UNIVERSITA' DEGLI STUDI DI PAVIA

Dipartimento di Biologia e Biotecnologie

"*L. Spallanzani*"

New molecular players in gastric carcinogenesis

Raefa Abou Khouzam

Dottorato di Ricerca in

Genetica, Biologia Molecolare e Cellulare

XXIX Ciclo – A.A. 2013-2016

UNIVERSITA' DEGLI STUDI DI PAVIA

Dipartimento di Biologia e Biotecnologie

"*L. Spallanzani*"

New molecular players in gastric carcinogenesis

Raefa Abou Khouzam

Supervised by Prof. Guglielmina Nadia Ranzani

Dottorato di Ricerca in
Genetica, Biologia Molecolare e Cellulare
XXIX Ciclo – A.A. 2013-2016

Table of contents

Abstract

Gastric cancer (GC) is a global health burden affecting nearly one million people per year worldwide. It is a heterogeneous disease with two major histological subtypes "intestinal" (IGC) and "diffuse" (DGC) varying in terms of both clinic-pathological profiles and molecular pathogenesis.
The differential contribution of environmental and genetic risk factors results in a GC profile in which the majority of cases occur sporadically, while only 1-3% are strictly hereditary and dominated by the Hereditary Diffuse Gastric Cancer (HDGC) syndrome. HDGC is an autosomal dominantly inherited cancer predisposition that is associated with a high lifetime risk for diffuse-type gastric cancer, coupled with an increased risk of lobular breast cancer in women. It is an aggressive disease that is primarily explained by germline alterations in the *CDH1* tumor suppressor gene. Despite the efforts, at least 50% of HDGC patients remain without a known genetic determinant. This large fraction of cases might be attributed to mechanisms of *CDH1* inactivation that are yet to be accounted for. Indeed, it has been recently demonstrated that intron 2 of the *CDH1* gene gives rise to a number of non-canonical transcripts, namely *CDH1-a*, *-1b*, *-j*, *-1b10*, none of which have been investigated in GC patients. Alternatively, the missing genetic heritability causative of HDGC could be distributed along other loci. In fact, the application of exome sequencing on such cases has recently unraveled germline lesions in diverse unexpected genes that could predispose to the disease.
Therefore, in the first part of my research, I set out to better characterize a cohort of 30 Italian HDGC patients already proven to be *CDH1* mutation-negative by a series of comprehensive molecular approaches. To that purpose, I started by amplifying the *CDH1* canonical transcript and the *CDH1a* and *CDH1j* non-canonical transcripts in the blood of the HDGC patients and normal controls in search for expression defects. In all cases the same expression patterns were detected, thus dismissing the involvement of *CDH1* defective splicing incurred by germline exonic or intronic variants in disease manifestation. Subsequently, as part of a much larger study in collaboration with IRST-IRCCS (Istituto Scientifico Romagnolo per lo Studio e la Cura dei Tumori Srl Istituto di Ricovero e Cura a Carattere Scientifico) of Meldola (FC-Italy), the 7 most recently recruited cases underwent exome sequencing of a panel of 94 cancer-related genes. I then performed an *in silico* analysis on all sub-polymorphic and rare non-synonymous variants arising from sequencing to uncover whether other genes could be causative of HDGC in the investigated cohort. Upon combining the results of the *in silico* analysis with supportive literature findings regarding the affected genes, two novel candidates for HDGC, namely *CDKN2A* and *SDHC*, could be put forth.
The exome sequencing results presented here highlight the impressive aspect of applying next generation sequencing on patients selected for a specific disease. In

particular, the emergence of unsuspected candidate genes that could somehow explain the disease in question opens the door to its better characterization. This, in turn, would aid in the identification of at risk individuals in affected families who are negative for *CDH1* mutations, thereby improving their counseling and management.

The *CDH1* gene encodes E-cadherin, a calcium-dependent transmembrane adhesion protein that plays multiple crucial roles in epithelial cell adhesion, survival, proliferation and migration. Functional loss of E-cadherin is a well-established molecular event that occurs during gastric cancer progression, leading to increased invasion of cancer cells to neighboring tissues and to metastasis. Despite the absence of *CDH1* germline mutations in patients with IGC, somatic alterations encompassing *CDH1* have been reported in both diffuse and intestinal sporadic gastric cancer cases. Indeed, the majority of GCs show an immunohistochemical aberrant pattern of E-cadherin expression, while complete protein loss is highly predominant in GCs of the diffuse histotype. Different mechanisms have been implicated in the negative regulation of *CDH1*, indicating the existence of factors acting at diverse levels that can either abrogate or subtly modulate E-cadherin expression in gastric cancer. Among those factors, are small regulatory RNAs (micro-RNAs); in addition, an intron-mediated mechanism of *CDH1* regulation has also been identified. In particular, intron 2, harboring an exceptionally high number of repetitive elements involved in exonization, can act as a *cis*-modulator of E-cadherin gene and protein expression. Among the novel non-canonical transcripts arising from this intron, *CDH1a* harbors properties that enable its translation into a protein isoform differing from the canonical E-cadherin in its N-terminal domain. Functional assays performed on gastric cancer transfected cells associated the overexpression of *CDH1a* with increased angiogenesis and invasion in the presence of the canonical transcript. These findings make *CDH1* gene transcripts likely players in gastric carcinogenesis of the intestinal type, where some level of E-cadherin expression is often retained.
On that basis, in the second part of my research I evaluated *CDH1* canonical and non-canonical transcripts, as well as *CDH1*-regulating miRNAs, in intestinal-type gastric cancer to provide a more holistic picture of *CDH1* deregulation in IGC. I initiated the study by setting up the optimized conditions in digital-PCR (dPCR) for the quantification of *CDH1* and *CDH1a* transcripts in tumor and normal tissue samples derived from 32 patients with IGC. dPCR was utilized since it is a sensitive method for detection of gene expression in settings where the target RNA is limited or present in quantity that approaches the limits of quantitative-PCR (qPCR) sensitivity. Following the application of this fine-tuned technique, I found a significant decrease in *CDH1* expression in tumors compared to the normal counterparts (P=0.001), which was especially evident in 76% of cases. In addition, *CDH1a* was detected at extremely low levels in 47% of tumors, but not in normal

mucosa. A trend was also observed of having more *CDH1* in tumors lacking the *CDH1a* transcript. Further investigations on the mechanistic effect of the presence of *CDH1a* in IGC tumors revealed that, despite what has been previously reported for GC cell lines, *CDH1a* did not associate with the expression levels of interferon inducible genes in IGC tissue. Regarding the evaluation of *CDH1*-regulating miRNAs, I first applied an *in silico* and literature driven approach to determine the miRNAs most likely to be involved in IGC. As a result, a list of 16 miRNAs that could potentially directly or indirectly regulate *CDH1* in IGC was produced. Two miRNAs from this list, *miR-92a* and *miR-101*, have been quantified thus far by reverse transcription qPCR. While neither miRNA correlated with the expression of *CDH1*, *miR-101* was found to be significantly lower in tumors compared to normal gastric mucosa (P=1.565x10^{-05}), which points towards its implication in gastric carcinogenesis in ways that surpass *CDH1* regulation.

