDH1*-coding mutations may harbor germline alterations in non-coding regions (Pinheiro et al. 2010). To address this, 90 *bona fide CDH1*-negative HDGC probands were screened by next-generation sequencing of the complete *CDH1* locus and 42 true rare heterozygous germline *CDH1* non-coding candidate variants that could be pathogenic were identified. The preliminary bioinformatics query predicted that a unique variant often interfered with the binding of several transcription factors, some of which are classical modulators of *CDH1* expression. Moreover, several variants occurred in sequences that are highly conserved among species or were located within DNase I hypersensitivity sites, which are expression regulatory regions. These predictions suggest that some *CDH1* non-coding variants may cause germline *CDH1* mono-allelic down-regulation and subsequent loss of protein expression. For this reason, the custom gene panel proposed in Aim 1 will screen for variants in the entire *CDH1* locus. We will further explore the functional relevance of the detected *CDH1* non-coding mutations in Aim 2.

Collaboration and Partnership with the International Gastric Cancer Linkage Consortium. This study will leverage our preliminary findings and the large resource of families available within the IGCLC, of which Dr.'s D Huntsman and C Oliveira have been members for over 15 years. The IGCLC is a GC reserch-supportive group, whose goals are to define the clinical criteria for FGC, outline management strategies for affected families and promote collaborative research across international institutions.

This group of leading physicians and scientists meet regularly to present important findings on clinical management, genetics, biology, pathology and treatment of familial GC and discuss prospective research in this area. A letter describing the current project and invitation to include *CDH1* mutation-negative families in our genetic screen has already been sent to 33 members of the IGCLC and other international researchers, clinicians and genetic counselors. We have already received positive feedback from many who would like to include their families into this study and will use the upcoming IGCLC meeting in Nijmegem to elicit further engagement and anticipate the inclusion of 500 *CDH1*-negative GC families worldwide.

Hypotheses

1. Families with FGC that are not attributable to *CDH1* coding mutations will have identifiable causative mutations in *CDH1* non-coding regions and other susceptibility genes that increase risk for disease.

2. Non-coding *CDH1* variants may cause monoallelic down-regulation leading to loss of E-cadherin expression and increased susceptibility to DGC.

Aims, Research Design and Methods

Through grant funding from the Canadian Institutes of Health Research to D. Huntsman, we were able to assess the contribution of susceptibility genes other than *CDH1* in our own cohort of 115 families (preliminary data, **Section 3.1**). Funding to C. Oliveira from The Portuguese Science Foundation (FCT) allowed the screening of the full CDH1-locus in 90 HDGC families (preliminary data, Section 3.2). Funding of this application will enable the screening of the majority of CDH1-mutations negative GC families worldwide and provide the necessary evidennce of the functional implications of CDH1 non-coding variants.

It will triage candidate families for other genetic studies such as whole genome or exome sequencing. All laboratory techniques, including next generation sequencing analyses using the Illumina MiSeq desktop platform, bioinformatic and statistical analyses methods and functional assays are well established in the labs to be used.

Aim 1A.

Screen >500 *CDH1*-negative gastric cancer families from a worldwide cohort against a custom gene panel using targeted multiplexed sequencing. Based on preliminary data, we have selected 21 genes along with the *CDH1* full locus to be included in our panel for screening *CDH1*-mutation negative HDGC and FIGC families in the IGCLC cohort (**Figure 6.2; Table 6.1)**. Additional genes from collaborators may be added if there is strong compelling evidence for their involvement in FGC development. Germline DNA, and tumour materials when applicable, will be shipped to the laboratory of Dr. Huntsman (BC Cancer Agency, Vancouver, BC). Based on previous use of the Illumina Design Studio software, we estimate >1000 amplicons will be generated covering the entire *CDH1*-locus and all coding (exonic) regions of the other genes. The Minimum required input DNA (50ng) will be confirmed through fluorometer quantification. Multiplexed sequencing analysis of germline DNA will be performed using the Illumina TSCA assay as described in preliminary work. Up to 96 samples can be multiplexed in a single run for simultaneous sequencing and data analysis on the MiSeq platform.

Aim 1B.

Quantify the impact of genes mutated in multiple families by conducting penetrance analyses. Recent data suggests some FGC cases can be attributed to germline truncating mutations in the cellular adhesion molecule, *CTNNA1*, and the well-established cancer susceptibility gene, *BRCA2* (Majewski et al. 2013). From Aim 1A, we anticipate the identification of additional families with mutations in these genes and others which will enable us to estimate the penetrance of mutations in these families. Pedigree information will be collected from genetic counselors from the referring centres and segregation analysis performed to estimate the penetrance of the mutations using the MENDEL program, as previously described (Lange et al. 1988).

Aim 2A.

Determine if single non-coding *CDH1* variants impair protein expression. HDGC families in which non-coding variants are identified, allele-specific expression analysis will be performed on germline RNA to identify occurrence of *CDH1* monoallelic expression. If presence of *CDH1* non-coding variant is associated with a germline monoallelic phenotype, segregation analysis will be conducted in other family members. Non-coding variants that segregate in the family and are associated with *CDH1* allele-specific expression will be selected for further analysis.

Aim 2B.

Determine the functional impact of non-coding variants. The functional impact of non-coding *CDH1* variants selected in Aim 2A, and variants for which segregation and allele-specific expression analysis are impossible to obtain, will be tested using a knock-in strategy (CRISPR) to produce variant/wild-type heterozygous cell lines (Hwang et al. 2013). The impact of these single heterozygous non-coding variants in *CDH1* allele specific expression will be tested in these cell lines. Those variants leading to monoallelic expression will be selected. Following prediction of the differential binding of transcription factors to non-coding mutant sequences, a RNAi-targeted approach will be used to evaluate the impact of depleting potential *CDH1*-negative regulators and restore *CDH1* biallelic expression. Electrophoretic mobility shift assay, coupled with cell extracts and specific antibodies, will be used to demonstrate the binding of specific *CDH1* modulators to non-coding sequences encompassing HDGC-associated variants. The identification of non-coding variants that impact *CDH1* expression by affecting the binding of regulatory molecules will constitute a strong argument towards including these sequences as novel amplicons in diagnostic approaches currently offered to HDGC patients.

Research Impact and Translational Pathway

Prophylactic total gastrectomy minimizes the risk of GC in those carrying pathogenic *CDH1* germline mutations. Options for further screening and clinical decision-making are severely hindered in families that meet IGCLC criteria for HDGC but test negative for *CDH1* mutations. The hopeful outcome of this proposed project is the identification of genetic susceptibilities to GC using a practical and efficient method of screening targeted regions of interest. Those families in whom variants are not identified can then be triaged for more resource-intensive methods, such as whole-exome or WGS. Functional studies of both CDH1 non-coding variants

and non-*CDH1* cancer predisposition genes (to be done in parallel) will enhance our understanding of both hereditary and sporadic GC. The results of this project could facilitate improved prevention and treatment strategies in the management familial gastrointestinal malignancies. The discovery and characterization of novel susceptibilities would also improve risk stratification in families in which pathogenic variants are identified. From a more practical perspective, these findings are likely to improve patient management through the identification of disease-causing abnormalities, which can be added to the conventional genetic screening methodology offered to prospective HDGC families.

Research Team

The team includes Drs. D Huntsman and C Oliveira (co-Principal Applicants) who have been active researchers in the field of HDGC since 1998 and have worked together on many projects. Collectively, they have published more than 60 papers in this area (18 of them in co-authorship) and have: (i) described the majority of *CDH1* mutations in HDGC families including atypical mutation types such as deletions (Brooks-Wilson et al. 2004; Suriano et al. 2005; Oliveira et al. 2009), founder mutations in Newfoundland (Kaurah et al. 2007) and germline promoter methylation (Pinheiro et al. 2010) (**Figure 1**); (ii) first described occult carcinomas in prophylactic total gastrectomy samples, thus establishing that prophylactic total gastrectomy is standard of care for this condition (Huntsman et al. 2001; Carneiro et al. 2004; Lee et al. 2010); (iii) contributed much of the counseling knowledge for HDGC (Fitzgerald et al. 2010; Caldas et al. 1999; Schrader et al. 2011) ; (iv) shown that women who have undergone prophylactic total gastrectomy can have successful pregnancies (Kaurah et al. 2010) ; and (v) demonstrated the heterogeneity of CDH1-second hits in primary and metastatic lesions from the same patient (Oliveira et al. 2009).