Taken together, our results support the notion that abnormal isoforms and transcripts' imbalance resulting from cryptic abnormalities along the *CDH1* locus, although subtly modulating E-cadherin expression, can still contribute to the carcinogenic process of the intestinal type.

Abbreviations

ACRG	Asian Cancer Research Group
AI	Allelic imbalance
aa	Amino acids
BLAST	Basic Local Alignment Search Tool
BQSR	Base Quality Score Recalibration
CagA	Cytotoxin-associated gene A
CAGE	5' cap analysis gene expression
CBD	β-catenin binding domain
CCC	Cadherin-catenin complex
CDS	Coding sequence
CHO	Chinese hamster ovary
CIN	Chromosomal instability
COX-2	Cyclooxygenase-2
CSV	Comma-separated values
CTCF	CCCTC-binding factor
CV	Coefficient of variation
ddPCR	Droplet digital PCR
DGC	Diffuse gastric cancer
dMIQE	Minimum Information for Publication of Quantitative Real-Time dPCR Experiments
dPCR	Digital-PCR
EBV	Epstein - Barr virus
EC	Extracellular Cadherin repeat
ECM	Extracellular matrix
EMT	Epithelial-to-Mesenchymal Transition
EODGC	Early Onset Diffuse Gastric Cancer
EPLIN	Epithelial Protein Lost In Neoplasm
ESCC	Esophageal squamous cell carcinoma
ESP6500	Exome Sequencing Project
EST	Expressed Sequence Tags
ExAC	Exome Aggregation Consortium
EZH2	Enhancer of Zeste Homolog 2
FAM	Fluorescein
FAMMM	Familial atypical mole multiple melanoma
FAP	Familial Adenomatous Polyposis
FDA	Food and Drug Administration
FDGC	Familial Diffuse Gastric Cancer
FFPE	Formalin-fixed paraffin-embedded
FGC	Familial Gastric Cancer
FIGC	Familial Intestinal Gastric Cancer
GAPPS	Gastric Adenocarcinoma and Proximal Polyposis of the Stomach
GATK	Genome Analysis Toolkit
GC	Gastric cancer
GEJ	Gastroesophageal junction
GI	Gastrointestinal
GIST	Gastrointestinal stromal tumors
GS	Genomically stable
GWAS	Genome wide association studies
H. pylori	*Helicobacter pylori*
HDGC	Hereditary Diffuse Gastric Cancer
HNPCC	Hereditary NonPolyposis Colorectal Cancer

HtrA	High-temperature requirement A
IDT	Integrated DNA Technologies
IGC	Intestinal gastric cancer
IGCLC	International Gastric Cancer Linkage Consortium
IGV	Integrative Genomics Viewer
IHC	Immunohistochemistry
Indels	Insertions deletions
JMD	Juxtamembrane domain
JPS	Juvenile Polyposis syndrome
LBC	Lobular breast cancer
LFS	Li-Fraumeni Syndrome
LOH	Loss of heterozygosity
LPS	Lipopolysaccharide
mAb	Monoclonal antibody
MAPK	Mitogen-activated protein kinase
MBL2	Mannose binding lectin
MET	Mesenchymal-to-Epithelial Transition
miRNA	microRNA
MLPA	Multiplex Ligation-dependent Probe Amplification
MPED	Membrane Proximal Extracellular Domain
MSI	Microsatellite instability
MSS	Microsatellite stable
MTC	Medullary thyroid carcinoma
MTIs	microRNA-target interactions
ncRNA	Noncoding RNA
NER	Nucleotide excision repair
NTC	No-template controls
PAI	Pathogenicity island
PC	Pancreatic cancer
PCR	Polymerase chain reaction
PGL	Paraganglioma
PHEO	Phaeochromocytoma
PJS	Peutz-Jeghers Syndrome
PRC2	Polycomb Repressive Complex 2
PRCs	Polycomb Repressive Complexes
rRNA	Ribosomal RNA
RTK	Receptor tyrosine kinase
RT-PCR	Reverse transcription PCR
RT-qPCR	Reverse transcription quantitative PCR
sCNAs	Somatic copy number alterations
SNP	Single nucleotide polymorphism
SNuPE	Single-Nucleotide Primer Extension
SNVs	Single Nucleotide Variants
SUZ12	Suppressor of Zeste 12 homolog
SVM	Support vector machine
TAMRA	Carboxytetra-methylrhodamine
TCGA	The Cancer Genome Atlas
TLR4	Toll-like receptor 4
TNF-A	Tumor necrosis factor
TSG	Tumor suppressor gene
TSS	Transcription start site
UCSC	University of California, Santa Cruz
UNG	Uracil-N glycosylase
VacA	Vacuolating cytoxin A
VUSs	Variants of Uncertain Significance
WHO	World Health Organization
YY1	Ying Yang 1

Introduction

Gastric cancer (GC) most commonly manifests in the innermost layer of the stomach in the form of an adenocarcinoma. It is a heterogeneous disease at the epidemiological, molecular and histopathological levels, with surgical resection being the only curative treatment.

1. Epidemiology of gastric cancer

GC is the fifth most commonly occurring cancer and the third most common cause of cancer related deaths, in both sexes, worldwide (Ferlay J *et al*., 2015). In 2012 around 140,000 people in Europe were diagnosed with this disease and more than three quarters of the afflicted died (Ferlay J *et al*., 2013). It is quite clear that sex and geographical origins significantly influence GC incidence and mortality rates (Figure 1).

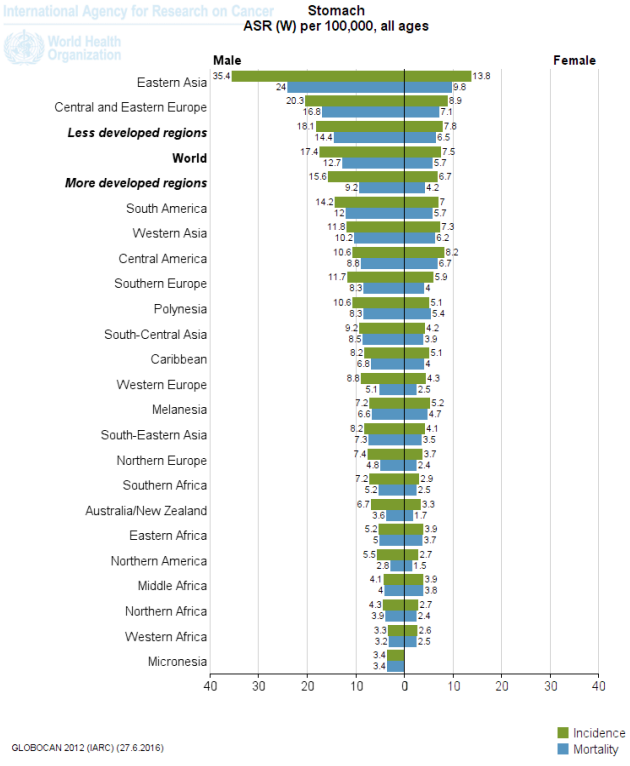

Figure 1. GC incidence and mortality rates from GLOBOCAN 2012. Numbers refer to the estimated age-standardized rates worldwide, ASR (W), per 100,000 individuals.

In particular, age-standardized incidence rates are two-fold less in women than men, with the latter ranging from 3.3 in Western Africa to 35.4 in Eastern Asia (Ferlay J *et al*., 2015). Geographically speaking, the incidence is lowest in Africa and North America and highest in Eastern Asia, Eastern Europe and South America. Those countries also share the lowest and highest mortality rates respectively (reviewed by Corso S and Giordano S, 2016). Strong variability in GC incidence has also been noted in different ethnic groups living in the same region, where, for example, Native Americans, Hispanics and African-Americans are significantly more affected than Caucasians in the United States (reviewed by Nagini S, 2012).

GC incidence and mortality rates have been steadily plummeting since the very first estimates in 1975 (Parkin DM *et al*., 1984), when GC was reported as the most commonly occurring malignancy in the world (Ferlay J *et al*., 2015). Such decline has been attributed to multiple factors, including changes in food storage techniques and dietary pattern (reviewed by Nagini S, 2012). Despite this encouraging turn of events, GC represents a major health burden in most countries, with an estimated 5-year relative survival rate of approximately 25%-30% (reviewed by Hudler P, 2015), which is predominantly due to the advanced stage of disease at diagnosis (reviewed by Lordick F *et al*., 2014).

2. Histopathological classification of gastric cancer

It has become apparent in recent years that gastric carcinomas can be classified based on the anatomical site in which they reside. According to the Japanese Gastric Cancer Association (2011), tumors residing within the different anatomical sites of the stomach, namely the fundus, corpus or body, and pyloric antrum can be grouped as proximal-third, middle-third and distal-third tumors, respectively (Figure 2).

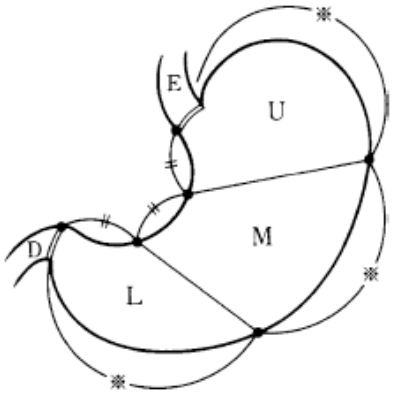

Figure 2. The three portions of the stomach. U upper third, M middle third, L lower third, E esophagus, D duodenum (from Japanese Gastric Cancer Association, 2011).

Alternatively, gastric tumors can simply be classified as proximal, when arising from the cardia region, and distal, when originating in non-cardia regions (reviewed by Lochhead P and El-Omar EM, 2008). However, the stomach cardia harbors the gastroesophageal junction (GEJ) and tumors residing in that anatomically mal-defined region are etiologically and epidemiologically distinct from true non-cardia gastric tumors (reviewed by Hu B *et al*., 2012; reviewed by van

Cutsem E *et al*., 2016). Therefore, to distinguish between true gastric and GEJ cancers, the TNM classification has introduced simplified categories based on the location of the tumor's epicenter and how far it extends into the GEJ. In particular, only carcinomas with an epicenter that is within 5 cm of the GEJ but that do not extend into the GEJ or esophagus; and those with an epicenter that is more than 5 cm distal to the GEJ can be designated as gastric (Edge SB *et al*., 2010; reviewed by van Cutsem E *et al*., 2016).

The majority (around 95%) of these gastric carcinomas are adenocarcinomas, while less than 5% of cases are ascribable to sarcomas and lymphomas, originating from the lymphatic system and connective tissue, respectively (reviewed by Tan P and Yeoh KG, 2015).

Gastric adenocarcinomas are reported to arise from the glandular, mucus secreting, epithelium, and they represent a histopathologically, molecularly and morphologically heterogeneous group of tumors (reviewed by Berlth F *et al*., 2014). This diversity is partly behind the presence of various histological classification systems for gastric cancer, the most common of which are the Laurén and the World Health Organization (WHO) schemes (reviewed by Yakirevich E and Resnick MB, 2013) (Table 1).

Lauren	World Health Organization 2010
Intestinal type	Papillary adenocarcinoma Tubular adenocarcinoma Mucinous adenocarcinoma
Diffuse type	Poorly cohesive carcinoma (including signet ring cell carcinoma and other variants)
Mixed type (equal intestinal and diffuse)	Mixed type, mixture of glandular (tubular/papillary) and poorly cohesive/signet ring
Indeterminate	Undifferentiated carcinoma
	Adenosquamous carcinoma Carcinoma with lymphoid stroma (medullary carcinoma) Hepatoid adenocarcinoma Squamous cell carcinoma

Table 1. Laurén and World Health Organization classification systems of gastric cancer (from Yakirevich E and Resnick MB, 2013).

In 1965 Laurén P classified gastric adenocarcinomas into two primary histological types, "intestinal" (IGC) and "diffuse" (DGC). Carcinomas exhibiting equal features of both histotypes were designated as mixed-type, and the rest as indeterminate.

IGC tumors account for around 54% of gastric adenocarcinomas (Polkowski W *et al*., 1999; reviewed by Hu B *et al*., 2012) and as their name indicates, they are characterized by a moderately- to well-differentiated glandular architecture that is reminiscent of colorectal adenocarcinomas (reviewed by van Cutsem E *et al*., 2016) (Figure 3A). They are also known to display tubular structures (reviewed by Lynch HT *et al*., 2005) and extracellular mucin (Yakirevich E and Resnick MB, 2013).
On the other hand, DGCs, accounting for 32% of cases (Polkowski W *et al*., 1999; reviewed by Hu B *et al*., 2012), lack gland formation and comprise poorly cohesive tumor cells. These cells are known to infiltrate the gastric wall resulting in its widespread thickening and rigidity, also known as *linitis plastica* (reviewed by Lynch HT *et al*., 2005). Another distinctive feature of DGC is the presence of "signet ring cells", which are rich in intracytoplasmic mucin that displaces the nucleus to the periphery (Laurén P, 1965; reviewed by Hu B *et al*., 2012; reviewed by Yakirevich E and Resnick MB, 2013; reviewed by van Cutsem E *et al*., 2016) (Figure 3B).

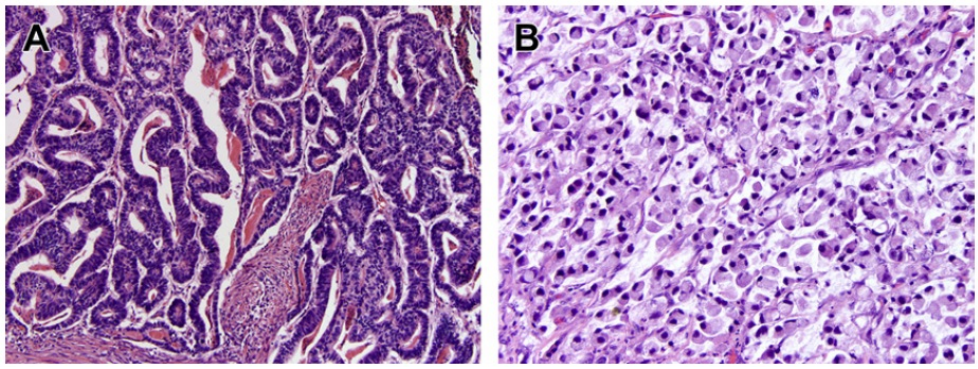

Figure 3. Hematoxylin-eosin stained photomicrographs of the two primary GC histotypes based on Laurén's classification. **A)** Intestinal-type gastric adenocarcinoma exhibiting gland formation. **B)** Diffuse-type gastric adenocarcinoma with signet ring cells (from Yakirevich E and Resnick MB, 2013).