Dr. Huntsman is a global leader in the application of cancer genetics to improve management and will lead the next-generation sequencing analysis in Aim 1A. Dr. Oliveira has contributed to the understanding of HDGC development through the identification of heterogeneous somatic alterations in HDGC tumours and recently presented the first tentative therapy for HDGC through the recovery of *CDH1* expression and function with read-through strategies. Samantha Hansford (graduate student) developed the gene panel and performed next-generation sequencing assay on the preliminary 115 families and has several years experience conducting genetic studies on HDGC. She will lead the ascertainment of GC families and perform the sequencing analysis. Pardeep Kaurah is an experienced gastric cancer genetic counselor and will collect pedigree information for penetrance analyses. Dr. H Li-Chang is a pathologist with sub-specialty training in gastrointestinal neoplasia and currently a pathology research fellow in the Huntsman lab. Dr. H Pinheiro did his PhD supervised by Dr.'s C Oliveira and D Huntsman and demonstrated the presence of germline *CDH1* mono-allelic expression in *CDH1*-negative families; he is currently a Post-doctoral fellow in the Oliveira lab and together with Dr. J Carvalho, (post-doctoral fellow in Oliveira's lab) will work on the functional analysis of potentially deleterious mutations in non-coding *CDH1* regions or other candidate genes. Dr. J Carvalho has recently found a link between the presence of somatic deletions of the *CDH1* gene in GC and family history of FIGC in C. Oliveira's lab (Corso et al. 2013).
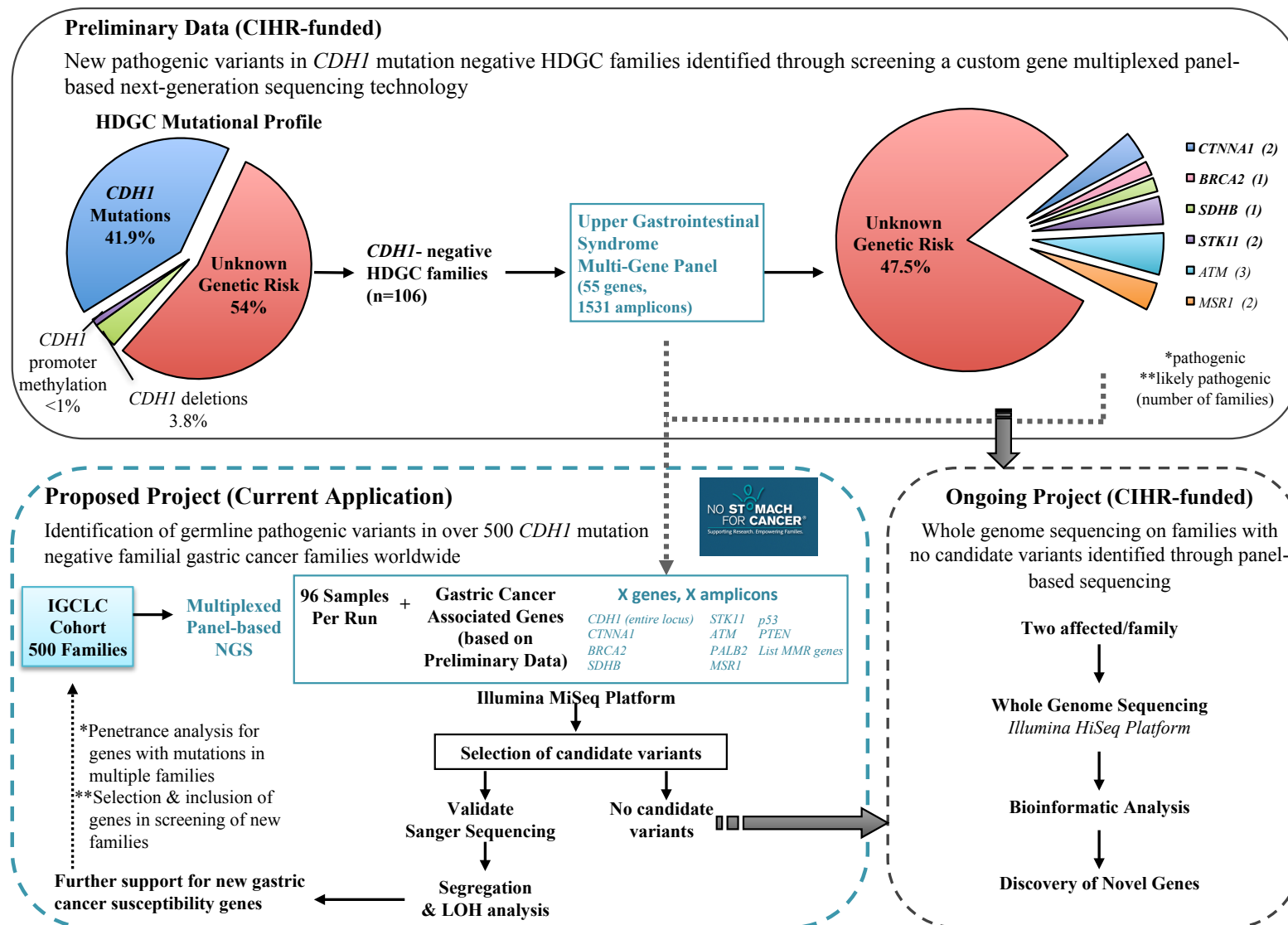
Figure 6.2 Preliminary data and workflow for No Stomach For Cancer grant proposed project.

Table 6.1 Genes selected for custom panel sequencing in proposed No Stomach For Cancer grant

Syndrome	Gene	Mutation Status	Penetrance
Carney-Stratakis	SDHB	Heterozygous	High
Colorectal Carcinoma & Polyposis	MUTYH	Homozygous	High
Esophgeal Adenocarcinoma / Barrett's Esophagous	MSR1	Heterozygous	Intermediate
Familial Adenomatous Polyposis	APC	Heterozygous	High
Gastric Cancer	BRCA2	Heterozygous	High
	CDH1*	Heterozygous	High
	CTNNA1	Heterozygous	High
	PSCA	Heterozygous	High
	PTEN	Heterozygous	High
Cowden Syndrome	PTEN	Heterozygous	High
	SDHB	Heterozygous	High
Gastrointestinal-Type Polyposis	MSH3	Heterozygous	High
Lynch Syndrome	EPCAM	Heterozygous	High
	MLH1	Heterozygous	High
	MSH2	Heterozygous	High
	MSH3	Heterozygous	High
	MSH6	Heterozygous	High
	PMS1	Heterozygous	High
	PMS2	Heterozygous	High
Pancreatic Cancer	ATM	Heterozygous	Intermediate
	PALB2	Heterozygous	Moderate
Peutz-Jeghers Syndrome	STK11	Heterozygous	High

Bibliography

Adzhubei, IA et al. A method and server for predicting damaging missense mutations. Nature Methods. 20120;7, 248–249.

Agilent Technologies (2012) HaloPlex target enrichment system for illumina sequencing (old version) v.A. Publication number: G9900-90000

Aparicio SA, Huntsman DG (2010) Does massively parallel DNA resequencing signify the end of histopathology as we know it? J Pathol 220(2):307–315. doi:10.1002/path.2636

Apostolou P & Fostira F. Review: Hereditary Breast Cancer: The era of new susceptibility genes. BioMed Research International. Volume 2013 (2013), Article ID 747318, 11 pages.