In the revised WHO classification system that was established later in 2010, gastric adenocarcinomas were typed, based on the predominant histological component, into four principle histopathological entities: tubular, papillary, poorly cohesive and mucinous (reviewed by Hu B *et al.*, 2012).
The tubular and papillary tumors, characterized by irregularly distended and elongated epithelial projections respectively, roughly embody the description of Laurén's intestinal type. On the other hand, the poorly cohesive entity, marked by a highly infiltrative mixture of signet and non-signet ring cells, embodies Laurén's diffuse type. With respect to mucinous tumors, they are defined by the presence of

extracellular mucin, occupying more than 50% of their volume. The tumor itself can be glandular or composed of irregular cell clusters, with signet-ring cells occasionally floating in the mucinous pools (reviewed by Yakirevich E and Resnick MB, 2013; reviewed by van Cutsem E *et al*., 2016).
Even though the WHO scheme has the advantage of harmonizing GC classification with that in other parts of the gut, the Laurén scheme is robust and simple, making it the histological classification method of choice for pathologists, epidemiologists and clinicians investigating GC (reviewed by Yakirevich E and Resnick MB, 2013; reviewed by van Cutsem E *et al*., 2016). Indeed, multiple studies have shown that IGC and DGC manifest via distinct carcinogenic pathways (reviewed by Yuasa Y, 2003; Gomceli I *et al*., 2012; Lee YS *et al*., 2014), with variable incidence rates that are sensitive to geographical location, age and sex (reviewed by Lochhead P and El-Omar EM, 2008; reviewed by Tan P and Yeoh KG, 2015).

3. Risk factors in gastric cancer

A complex interplay between environmental and genetic risk factors underlies gastric carcinogenesis (reviewed by McLean MH and El-Omar EM, 2014). Indeed, the differential contribution of these factors results in a GC profile, in which the majority of cases occur sporadically, that is without an evident family history; 10% show familial clustering, where at least two GC cases are diagnosed in the same family; and only 1-3% are strictly hereditary (Figure 4).

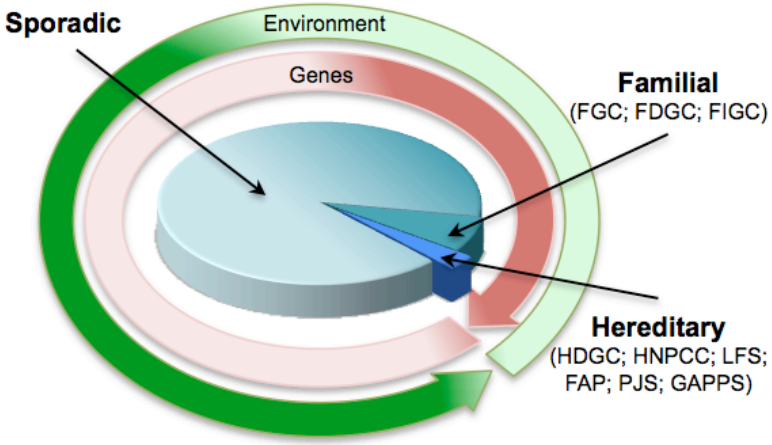

Figure 4. Pie chart showing the proportion of sporadic, familial and hereditary GC cases. Graduated shading of the arrows corresponds to the relevance of environmental (green) and genetic (red) factors in cancer etiology.

Similarly to other cancer types, environmental risk factors associated with diet and lifestyle can gravely contribute to GC development (reviewed by Ang TL and Fock KM, 2014). Unequivocal epidemiological associations exist between augmented disease risk and the high consumption of salts (D'Elia L *et al*., 2012), pickled food (Ren JS *et al*., 2012), nitrites (Song P *et al*., 2015) and low intake of fruits and vegetables (Lunet N *et al*., 2005; Bertuccio P *et al*., 2013), as well as smoking (Ladeiras-Lopes R *et al*., 2008). In addition to that, Epstein - Barr virus (EBV) is the etiological agent in 5-10% of GCs (Chen XZ *et al*., 2015). All these factors underlie sporadic gastric cancer and have a higher contribution in the risk of IGC than DGC, which is more of a "genetically-based" disease (reviewed by Lynch HT *et al*., 2005; reviewed by Ang TL and Fock KM, 2014)
Nonetheless, the most relevant environmental risk factor for sporadic GC is the gram-negative *Helicobacter pylori* (*H. pylori*); in keeping with the complexity of this cancer, low-penetrance genetic variants can additionally contribute to augment or reduce "sporadic" GC risk.

3.1. *H. pylori*

H. pylori-induced inflammation and injury account for 75% of the global burden of GC (reviewed by Herrera V and Parsonnet J, 2009), making it a primary risk factor in this disease. It is more strictly associated with the development of intestinal-type GC through the induction of chronic inflammation and the secretion of virulence factors, which promote the disruption of tight junctions and loss of cell polarity. In this way, *H. pylori* transforms the normal gastric mucosa by altering gene expression and inducing genetic and epigenetic changes in the cells (reviewed by Wadhwa R *et al*., 2013) (Figure 5).

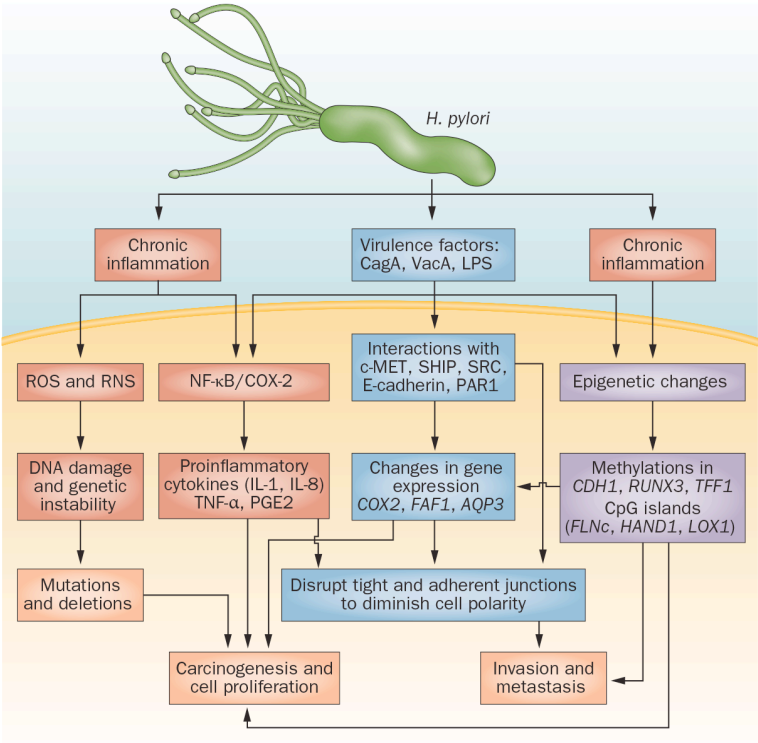

Figure 5. Molecular carcinogenesis of *Helicobacter pylori* in gastric cancer. *H. pylori* and its several virulence factors, such as CagA (cytotoxin-associated gene A), interact with gastric epithelial cells to induce chronic inflammation, mucosal damage and multiple alterations in gene expression and genetic and epigenetic changes, eventually leading to gastric carcinogenesis. VacA: vacuolating cytoxin A; LPS: lipopolysaccharide; ROS: reactive oxygen species; RNS: reactive nitrogen species; CpG islands: areas of cytosine and guanine repeats (from Wadhwa R *et al*., 2013).

This class I carcinogen colonizes the stomach of up to 50% of the world's population, however, only 1-2% develop gastric cancer (Herrera V and Parsonnet J, 2009). Such a disparity is due to a network of interacting components pertaining to bacterial and host genetics, as well as environmental factors. Indeed, not all bacterial strains are carcinogenic and thus far only isolates expressing the cag pathogenicity island (PAI) and those possessing the *vacA* s1/m1 genetic locus, have been shown to increase GC risk (reviewed by Amieva M and Peek RM Jr, 2016). From the infected host's perspective, the presence of certain polymorphisms can considerably modulate GC risk, including variants in genes encoding: the pro-inflammatory interleukins, IL-1 and IL10; chemokine, IL-8; the tumor necrosis factor, TNF-A; toll-like receptor 4 (TLR4) and the mannose binding lectin (MBL2), involved in systemic and mucosal innate immunity (El-Omar EM *et al*.,

2003; Ohyauchi M *et al*., 2005; Baccarelli A *et al*., 2006; reviewed by Amieva MR and El-Omar EM, 2008; reviewed by Amieva M and Peek RM Jr, 2016). Moreover, the risk for gastric cancer jumps up to 87-fold over baseline when high-risk host genotypes are combined with the cancer associated vacA or cag bacterial genotypes (Figueiredo C *et al.,* 2002; reviewed by Amieva M and Peek RM Jr, 2016). The factors contributing to the pathogenesis of *H. pylori*-induced gastric cancer are depicted in figure 6.

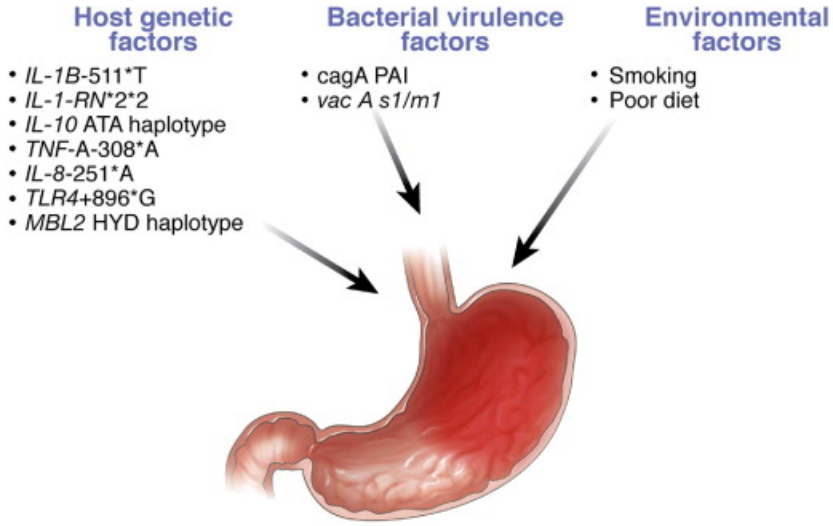

Figure 6. Contribution of host genetics, bacterial, and environmental factors to the pathogenesis of *H. pylori*-induced gastric cancer. These factors combine to create an intragastric milieu characterized by chronic inflammation, bacterial overgrowth, and sustained genotoxic stress. This could ultimately lead to gastric adenocarcinoma (adapted from Amieva MR and El-Omar EM, 2008).

Nonetheless, neither human nor *H. pylori* genetic variations, alone, are sufficient to determine disease susceptibility. Interestingly, recent co-evolutionary studies have reported that a genetic mismatch in terms of host and pathogen ancestries can also account for differences in the degree of gastric injuries in admixed populations (Kodaman N *et al.,* 2014; reviewed by Amieva M and Peek RM Jr, 2016).
In this complex scenario, an evaluation of both host and bacterial genomes can identify people at highest risk for developing GC. Hence, the integration of such measures in disease management would better guide physicians in selecting, for example, prime candidates for antibacterial intervention.

3.2. Single nucleotide polymorphisms (SNPs)

An individual's susceptibility to developing sporadic gastric cancer is not only influenced by the aforementioned environmental factors, but is also associated with polymorphic, low-penetrance genes. Indeed, a myriad of genome wide association studies (GWAS) and candidate-gene-based approaches have unraveled the importance of common risk alleles in disease susceptibility (reviewed by Hudler P, 2015). In a recent paper, Mocellin S and coworkers (2015) performed a comprehensive overview and quantitative summary of the published data on genetic susceptibility to sporadic gastric carcinomas. They investigated 2841 variants and found a highest level of evidence for 11 polymorphisms in 10 genes. Table 2 reports all high-quality biomarkers showing either high or intermediate level of evidence for the association with gastric cancer in general, and in specific subgroups defined by primary tumor site (cardia *vs.* non-cardia), histotype (intestinal *vs.* diffuse), ethnicity (Asian *vs.* Caucasian) and *H. pylori* infection status (positive *vs.* negative).