Barber ME, Save V, Carneiro F, Dwerryhouse S, Lao-Sirieix P, Hardwick RH, Caldas C, Fitzgerald RC (2008b) Histopathological and molecular analysis of gastrectomy specimens from heredi- tary diffuse gastric cancer patients has implications for endoscopic surveillance of individuals at risk. J Pathol 216:286–294. doi:10.1002/path.2415

Benjamini, Yoav; Hochberg, Yosef. Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society, Series B. 1995;57 (1): 289–300.

Bernstein, J. L., Teraoka, S., Southey, M. C., Jenkins, M. A., Andrulis, I. L., Knight, J. A., John, E. M., Lapinski, R., Wolitzer, A. L., Whittemore, A. S., West, D., Seminara, D., Olson, E. R., Spurdle, A. B., Chenevix-Trench, G., Giles, G. G., Hopper, J. L., Concannon, P. Population-based estimates of breast cancer risks associated with *ATM* gene variants c.7271T-G and c.1066-6T-G (IVS10-6T-G) from the breast cancer family registry. Hum. Mutat. 27: 1122-1128, 2006.

Blanco A, de la Hoya M, Osorio A, Diez O, Miramar MD, et al. (2013) Analysis of *PALB2* Gene in BRCA1/*BRCA2* Negative Spanish Hereditary Breast/Ovarian Cancer Families with Pancreatic Cancer Cases. PLoS ONE 8(7): e67538. doi:10.1371/journal.pone.0067538

BOOK: Bosman, F. T., Carneiro, F., Hruban, R. H. & Theise, N. D. (Eds) WHO Classification of Tumors of the Digestive System 4th edn Vol. 3 (IARC Press, Lyon, 2010). Pages 48-63

Borch, K., Jönsson, B., Tarpila, E., Franzén, T., Berglund, J., Kullman, E., & Franzén, L. (2000). Changing pattern of histological type, location, stage and outcome of surgical treatment of gastric carcinoma. The British Journal of Surgery, 87(5), 618–626. doi:10.1046/j.1365-2168.2000.01425.x

Bringuier, P. P., Umbas, R., Schaafsma, H. E., Karthaus, H. F., Debruyne, F. M., & Schalken, J. A. (1993). Decreased E-cadherin immunoreactivity correlates with poor survival in patients with bladder tumors. Cancer Research, 53(14), 3241–3245.

Brooks-Wilson, A. R., Kaurah, P., Suriano, G., Leach, S., Senz, J., Grehan, N., et al. (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(7), 508–517.

Bybee SM, Bracken-Grissom H, Haynes BD, Hermansen RA, Byers RL, Clement MJ, Udall JA, Wilcox ER, Crandall KA (2011) Targeted amplicon sequencing (TAS): a scalable next-gen approach to multilocus, multitaxa phylogenetics. Genome Biol Evol 3:1312–1323. doi:10.1093/ gbe/evr106

Byrnes GB, Southey MC, Hopper JL. Are the so-called low penetrance breast cancer genes, *ATM*, BRIP1, *PALB2* and CHEK2, high risk for women with strong family histories? Breast Cancer Res. 2008;10(3):208. doi: 10.1186/bcr2099. Epub 2008 Jun 5.

Caldas C, Carneiro F, Lynch HT, et al. Familial gastric cancer: overview and guidelines for management. J Med Genet 1999;36:873e80.

Calva-Cerqueira D, Dahdaleh FS, Woodfield G, Chinnathambi S, Nagy PL, Larsen-Haidle J, Weigel RJ, Howe JR (2010) Discovery of the BMPR1A promoter and germline mutations that cause juvenile polyposis. Hum Mol Genet 19(23):4654–4662

Carneiro F, Huntsman DG, Smyrk TC, Owen DA, Seruca R, Pharoah P, Caldas C, Sobrinho-Simões M. Model of the early development of diffuse gastric cancer in E-cadherin mutation carriers and its implications for patient screening. J Pathol. 2004 Jun;203(2):681-7.

Carneiro F, Oliveira C, Suriano G, Seruca R. Molecular pathology of familial gastric cancer, with an emphasis on hereditary diffuse gastric cancer. J Clin Pathol. 2008 Jan;61(1):25-30. Epub 2007 May 18. Review.

Carneiro, F. REVIEW: Hereditary gastric cancer. Pathology 2012 · [Suppl 2] 33:231–234 DOI 10.1007/s00292-012-1677-6

Chan B, Facio FM, Eidem H, Hull SC, Biesecker LG, Berkman BE (2012) Genomic inheritances: disclosing individual research results from whole-exome sequencing to deceased participants' relatives. Am J Bioeth 12(10):1–8. doi:10.1080/15265161.2012.699138

Charlton A, Blair V, Shaw D, et al. Hereditary diffuse gastric cancer: predominance of multiple foci of signet ring cell carcinoma in distal stomach and transitional zone.Gut.2004;53:814–820.

Chenevix-Trench G, Spurdle AB, Gatei M, Kelly H, Marsh A, Chen X, Donn K, Cummings M, Nyholt D, Jenkins MA, et al. Dominant negative *ATM* mutations in breast cancer families. J Natl Cancer Inst. 2002 Feb 6; 94(3):205-15.

Choi Y, Sims GE, Murphy S, Miller JR, Chan AP. Predicting the Functional Effect of Amino Acid Substitutions and Indels. PLoS ONE. 2012;7(10): e46688.

Choi Y. A Fast Computation of Pairwise Sequence Alignment Scores Between a Protein and a Set of Single-Locus Variants of Another Protein. In Proceedings of the ACM Conference on Bioinformatics, Computational Biology and Biomedicine. 2012;ACM, New York, NY, USA, 414-417.

Chun, N., & Ford, J. M. (2012). Genetic testing by cancer site: stomach. Cancer Journal (Sudbury, Mass.), 18(4), 355–363. doi:10.1097/PPO.0b013e31826246dc

Conacci-Sorrell, M., Zhurinsky, J., & Ben-Ze'ev, A. (2002). The cadherin-catenin adhesion system in signaling and cancer. The Journal of Clinical Investigation, 109(8), 987–991. doi:10.1172/JCI15429

Coonrod EM, Durtschi JD, Margraf RL, Voelkerding KV (2012) Developing genome and exome sequencing for candidate gene identification in inherited disorders. Arch Pathol Lab Med. doi:10.5858/arpa.2012-0107-RA

Corso G, Carvalho J, Marrelli D, et al. Somatic mutations and deletions of the E-cadherin gene predict poor survival of patients with gastric cancer. J Clin Oncol. 2013 Mar 1;31(7):868-75. doi: 10.1200/JCO.2012.44.4612.

Corso G., Marelli D. & Roviello F. (2013) Clinical Management of Familial Gastric Cancer. In G. Corso & F. Roviello (eds.), Spotlight on Familial and Hereditary Gastric Cancer (183-190)

Dicken, B. J., Bigam, D. L., Cass, C., Mackey, J. R., Joy, A. A., & Hamilton, S. M. (2005). Gastric adenocarcinoma: review and considerations for future directions. Annals of Surgery, 241(1), 27–39.

Duncavage EJ, Abel HJ, Szankasi P, Kelley TW, Pfeifer JD (2012) Targeted next generation sequencing of clinically significant gene mutations and translocations in leukemia. Mod Pathol 25(6):795–804. doi:10.1038/modpathol.2012.29

Ejima, Y., Yang, L., & Sasaki, M. S. (2000). Aberrant splicing of the *ATM* gene associated with shortening of the intronic mononucleotide tract in human colon tumor cell lines: a novel mutation target of microsatellite instability. International Journal of Cancer. Journal International Du Cancer, 86(2), 262–268.

Epidemiol Rev. 1986;8:1-27. The decline in gastric cancer: epidemiology of an unplanned triumph. Howson CP, Hiyama T, Wynder EL.PMID: 3533579

Erkko H, Dowty JG, Nikkila J, et al. Penetrance analysis of the *PALB2* c.1592delT founder mutation. Clin Cancer Res. 2008; Jul 15;14(14):4667-71. doi: 10.1158/1078-0432.CCR-08-0210.