Gene	Variant ID	Chr	Risk Allele	Datasets	Cases	Controls	Subgroup	sOR	LL	UL	p Value	Venice criteria	Level of evidence	FPRP (10E-3)	FPRP (10E-6)
MUC1	rs2070803	1	A	6	2376	4903	Diffuse	0.585	0.526	0.650	<1.00E-20	AAA	High	0.00	0.00
MTX1	rs2075570	1	G	5	2299	4603	Diffuse	0.587	0.525	0.655	<1.00E-20	AAA	High	0.00	0.00
PRKAA1	rs13361707	5	C	5	7587	7891	Non-cardia	1.406	1.334	1.482	<1.00E-20	AAA	High	0.00	0.00
PLCE1	rs2274223	10	G	5	4869	16 497	Cardia	1.565	1.486	1.648	<1.00E-20	AAA	High	0.00	0.00
PSCA	rs2294008	8	T	7	5690	6975	Non-cardia	1.329	1.254	1.408	<1.00E-20	AAA	High	0.00	0.00
TGFBR2	rs3087465	3	A	4	2323	2347	Asian	0.691	0.620	0.769	1.54E-11	AAA	High	0.00	0.53
PKLR	rs3762272	1	G	3	1362	3101	Diffuse	0.710	0.631	0.799	1.42E-08	AAA	High	0.03	0.97
PSCA	rs2976392	8	G	4	2602	4443	Intestinal	0.808	0.751	0.870	1.42E-08	AAA	High	0.00	0.36
GSTP1	rs1695	11	G	11	2425	4731	Asian	1.191	1.092	1.299	7.88E-05	AAA	High	0.12	0.99
CASP8	rs3834129	2	del	3	616	1085	Mixed	0.732	0.617	0.869	3.48E-04	AAA	High	0.93	1.00
TNF	rs1799724	6	T	10	2136	3817	Mixed	1.173	1.047	1.314	0.006	AAA	High	0.89	1.00
PSCA	rs2294008	8	T	13	7380	26 388	Intestinal	1.284	1.206	1.367	5.33E-15	ABA	Intermediate	0.00	0.00
TRIM46	rs3814316	1	C	3	1362	3101	Diffuse	0.582	0.506	0.670	3.71E-14	AAB	Intermediate	0.00	0.79
THBS3	rs2066981	1	C	3	1362	3101	Diffuse	0.585	0.507	0.674	1.57E-13	AAB	Intermediate	0.00	0.22
ABO	rs8176719	9	A (vs O)	15	8737	1 069 256	Caucasian	1.225	1.161	1.293	1.92E-13	ABA	Intermediate	0.00	0.00
MUC1	rs2070803	1	A	6	3886	4903	Asian	0.660	0.590	0.738	2.55E-13	ABA	Intermediate	0.00	0.01
KRTCAP2	rs4971088	1	T	3	1362	3101	Diffuse	0.608	0.526	0.703	1.60E-11	AAB	Intermediate	0.02	0.96
MTX1	rs2075570	1	G	5	3586	4603	Asian	0.672	0.596	0.759	1.44E-10	ABA	Intermediate	0.01	0.90
PPARG	rs1805192	3	G	5	546	903	Mixed	2.265	1.695	3.027	3.20E-08	BAA	Intermediate	0.78	1.00
PPARG	rs1805192	3	G	3	227	598	Caucasian	2.149	1.536	3.007	8.03E-06	BAA	Intermediate	0.96	1.00
PSCA	rs2294008	8	T	4	1187	3162	Caucasian	1.337	1.177	1.519	8.51E-06	ABA	Intermediate	0.45	0.99
PSCA	rs2976391	8	A	4	1921	2504	Asian	0.725	0.625	0.840	2.00E-05	ABA	Intermediate	0.71	1.00
FAM189B	rs2072647	1	G	3	1362	3101	Diffuse	0.727	0.626	0.844	2.82E-05	ABA	Intermediate	0.44	0.99
PTGS2	rs20417	1	C	6	1464	3000	Asian	1.773	1.356	2.318	2.85E-05	BBB	Intermediate	0.92	1.00
THBS3	rs2066981	1	C	3	2649	3101	Asian	0.721	0.615	0.845	5.63E-05	ABA	Intermediate	0.60	0.99
IL1B	rs1143634	2	T	4	434	450	HP positive	1.721	1.321	2.243	5.84E-05	BAB	Intermediate	0.93	1.00
IL17F	rs763780	6	G	4	1760	2362	Asian	1.288	1.138	1.458	6.28E-05	AAB	Intermediate	0.63	0.99
YY1AP1	rs3738590	1	T	3	1362	3101	Diffuse	0.750	0.651	0.864	6.78E-05	ABA	Intermediate	0.77	1.00
KRTCAP2	rs4971088	1	T	3	2649	3101	Asian	0.725	0.618	0.852	9.04E-05	ABA	Intermediate	0.87	1.00
TNF	rs1800629	6	A	38	7896	12 982	Mixed	1.166	1.074	1.267	2.67E-04	ABB	Intermediate	0.26	0.99
TLR4	rs4986790	9	G	3	774	1129	Non-cardia	2.022	1.351	3.026	6.17E-04	BBA	Intermediate	0.99	1.00
MDM2	rs2279744	12	G	4	778	2272	Cardia	1.384	1.133	1.690	0.001	ABB	Intermediate	0.97	1.00
TLR4	rs4986790	9	G	8	1859	3276	Caucasian	1.475	1.159	1.878	0.002	BBB	Intermediate	0.98	1.00
TNF	rs1800629	6	A	19	3516	6604	Caucasian	1.187	1.065	1.322	0.002	ABA	Intermediate	0.76	1.00
IL1B	rs16944	2	T	9	644	4867	Cardia	1.199	1.066	1.349	0.003	AAB	Intermediate	0.83	1.00
IL8	rs4073	4	A	8	1058	2620	Intestinal	1.240	1.077	1.428	0.003	ABA	Intermediate	0.95	1.00
ABO	rs8176719	9	A (vs AB)	13	4696	620 332	Caucasian	1.234	1.068	1.427	0.004	ABB	Intermediate	0.96	1.00
TNF	rs1799724	6	T	7	1503	2517	Asian	1.189	1.043	1.355	0.010	BAA	Intermediate	0.94	1.00
TLR4	rs4986790	9	G	9	1919	3438	Mixed	1.433	1.089	1.886	0.010	BBB	Intermediate	0.99	1.00
IL10	rs1800871	1	T	3	78	399	Diffuse	0.567	0.367	0.875	0.010	BAB	Intermediate	0.99	1.00
EPHX1	rs1051740	1	C	3	427	1401	Caucasian	1.236	1.049	1.456	0.011	BAB	Intermediate	0.98	1.00
IGFBP3	rs2854744	7	C	3	1100	1338	Mixed	1.181	1.026	1.360	0.020	BAB	Intermediate	0.98	1.00
TP53	rs17878362	17	ins	3	342	635	Mixed	1.373	1.051	1.793	0.020	BAB	Intermediate	0.99	1.00
PRKAA1	rs13361707	5	C	3	1113	5261	Cardia	1.177	1.025	1.350	0.020	AAB	Intermediate	0.97	1.00
TIMP2	rs2277698	17	T	3	692	685	Mixed	0.807	0.669	0.975	0.026	BAB	Intermediate	0.99	1.00
IL8	rs2227306	4	T	4	348	1040	Asian	1.225	1.017	1.476	0.032	BAB	Intermediate	0.99	1.00
TNF	rs1800629	6	A	7	1239	4520	Cardia	1.180	1.010	1.379	0.037	BAB	Intermediate	0.98	1.00
TNF	rs1800629	6	A	8	1338	3836	Diffuse	1.262	1.010	1.578	0.041	BBB	Intermediate	0.99	1.00
KIAA0907	rs625658	1	C	3	1362	3101	Diffuse	0.842	0.713	0.993	0.042	ABB	Intermediate	0.99	1.00

Venice criteria: A (high), B (moderate), C (weak) credibility for three parameters (amount of evidence, heterogeneity and bias; see text for more details); level of evidence: overall level of summary evidence according to the Venice criteria.
Chr, chromosome; FPRP, false positive report probability (at two levels of prior probability, see text for more details); HP, *Helicobacter pylori*; LL, 95% lower level; sOR, summary OR; UL, 95% upper level.

Table 2. Genetic variants significantly associated with GC risk with a high or intermediate level of summary evidence (from Mocellin S *et al*., 2015).

In their work, the authors further distributed the data based on main gene function, which revealed that a great percentage of these variants are attributed to genes involved in adhesion/invasiveness (14%) or immunity/inflammation (36%) (Figure

7). Interestingly, this finding is in line with the eminent role of *H. pylori* in GC and can add another explanation for the discrepancy between the high prevalence of *H. pylori* infection in the world and the comparably low gastric cancer incidence.

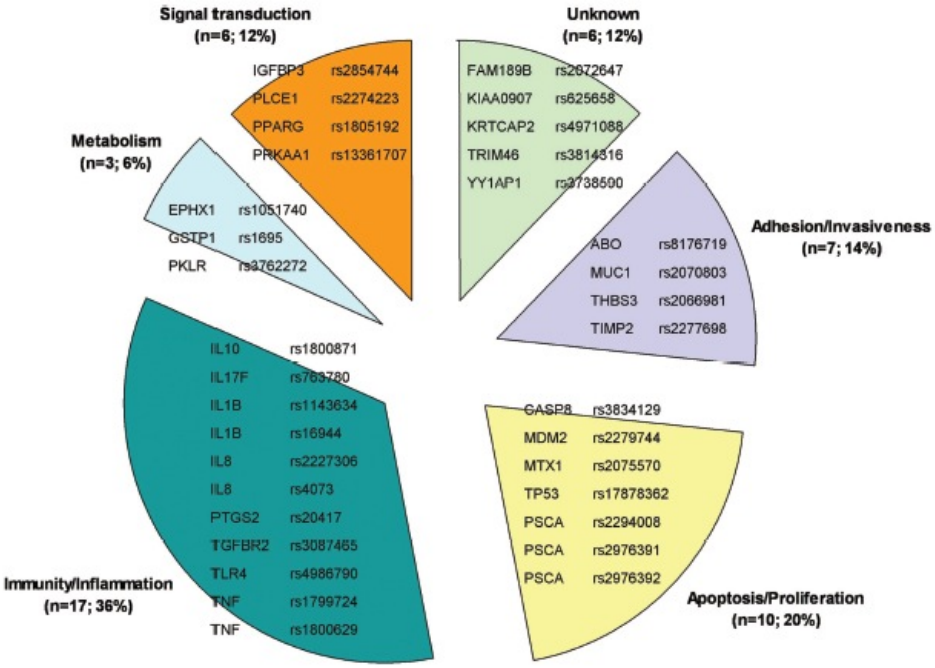

Figure 7. SNPs significantly associated with GC risk distributed based on main gene function (from Mocellin S *et al*., 2015).

Regarding SNPs involved in other gene functions, a recent systemic analysis of 39 SNPs in 8 core genes of the nucleotide excision repair (NER) pathway in a Chinese Han population, determined the presence of a significant association between two polymorphisms, rs2607775 and rs830083, and an increased susceptibility to GC. In addition, the coexistence of risk variants in these two SNPs, which reside in the *XPC* and *DDB2* DNA damage recognition genes of the NER pathway respectively, seemed to confer an even higher GC risk (OR=3.05) (Liu J *et al*., 2016). Furthermore, candidate gene approaches have hinted at the association of polymorphisms in the *XPG* DNA repair gene with GC risk, again in a Chinese population (Feng YB *et al*., 2016; Zhou RM *et al*., 2016). Such associations, however, remain to be validated in larger cohorts.

In addition to their influence on the manifestation of gastric cancer, multiple genetic variants have been found to affect disease prognosis and clinical outcome (Liu J *et al*., 2016; Liu R *et al*., 2016; Jia ZF *et al*., 2016; Nie X *et al*., 2016).

Nonetheless, it is important to keep in mind that the effect of low-penetrance alleles varies depending on the population studied and on the presence of other genetic variations and/or environmental factors (reviewed by Hudler P, 2015). As such heterogeneous studies have given conflicting results on the involvement of the same variants in GC (Cao Z *et al*., 2014; Xu J *et al*., 2014). Moreover, while the attractiveness of SNPs lies in their accessibility for simple and robust screening by using cost-effective PCR-based methods, there are multiple obstacles hindering SNP profiling from becoming clinically relevant (reviewed by Hudler P, 2015). The most critical of these barriers are fine-mapping the causal SNP, which is often obscured by the presence of linkage disequilibrium, and determining its potential functional significance. Therefore, instead of continuing to put forth novel probable associations, more effort needs to be made in validating existing ones and understanding their precise contribution to gastric carcinogenesis. Only then can the clinical assessment of such variants be efficiently integrated in patient stratification and disease management.

4. Molecular pathogenesis of gastric cancer

Gastric cancer is a complex disease that often develops following the accumulation of various “driver” genetic and epigenetic anomalies, which result in the disruption of key cellular pathways pertaining to signal transduction, cell adhesion and differentiation, as well as DNA repair and apoptosis (reviewed by Tan P and Yeoh KG, 2015) (Figure 8) .

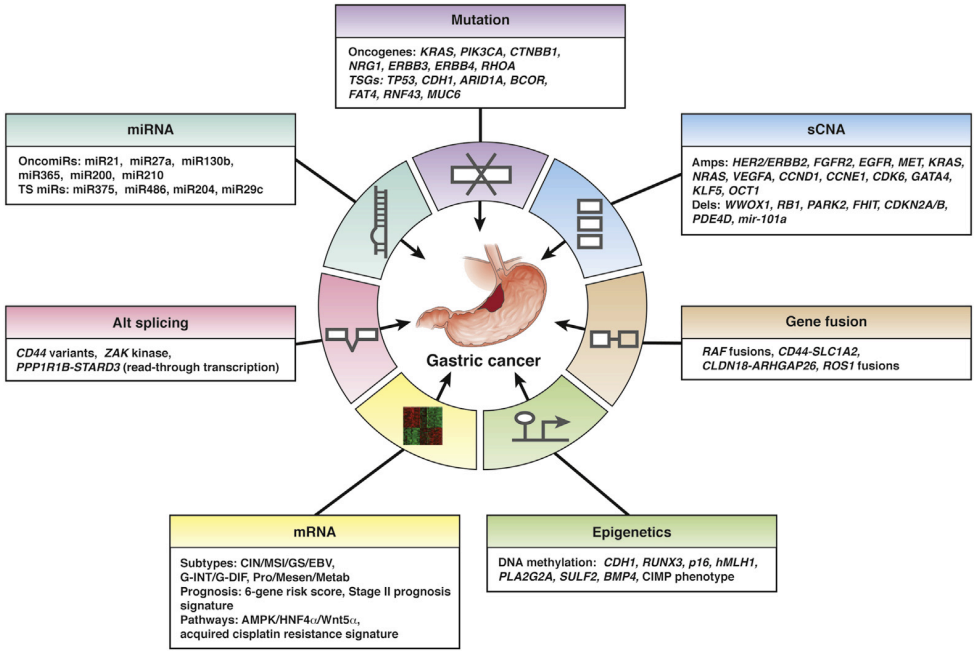

Figure 8. Genomic features of gastric cancer. Anomalies comprise gene mutations, somatic copy number alterations (sCNAs) and structural variations, altered methylation profiles, and mRNAs and noncoding RNAs (ncRNAs) transcriptional variations (from Tan P and Yeoh KG, 2015).

With respect to intestinal-type GC, it is known to develop as a consequence of chronic inflammation, which as was just explained, is most often ensued by *H. pylori* infection. In susceptible individuals, resulting gastritis would progress to intestinal metaplasia that can subsequently end in a full-blown adenocarcinoma (reviewed by Hayakawa Y *et al*., 2016). Figure 9 depicts all what is known thus far regarding the causes and pathogenesis of IGC, which include host genetics, environmental factors and acquired molecular events.