Ferla R, Calò V, Cascio S, et al. Founder mutations in BRCA1 and *BRCA2* genes. Ann Oncol. 2007 Jun;18 Suppl 6:vi93-8.

Fernandes, P. H., Saam, J., Peterson, J., Hughes, E., Kaldate, R., Cummings, S., Theisen, A., Chen, S., Trost, J. and Roa, B. B. (2014), Comprehensive sequencing of *PALB2* in patients with breast cancer suggests *PALB2* mutations explain a subset of hereditary breast cancer. Cancer. doi: 10.1002/cncr.28504

Ferreira, A. C., Suriano, G., Mendes, N., Gomes, B., Wen, X., Carneiro, F., et al. (2012). E-cadherin impairment increases cell survival through Notch-dependent upregulation of Bcl-2. Human Molecular Genetics, 21(2), 334–343. doi:10.1093/hmg/ddr469

Figer A, Irmin L, Geva R, et al. The rate of the 6174delT founder Jewish mutation in *BRCA2* in patients with non-colonic gastrointestinal tract tumors in Israel. Br J Cancer. 2001 Feb;84(4):478-81.

Figueiredo, J., Söderberg, O., Simões-Correia, J., Grannas, K., Suriano, G., & Seruca, R. (2013). The importance of E-cadherin binding partners to evaluate the pathogenicity of E-cadherin missense mutations associated to HDGC. European Journal of Human Genetics : EJHG, 21(3), 301–309. doi:10.1038/ejhg.2012.159

Fitzgerald RC, Hardwick R, Huntsman D, et al; International Gastric Cancer Linkage Consortium. Hereditary diffuse gastric cancer: updated consensus guidelines for clinical management and directions for future research. J Med Genet. 2010 Jul;47(7):436-44. doi: 10.1136/jmg.2009.074237.

Forcet C, Etienne-Manneville S, Gaude H, et al. Functional analysis of Peutz-Jeghers mutations reveals that the LKB1 C-terminal region exerts a crucial role in regulating both the AMPK pathway and the cell polarity. Hum Mol Genet. 2005 May 15;14(10):1283-92. Epub 2005 Mar 30.

Gall, T. M. H., & Frampton, A. E. (2013). Gene of the month: E-cadherin (*CDH1*). Journal of Clinical Pathology, 66(11), 928–932. doi:10.1136/jclinpath-2013-201768

Gaston & Hansford et al. (2014) Manuscript under review.

Giardiello, F. M., Brensinger, J. D., Tersmette, A. C., Goodman, S. N., Petersen, G. M., Booker, S. V., et al. (2000). Very high risk of cancer in familial Peutz-Jeghers syndrome. Gastroenterology, 119(6), 1447–1453.

Gilad, S., Chessa, L., Khosravi, R., Russell, P., Galanty, Y., Piane, M., Gatti, R. A., Jorgensen, T. J., Shiloh, Y., Bar-Shira, A. Genotype-phenotype relationships in ataxia-telangiectasia and variants. Am. J. Hum. Genet. 62: 551-561, 1998.

Global cancer statistics. Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D. CA Cancer J Clin. 2011 Mar-Apr;61(2):69-90. doi: 10.3322/caac.20107. Epub 2011 Feb 4. Erratum in: CA Cancer J Clin. 2011 Mar-Apr;61(2):134.

Goldgar DE, Healey S, Dowty JG, et al. Rare variants in the *ATM* gene and risk of breast cancer. Breast Cancer Res. 2011 Jul 25;13(4):R73. doi: 10.1186/bcr2919.

Gore R. Gastrointestinal cancer. Radiol Clin North Am. 1997;35:295–310.Borch:2000es

Goya R, Sun MG, Morin RD, Leung G, Ha G, Wiegand KC, Senz J, Crisan A, Marra MA, Hirst M, Huntsman D, Murphy KP, Aparicio S, Shah SP. SNVMix: predicting single nucleotide variants from next-generation sequencing of tumors. Bioinformatics. 2010 Mar 15;26(6):730-6.

Guilford P, Hopkins J, Harraway J, et al. E-cadherin germline mutations in familial gastric cancer. Nature 1998;392:402e5.

Guilford P, Humar B, Blair V. Hereditary diffuse gastric cancer: translation of *CDH1* germline mutations into clinical practice. Gastric Cancer. 2010 Mar;13(1):1-10. doi: 10.1007/s10120-009-0531-x.

Hackenson D, Edelman DA, McGuire T, Weaver DW, Webber JD (2010) Prophylactic laparo-scopic gastrectomy for hereditary diffuse gastric cancer: a case series in a single family. JSLS 14(3):348–352. doi:10.4293/108680810X12924466007449

Hansford S et al. (2014) Manuscript in preparation.

Harris JK, Sahl JW, Castoe TA, Wagner BD, Pollock DD, Spear JR. Comparison of normalization methods for construction of large, multiplex amplicon pools for next-generation sequencing. Appl Environ Microbiol. 2010 Jun;76(12):3863-8. doi: 10.1128/AEM.02585-09. Epub 2010 Apr 23.

(1) Hearle NC, Rudd MF, Lim W, Murday V, Lim AG, Phillips RK et al. (2006). Exonic *STK11* deletions are not a rare cause of Peutz–Jeghers syndrome. J Med Genet 43: e15.

(2) Hearle N, Schumacher V, Menko FH, Olschwang S, Boardman LA, Gille JJP, et al. (2006). Frequency and spectrum of cancers in the Peutz-Jeghers syndrome. Clinical Cancer Research: an Official Journal of the American Association for Cancer Research, 12(10), 3209–3215. doi:10.1158/1078-0432.CCR-06-0083

Hebbard PC, Macmillan A, Huntsman D, Kaurah P, Carneiro F, Wen X, Kwan A, Boone D, Bursey F, Green J, Fernandez B, Fontaine D, Wirtzfeld DA (2009) Prophylactic total gastrectomy (PTG) for hereditary diffuse gastric cancer (HDGC): the Newfoundland experience with 23 patients. Ann Surg Oncol 16:1890–1895. doi:10.1245/s10434-009-0471-z

Hennekam RC, Biesecker LG (2012) Next-generation sequencing demands next-generation phe-notyping. Hum Mutat 33(5):884–886. doi:10.1002/humu.22048

Hundahl SA, Phillips JL, Menck HR. The National Cancer Data Base Report on poor survival of US gastric carcinoma patients treated with gastrectomy: fifth edition American Joint

Committee on Cancer staging, proximal disease, and the 'different disease' hypothesis. Cancer. 2000;88:921–932.

Huntsman D, Carneiro F, Lewis F, MacLeod P, Hayashi A, Monaghan K, Maung R, Seruca R, Jackson C, Caldas C. Prophylactic gastrectomy in patients with deleterious E-cadherin gene mutation. Gastroenterol Clin Biol. 2001 Oct;25(10):931-2.

Hwang WY, Fu Y, Reyon D, Maeder ML, Tsai SQ, Sander JD, Peterson RT, Yeh JR, Joung JK. 2013. Efficient genome editing in zebrafish using a CRISPR-Cas system. *Nat Biotechnol* 31:227–229.

Ingham SL, Sperrin M, Baildam A, et al. Risk-reducing surgery increases survival in BRCA1/2 mutation carriers unaffected at time of family referral. Breast Cancer Res Treat. 2013 Nov 20.

Iriyama T, Takeda K, Nakamura H, Morimoto Y, Kuroiwa T, et al. (2009) *ASK1* and *ASK2* differentially regulate the counteracting roles of apoptosis and inflammation in tumorigenesis. The EMBO Journal 28: 843–853. doi:10.1038/emboj.2009.32.

Jakubowska A, Nej K, Huzarski T, Scott RJ, Lubiński J. *BRCA2* gene mutations in families with aggregations of breast and stomach cancers. Br J Cancer. 2002 Oct 7;87(8):888-91.

Jakubowska A, Scott R, Menkiszak J et al (2003) A high frequency of *BRCA2* gene mutations in Pol- ish families with ovarian and stomach cancer. Eur J Hum Genet 11:955–958

Janavicius R. Founder BRCA1/2 mutations in the Europe: implications for hereditary breast-ovarian cancer prevention and control. EPMA J. 2010 Sep;1(3):397-412. doi: 10.1007/s13167-010-0037-y. Epub 2010 Jun 27.

Jeanes, A., Gottardi, C. J., & Yap, A. S. (2008). Cadherins and cancer: how does cadherin dysfunction promote tumor progression? Oncogene, 27(55), 6920–6929. doi:10.1038/onc.2008.343

Johannsson O, Loman N, Möller T, Kristoffersson U, Borg A, Olsson H. Incidence of malignant tumors in relatives of BRCA1 and *BRCA2* germline mutation carriers. Eur J Cancer. 1999 Aug;35(8):1248-57.

Jones EG (1964) Familial gastric cancer. N Z Med J 63:287–296

Jones S, Hruban RH, Kamiyama M, et al. Exomic sequencing identifies *PALB2* as a pancreatic cancer susceptibility gene. Science. 2009; 324:217.

Joo, Y.-E., Rew, J.-S., Choi, S.-K., Bom, H.-S., Park, C.-S., & Kim, S.-J. (2002). Expression of e-cadherin and catenins in early gastric cancer. Journal of Clinical Gastroenterology, 35(1), 35–42.

Kang B, Guo RF, Tan XH, et al. Expression status of ataxia-telangiectasia-mutated gene correlated with prognosis in advanced gastric cancer. Mutat Res 2008;638:17–25.

Kaurah et al. (2014). Manuscript in preparation.

Kaurah P, MacMillan A, Boyd N, Senz J, De Luca A, Chun N, Suriano G, Zaor S, Van Manen L, Gilpin C, Nikkel S, Connolly-Wilson M, Weissman S, Rubinstein WS, Sebold C, Greenstein R, Stroop J, Yim D, Panzini B, McKinnon W, Greenblatt M, Wirtzfeld D, Fontaine D, Coit D, Yoon S, Chung D, Lauwers G, Pizzuti A, Vaccaro C, Redal MA, Oliveira C, Tischkowitz M, Olschwang S, Gallinger S, Lynch H, Green J, Ford J, Pharoah P, Fernandez B, Huntsman D (2007) Founder and recurrent *CDH1* mutations in families with hereditary diffuse gastric cancer. JAMA 297(21):2360–2372

Kaurah P1, Fitzgerald R, Dwerryhouse S, Huntsman DG. Pregnancy after prophylactic total gastrectomy. Fam Cancer. 2010 Sep;9(3):331-4. doi: 10.1007/s10689-009-9316-y.

Kempers MJ, Kuiper RP, Ockeloen CW, et al. Risk of colorectal and endometrial cancers in *EPCAM* deletion-positive Lynch syndrome: a cohort study. Lancet Oncol. 2011;12:49–55.

Kim JW, Im SA, Kim MA, Cho HJ, Lee DW, Lee KH, Kim TY, Han SW, Oh DY, Lee HJ, Kim TY, Yang HK, Kim WH, Bang YJ. Ataxia-telangiectasia-mutated protein expression with microsatellite instability in gastric cancer as prognostic marker. Int J Cancer. 2014 Jan 1;134(1):72-80. doi: 10.1002/ijc.28245.

Kluijt, I., Sijmons, R. H., Hoogerbrugge, N., Plukker, J. T., de Jong, D., Van Krieken, J. H., et al. (2012, September). Familial gastric cancer: guidelines for diagnosis, treatment and periodic surveillance. Familial Cancer. doi:10.1007/s10689-012-9521-y

Koslov ER, Maupin P, Pradhan D, Morrow JS, Rimm DL. Alpha-catenin can form asymmetric homodimeric complexes and/or heterodimeric complexes with beta-catenin. J Biol Chem. 1997 Oct 24;272(43):27301-6.

Kote-Jarai Z, Leongamornlert D, Saunders E, Tymrakiewicz M, Castro E, Mahmud N, Guy M, Edwards S, O'Brien L, Sawyer E, Hall A, Wilkinson R, Dadaev T, Goh C, Easton D; UKGPCS Collaborators, Goldgar D, Eeles R. *BRCA2* is a moderate penetrance gene contributing to young-onset prostate cancer: implications for genetic testing in prostate cancer patients. Br J Cancer. 2011 Oct 11;105(8):1230-4. doi: 10.1038/bjc.2011.383.

Ku CS, Naidoo N, Pawitan Y (2011) Revisiting Mendelian disorders through exome sequencing. Hum Genet 129(4):351–370. doi:10.1007/s00439-011-0964-2

Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat Protoc. 2009;4(7):1073-81.

La Vecchia C, Negri E, Franceschi S, et al. Family history and the risk of stomach and colorectal cancer. Cancer. 1992;70:50–55.

Lange K, Weeks D, Boehnke M. Programs for Pedigree Analysis: MENDEL, FISHER, and dGENE. Genet Epidemiol. 1988;5(6):471-2.

Lauren P. The two histological main types of gastric carcinoma: diffuse and so- called intestinal-type carcinoma. an attempt at a histo-clinical classification. Acta Pathol Microbiol Scand 1965;64:31e49.

Lee AF, Rees H, Owen DA, Huntsman DG. Periodic acid-schiff is superior to hematoxylin and eosin for screening prophylactic gastrectomies from CDH1 mutation carriers. Am J Surg Pathol. 2010 Jul;34(7):1007-13. doi: 10.1097/PAS.0b013e3181e28985.

Li H, Handsaker B, Wysoker A, et al. The Sequence alignment/map (SAM) format and SAMtools. Bioinformatics. 2009;25, 2078-9.

Li H. and Durbin R. Fast and accurate short read alignment with Burrows-Wheeler Transform. Bioinformatics. 2009;25:1754-60.

Li, Y.-J., Meng, Y.-X., & Ji, X.-R. (2003). Relationship between expressions of E-cadherin and alpha-catenin and biological behaviors of human pancreatic cancer. Hepatobiliary & Pancreatic Diseases International : HBPD INT, 2(3), 471–477.

Lynch HT, Lynch JF, Snyder CL, Riegert-Johnson D. *EPCAM* deletions, Lynch syndrome, and cancer risk. The Lancet Oncology, Volume 12, Issue 1, January 2011, Pages 5-6

Lynch HT, Smyrk TC, Watson P et al (1993) Genetics, natural history, tumor spectrum, and pathology of hereditary nonpolyposis colorectal cancer: an up- dated review. Gastroenterology 104:1535–1549

Lynch HT, Smyrk TC, Watson P et al (1993) Genetics, natural history, tumor spectrum, and pathology of hereditary nonpolyposis colorectal cancer: an up- dated review. Gastroenterology 104:1535–1549

Maier C, Vesovic Z, Bachmann N, et al. Germline mutations of the *MSR1* gene in prostate cancer families from Germany. Hum Mutat. 2006 Jan;27(1):98-102.

Marmorstein LY, Ouchi T, Aaronson SA. The *BRCA2* gene product functionally interacts with p53 and RAD51. Proc. Natl. Acad. Sci. USA, 95 (1998), pp. 13869–13874

Marsh DJ, Coulon V, Lunetta KL, Rocca-Serra P, Dahia PLM, Zheng Z, Liaw D, Caron S, Duboue B, Lin AY, Richardson AL, Bonnetblanc JM, et al. Mutation spectrum and genotype-phenotype analyses in Cowden disease and Bannayan-Zonana syndrome, two hamartoma syndromes with germline PTEN mutation. Hum. Molec. Genet. 7: 507-515, 1998.

Masciari S., Larsson N., Senz J., Boyd N., Kaurah P., Kandel M.J., Harris L.N., Pinheiro H.C., Troussard A., Miron P., et al. Germline E-cadherin mutations in familial lobular breast cancer. J. Med. Genet. 2007;44:726–731.

Mateus, A. R., Simões-Correia, J., Figueiredo, J., Heindl, S., Alves, C. C., Suriano, G., et al. (2009). E-cadherin mutations and cell motility: a genotype-phenotype correlation. Experimental Cell Research, 315(8), 1393–1402. doi:10.1016/j.yexcr.2009.02.020

Matsuura, K., Kawanishi, J., Fujii, S., Imamura, M., Hirano, S., Takeichi, M., & Niitsu, Y. (1992). Altered expression of E-cadherin in gastric cancer tissues and carcinomatous fluid. British Journal of Cancer, 66(6), 1122–1130.

McKinnon, Naud, Ashikaga, Colletti, and Wood. Results of an Intervention for Individuals and Families with BRCA Mutations: A Model for Providing Medical Updates and Psychosocial Support Following Genetic Testing. Journal of Genetic Counseling, Vol. 16, No. 4, (2007) DOI: 10.1007/s10897-006-9078-8

Meldrum C, Doyle MA, Tothill RW (2011) Next-generation sequencing for cancer diagnostics: a practical perspective. Clin Biochem Rev 32(4):177–195

Moran, A., O'Hara, C., Khan, S., Shack, L., Woodward, E., Maher, E. R., et al. (2012). Risk of cancer other than breast or ovarian in individuals with BRCA1 and *BRCA2* mutations. Familial Cancer, 11(2), 235–242. doi:10.1007/s10689-011-9506-2

Morrogh M, Andrade VP, Giri D, et al. Cadherin-catenin complex dissociation in lobular neoplasia of the breast. Breast Cancer Res Treat. 2012 Apr;132(2):641-52. doi: 10.1007/s10549-011-1860-0. Epub 2011 Nov 13.

Myllykangas S, Buenrostro J, Ji HP (2012) Overview of sequencing technology platforms. In: Rodríguez-Ezpeleta N, Hackenberg M, Aransay AM (eds) Bioinformatics for high throughput sequencing. Springer, New York. doi:10.1007/978-1-4614-0782-9_2

Nakopoulou L, Gakiopoulou-Givalou H, Karayiannakis AJ, et al. Abnormal alpha-catenin expression in invasive breast cancer correlates with poor patient survival. Histopathology. 2002 Jun;40(6):536-46.

Nam JH, Choi IJ, Cho SJ, Kim CG, Jun JK, Choi KS, Nam BH, Lee JH, Ryu KW, Kim YW (2012) Association of the interval between endoscopies with gastric cancer stage at diagnosis in a region of high prevalence. Cancer 118:4953–4960. doi:10.1002/cncr.27495

Narod, S. A., & Offit, K. (2005). Prevention and management of hereditary breast cancer. J Clin Oncol, 23, 1656–1663.

Ng PC, Henikoff S. Predicting the Effects of Amino Acid Substitutions on Protein Function Annu Rev Genomics Hum Genet. 2006;7:61-80.

Ng PC, Henikoff S. SIFT: predicting amino acid changes that affect protein function. Nucleic Acids Res. 2003 Jul 1;31(13):3812-4.

Ng PC, Kirkness EF (2010) Whole genome sequencing. Methods Mol Biol 628:215–226

Ni Y, Zbuk KM, Sadler T, et al. Germline mutations and variants in the succinate dehydrogenase genes in Cowden and Cowden-like syndromes. Am J Hum Genet. 2008 Aug;83(2):261-8. doi: 10.1016/j.ajhg.2008.07.011.

Nigam, A. K., Savage, F. J., Boulos, P. B., Stamp, G. W., Liu, D., & Pignatelli, M. (1993). Loss of cell-cell and cell-matrix adhesion molecules in colorectal cancer. British Journal of Cancer, 68(3), 507–514.

Norton JA, Ham CM, Van Dam J, et al. *CDH1* truncating mutations in the E-cadherin gene: an indication for total gastrectomy to treat hereditary diffuse gastric cancer. Ann Surg 2007;245:873e9.

Nozawa H, Oda E, Ueda S, Tamura G, Maesawa C, Muto T, Taniguchi T, Tanaka N (1998) Functionally inactivating point mutation in the tumor-suppressor IRF-1 gene identified in human gastric cancer. Int J Cancer 77(4):522–527

Oka, H., Shiozaki, H., Kobayashi, K., Inoue, M., Tahara, H., Kobayashi, T., et al. (1993). Expression of E-cadherin cell adhesion molecules in human breast cancer tissues and its relationship to metastasis. Cancer Research, 53(7), 1696–1701.

(1) Oliveira C, Senz J, Kaurah P. Germline *CDH1* deletions in hereditary diffuse gastric cancer families. Hum Mol Genet. 2009 May;18(9):1545-55.

(2) Oliveira C, Sousa S, Pinheiro H, et al. Quantification of epigenetic and genetic 2nd hits in *CDH1* during hereditary diffuse gastric cancer syndrome progression. Gastroenterology. 2009 Jun;136(7):2137-48.

Oliveira, C, Pinheiro, H., Figueiredo, J., Seruca, R., & Carneiro, F. (2013). E-cadherin alterations in hereditary disorders with emphasis on hereditary diffuse gastric cancer. Progress in Molecular Biology and Translational Science, 116, 337–359. doi:10.1016/B978-0-12-394311-8.00015-7

Orloff M, Peterson C, He X, et al. Germline mutations in *MSR1*, ASCC1, and CTHRC1 in patients with Barrett esophagus and esophageal adenocarcinoma. JAMA. 2011 Jul 27;306(4):410-9. doi: 10.1001/jama.2011.1029.

Orloff, M. S., He, X., Peterson, C., Chen, F., Chen, J.-L., Mester, J. L., Eng, C. Germline *PIK3CA* and *AKT1* mutations in Cowden and Cowden-like syndromes. Am. J. Hum. Genet. 92: 76-80, 2013.

Palli D, Russo A, Decarli A (2001) Dietary patterns, nutrient intake and gastric cancer in a high-risk area of Italy. Cancer Causes Control 12(2):163–172

Pandalai PK, Lauwers GY, Chung DC, Patel D, Yoon SS (2011) Prophylactic total gastrectomy for individuals with germline *CDH1* mutation. Surgery 149(3):347–355. doi:10.1016/j.surg. 2010.07.005

Pannone G, Santoro A, Feola A, Bufo P, Papagerakis P, Lo Muzio L, Staibano S, Ionna F, Longo F, Franco R, Aquino G, Contaldo M, De Maria S, Serpico R, De Rosa A, Rubini C, Papagerakis S, Giovane A, Tombolini V, Giordano A, Caraglia M, Di Domenico M. The Role Of E-Cadherin Down-Regulation In Oral Cancer: *CDH1* Gene Expression And Epigenetic Blockage. Curr Cancer Drug Targets. 2013 Nov 25. [Epub ahead of print]

Park D, Kåresen R, Axcrona U, Noren T, Sauer T. Expression pattern of adhesion molecules (E-cadherin, alpha-, beta-, gamma-catenin and claudin-7), their influence on survival in primary breast carcinoma, and their corresponding axillary lymph node metastasis. APMIS. 2007 Jan;115(1):52-65.

Park WS, Lee JH, Shin MS, Park JY, Kim HS, Lee JH, Kim YS, Lee SN, Xiao W, Park CH, Lee SH, Yoo NJ, Lee JY. Inactivating mutations of the caspase-10 gene in gastric cancer. Oncogene 21: 2919-2925, 2002.

Pharoah PD, Guilford P, Caldas C. Incidence of gastric cancer and breast cancer in *CDH1* (E-cadherin) mutation carriers from hereditary diffuse gastric cancer families. Gastroenterology 2001;121:1348e53.

Pinheiro H, Bordeira-Carriço R, Seixas S, Carvalho J, Senz J, Oliveira P, Inácio P, Gusmão L, Rocha J, Huntsman D, Seruca R, Oliveira C. Allele-specific *CDH1* down regulation and hereditary diffuse gastric cancer. Hum Mol Genet. 2010 Mar;19(5):943-52.

Pinheiro H, Carvalho J, Oliveira P. Transcription initiation arising from E-cadherin/*CDH1* intron2: a novel protein isoform that increases gastric cancer cell invasion and angiogenesis. Hum Mol Genet. 2012 Oct;21(19):4253-69.

Pomraning KR, Smith KM, Bredeweg EL, Connolly LR, Phatale PA, Freitag M (2012) Library preparation and data analysis packages for rapid genome sequencing. Methods Mol Biol 944:1–22. doi:10.1007/978-1-62703-122-6_1

Raffan E, Semple RK (2011) Next generation sequencing–implications for clinical practice. Br Med Bull 99:53–71. doi:10.1093/bmb/ldr02

Rahman N, Seal S, Thompson D, Kelly P, Renwick A, Elliott A, Reid S, Spanova K, Barfoot R, Chagtai T et al.: *PALB2*, which encodes a *BRCA2*-interacting protein, is a breast cancer susceptibility gene. Nat Genet 2007, 39:165-167.

Ratheesh A. & Yap AS. (2012) A bigger picture: classical cadherins and the dynamic actin cytoskeleton. Nat Rev Mol Cell Biol 13(10):673–679. doi:10.1038/nrm3431

Renwick A, Thompson D, Seal S, Kelly P, Chagtai T, et al. (2006) *ATM* mutations that cause ataxia-telangiectasia are breast cancer susceptibility alleles. Nat Genet 38: 873–875.

Renwick, A., Thompson, D., Seal, S., Kelly, P., Chagtai, T., Ahmed, M., North, B., Jayatilake, H., Barfoot, R., Spanova, K., McGuffog, L., Evans, D. G., Eccles, D., The Breast Cancer Susceptibility Collaboration (UK), Easton, D. F., Stratton, M. R., Rahman, N. *ATM* mutations that cause ataxia-telangiectasia are breast cancer susceptibility alleles. Nature Genet. 38: 873-875, 2006.

Risch HA, McLaughlin JR, Cole DE, et al. Prevalence and penetrance of germline BRCA1 and *BRCA2* mutations in a population series of 649 women with ovarian cancer. Am J Hum Genet. 2001 Mar;68(3):700-10. Epub 2001 Feb 15.

Rocco A, Schandl L, Chen J, Wang H, Tulassay Z, McNamara D, Malfertheiner P, Ebert MP (2003) Loss of FHIT protein expression correlates with disease progression and poor differentiation in gastric cancer. J Cancer Res Clin Oncol 129(2):84–88

Rogers WM, Dobo E, Norton JA, Van Dam J, Jeffrey RB, Huntsman DG, Kingham K, Chun N, Ford JM, Longacre TA (2008) Risk-reducing total gastrectomy for germline mutations in E-cadherin (*CDH1*): pathologic findings with clinical implications. Am J Surg Pathol 32:799–809. doi:10.1097/PAS.0b013e31815e7f1a

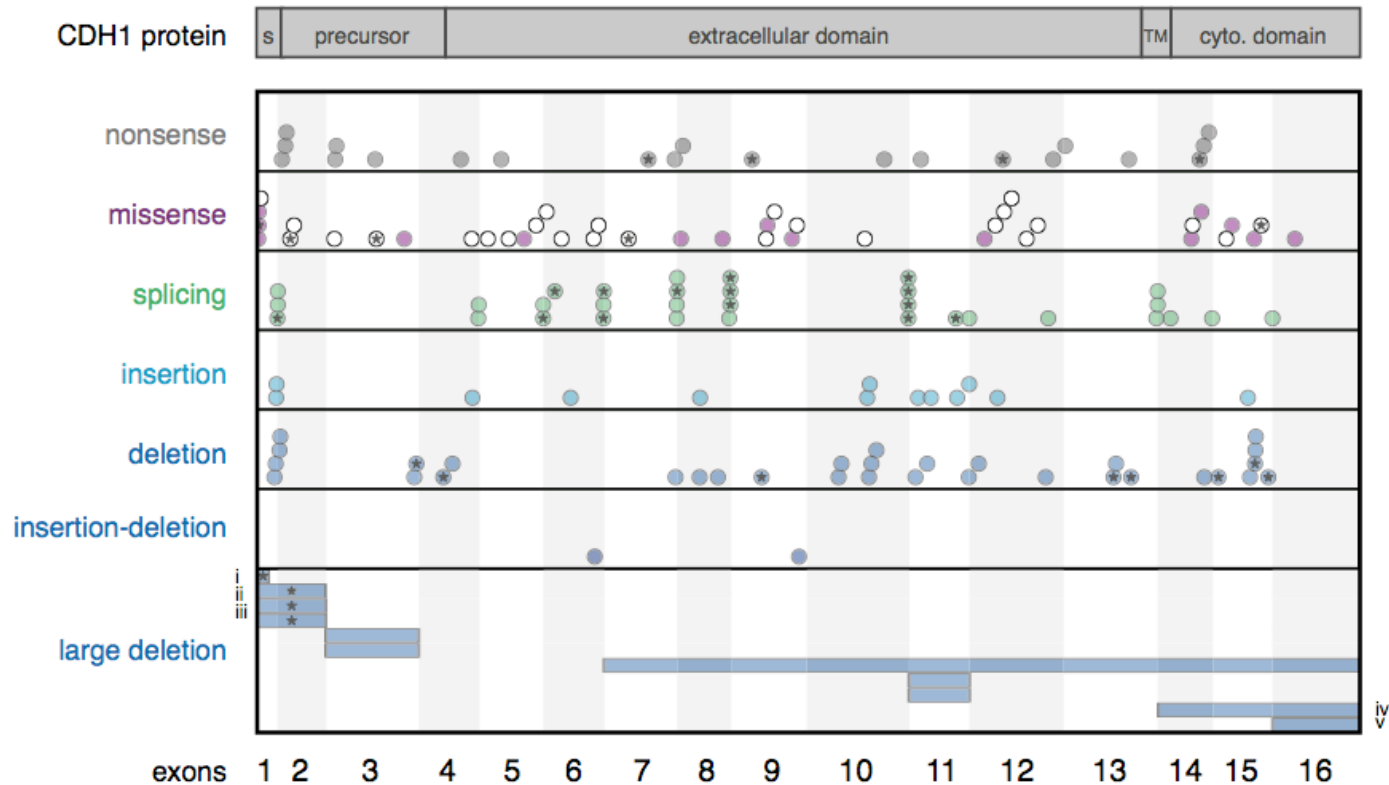Rotman G and Shiloh Y. *ATM*: from gene to function. Hum Mol Genet. 1998;7(10):1555-63.

Sanger F (1988) Sequences, sequences, and sequences. Annu Rev Biochem 57:1–28. doi:10.1146/annurev.bi.57.070188.000245

Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci 74:5463–5467

Savitsky, K., Bar-Shira, A., Gilad, S., Rotman, G., Ziv, Y., Vanagaite, L., Tagle, D. A., Smith, S., Uziel, T., Sfez, S., Ashkenazi, M., Pecker, I., and 18 others. A single ataxia telangiectasia gene with a product similar to PI-3 kinase. Science 268: 1749-1753, 1995.

Schrader KA, Heravi-Moussavi A, Waters PJ, Senz J, Whelan J, Ha G, Eydoux P, Nielsen T, Gallagher B, Oloumi A, Boyd N, Fernandez BA, Young TL, Jones SJ, Hirst M, Shah SP, Marra MA, Green J, Huntsman DG (2011) Using next-generation sequencing for the diagnosis of rare disorders: a family with retinitis pigmentosa and skeletal abnormalities. J Pathol 225:12–18

Schrader KA, Masciari S, Boyd N, et al. Germline mutations in *CDH1* are infrequent in women with early-onset or familial lobular breast cancers. J Med Genet. 2011 Jan;48(1):64-8. doi: 10.1136/jmg.2010.079814. Epub 2010 Oct 4.

Shah S, Morin RD, Khattra J, et al. Mutational evolution in a lobular breast tumour profiled at single nucleotide resolution. Nature. 2009;461, 809-813.

Sharer N, Schwarz M, Malone G, Howarth A, Painter J, Super M, Braganza J. Mutations of the cystic fibrosis gene in patients with chronic pancreatitis. N Engl J Med. 1998 Sep 3; 339(10):645-52.

Shen L, Yin Z-H, Wan Y, Zhang Y, Li K, & Zhou B-S (2012). Association between *ATM* polymorphisms and cancer risk: a meta-analysis. Molecular Biology Reports, 39(5), 5719–5725. doi:10.1007/s11033-011-1381-2

Shiloh, Y. (2003). *ATM* and related protein kinases: safeguarding genome integrity. Nature Reviews. Cancer, 3(3), 155–168. doi:10.1038/nrc1011

Shimoyama, Y., & Hirohashi, S. (1991). Expression of E- and P-cadherin in gastric carcinomas. Cancer Research, 51(8), 2185–2192.

Shinmura K, Goto M, Tao H et al (2005) A novel *STK11* germline mutation in two siblings with Peu- tz-Jeghers syndrome complicated by primary gas- tric cancer. Clin Genet 67:81–86

Simões-Correia, J., Figueiredo, J., Oliveira, C., van Hengel, J., Seruca, R., van Roy, F., & Suriano, G. (2008). Endoplasmic reticulum quality control: a new mechanism of E-cadherin regulation and its implication in cancer. Human Molecular Genetics, 17(22), 3566–3576. doi:10.1093/hmg/ddn249

Slater EP, Langer P, Niemczyk E, et al. *PALB2* mutations in European familial pancreatic cancer families. Clin Genet. 2010;78:490- 494.

Smith AM, Heisler LE, St Onge RP, Farias-Hesson E, Wallace IM, Bodeau J, Harris AN (2010) Highly-multiplexed barcode sequencing: an efficient method for parallel analysis of pooled samples. Nucleic Acids Res 38(13):e142

Southey MC, Teo ZL, Dowty JG, et al. A *PALB2* mutation associated with high risk of breast cancer. Breast Cancer Res. 2010;12: R109.

Suriano G, Yew S, Ferreira P, Senz J, Kaurah P, Ford JM, Longacre TA, Norton JA, Chun N, Young S, Oliveira MJ, Macgillivray B, Rao A, Sears D, Jackson CE, Boyd J, Yee C, Deters C, Pai GS, … Gronberg H, Gallinger S, Seruca R, Lynch H, Huntsman DG. Characterization of a recurrent germ line mutation of the E-cadherin gene: implications for genetic testing and clinical management. Clin Cancer Res. 2005 Aug 1;11(15):5401-9.

Takahashi M, Sakayori M, Takahashi S et al (2004) A novel germline mutation of the LKB1 gene in a pa- tient with Peutz-Jeghers syndrome with early-on- set gastric cancer. J Gastroenterol 39:1210–1214

Takeda K, Shimozono R, Noguchi T, Umeda T, Morimoto Y, et al. (2007) Apoptosis signal-regulating kinase (ASK) 2 functions as a mitogen-activated protein kinase kinase kinase in a heteromeric complex with ASK1. J Biol Chem 282: 7522–7531. doi:10.1074/jbc.M607177200.

Thompson, D., Duedal, S., Kirner, J., McGuffog, L., Last, J., Reiman, A., Byrd, P., Taylor, M., Easton, D. F. Cancer risks and mortality in heterozygous *ATM* mutation carriers. J. Nat. Cancer Inst. 97: 813-822, 2005.

Tulinius H, Olafsdottir GH, Sigvaldason H, et al. The effect of a single *BRCA2* mutation on cancer in Iceland. J Med Genet. 2002 Jul;39(7):457-62.

Umbas, R., Schalken, J. A., Aalders, T. W., Carter, B. S., Karthaus, H. F., Schaafsma, H. E., et al. (1992). Expression of the cellular adhesion molecule E-cadherin is reduced or absent in high-grade prostate cancer. Cancer Research, 52(18), 5104–5109.

Van Roy, F. & Berx, G. (2008). The cell-cell adhesion molecule E-cadherin. Cellular and Molecular Life Sciences : CMLS, 65(23), 3756–3788. doi:10.1007/s00018-008-8281-1

Varley JM, McGowan G, Thorncroft M et al (1995) An extended Li–Fraumeni kindred with gastric car- cinoma and a codon 175 mutation in TP53. J Med Genet 32:942–945

Volikos E, Robinson J, Aittomäki K, et al. LKB1 exonic and whole gene deletions are a common cause of Peutz-Jeghers syndrome. J Med Genet. 2006 May;43(5):e18.

Wainberg, S., & Husted, J. (2004). Utilization of screening and preventive surgery among unaffected carriers of a BRCA1 or *BRCA2* gene mutation. Can Epidemiol Biomarkers Prev, 13(12), 1989–1995.

Wang JL, Yang X, Xia K, Hu ZM, Weng L, Jin X, Jiang H, Zhang P, Shen L, Guo JF, Li N, Li YR, Lei LF, Zhou J, Du J, Zhou YF, Pan Q, Wang J, Wang J, Li RQ, Tang BS (2010) TGM6 identi- fied as a novel causative gene of spinocerebellar ataxias using exome sequencing. Brain 133(Pt 12):3510–3518

Wang L, McDonnell SK, Cunningham JM, et al. No association of germline alteration of *MSR1* with prostate cancer risk. Nat Genet. 2003;35(2):128–129.

Xu J, Sauvageot J, Ewing CM, et al. Germline ATBF1 mutations and prostate cancer risk. Prostate. 2006;66(10):1082–1085.

Xu J, Zheng SL, Komiya A, et al. Germline mutations and sequence variants of the macrophage scavenger receptor-1 gene are associated with prostate cancer risk. Nat Genet. 2002;32(2):321–325.

Yi TZ, Guo J, Zhou L, Chen X, Mi RR, Qu QX, Zheng JH, Zhai L. Prognostic value of E-cadherin expression and *CDH1* promoter methylation in patients with endometrial carcinoma. Cancer Invest. 2011 Jan;29(1):86-92. doi: 10.3109/07357907.2010.512603. Epub 2010 Sep 27.

Yoon KA, Ku JL, Choi HS, et al. Germline mutations of the *STK11* gene in Korean Peutz-Jeghers syndrome patients. Br J Cancer. 2000 Apr;82(8):1403-6.

Yoshida K, Miki Y. Role of BRCA1 and *BRCA2* as regulators of DNA repair, transcription, and cell cycle in response to DNA damage. Cancer Sci. 2004 Nov;95(11):866-71.

Yoshida-Noro, C., Suzuki, N., & Takeichi, M. (1984). Molecular nature of the calcium-dependent cell-cell adhesion system in mouse teratocarcinoma and embryonic cells studied with a monoclonal antibody. Developmental Biology, 101(1), 19–27.

Zanghieri G, Di Gregorio C, Sacchetti C, Fante R, Sassatelli R, Cannizzo G, Carriero A, Ponz de Leon M. Familial occurrence of gastric cancer in the 2-year experience of a population-based registry. Cancer. 1990 Nov 1;66(9):2047-51.

Zhang L, Jia G, Li W-M, Guo R-F, Cui J-T, Yang L, & Lu Y-Y. (2004). Alteration of the *ATM* gene occurs in gastric cancer cell lines and primary tumors associated with cellular response to DNA damage. Mutation Research, 557(1), 41–51.

Appendix 1. Mapping of Germline CDH1 Mutations Described To-Date by Mutation Type and Location on Transcript/Protein (Kaurah et al. Manuscript in preparation)