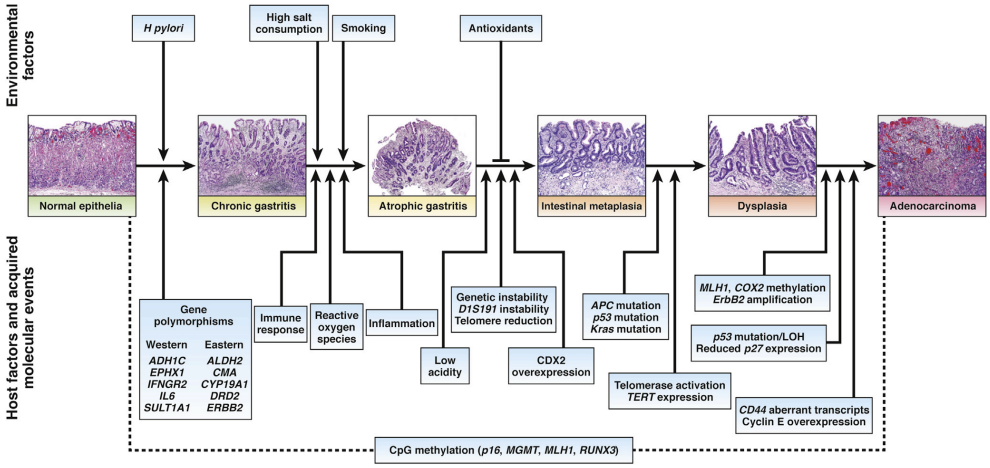

Figure 9. Cause and pathogenesis of intestinal-type GC. A summary of current knowledge of the cause and pathogenesis of intestinal type GC is shown, including host and environmental factors as well as acquired molecular events (Tan P and Yeoh KG, 2015).

On the other hand, regarding diffuse-type GC, it has been shown to develop *de novo*, mainly due to deregulation of the *CDH1* gene in epithelial cells (reviewed by Wadhwa R et *al.*, 2013). Unlike IGC, no identifiable precursor lesions have been identified for DGC, and the molecular events giving rise to these tumors have been found to be, for the most part, distinct from those involved in the carcinogenesis of the intestinal histotype. This distinction has been recently highlighted by the discovery that the two histotypes represent diverging molecular subtypes of GC (Cancer Genome Atlas Research Network, 2014).

5. Molecular classification of gastric cancer

In an effort to better delineate gastric tumors at the molecular level, The Cancer Genome Atlas (TCGA) research network (2014) applied an unbiased informatics approach using 6 molecular platforms (whole exome sequencing, DNA methylation analysis, evaluation of somatic copy number alterations, mRNA and miRNA sequencing, and phosphoproteomics) on 295 primary gastric adenocarcinomas (reviewed by Corso S and Giordano S, 2016). This cumbersome work resulted in the identification of not just two, but four, molecularly discrete GC subtypes (Figure 10).

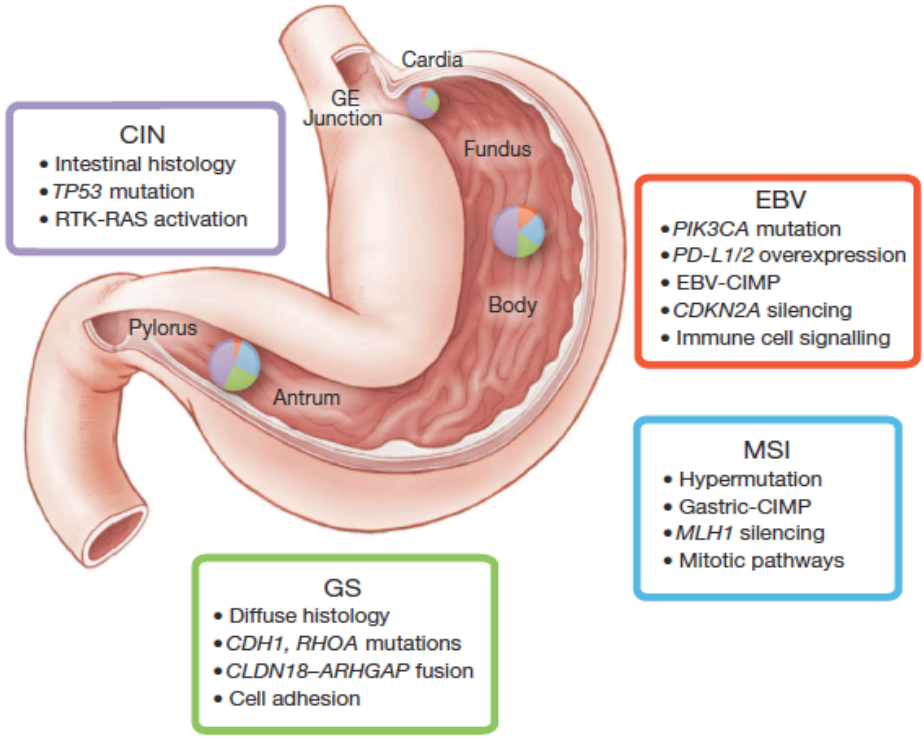

Figure 10. Key features associated with the four molecular GC subtypes. The distribution of the subtypes in tumors obtained from distinct regions of the stomach is represented by inset charts (adapted from Cancer Genome Atlas Research Network, 2014).

The first subtype harbored 9% of the tumors studies and described those positive for EBV infection. Tumors were mainly characterized by profound DNA promoter hypermethylation and the presence of activating *PIK3CA* mutations (in 80% of cases). They also had an increased tendency of manifesting in the fundus or body of the stomach (reviewed by Corso S and Giordano S, 2016).

The second subtype comprised around 22% of tumors and included those with microsatellite instability (MSI) that presented with a distinct pattern of DNA hypermethylation compared to the first subgroup; for example, a key component of the DNA mismatch repair system, *MLH1*, was hypermethylated only in MSI tumors, but not in EBV+ ones (reviewed by Corso S and Giordano S, 2016). The most notable feature of MSI tumors was their elevated somatic mutation rates, with recurring mutations of *PIK3CA* and *ERBB3* in 42% and 26% of cases respectively (reviewed by Lordick F and Janjigian YY, 2016).

IGC tumors were found to cluster in the third subtype, which incidentally included the majority of studied cases (50%) and was characterized by chromosomal

instability (CIN). These tumors presented marked aneuploidy and amplification of receptor tyrosine kinase (RTK) genes, of which *HER2* (24%) was most evident. CIN tumors also harbored the highest frequency of p53 mutations among the non-hypermutated cancers.
On the other hand, 73% of the DGC tumors in this study were enriched for in the fourth and final subtype that defined genomically stable (GS) tumors (reviewed by Van Cutsem E *et al*., 2016). These tumors comprised 20% of cases and lacked all aforementioned features of aneuploidy, excessive hypermethylation and hypermutation. They were instead characterized by frequent alterations in genes responsible for cell adhesion, including *CDH1* (26%), *RHOA* (15%) and *CLDN18*/*ARHGAP* (15%) (Wang K *et al*., 2014; reviewed by Corso S and Giordano S, 2016).
While the TCGA classifier might be influential in the identification of putative and beneficial target therapies for gastric cancer, it did not correlate well with disease prognosis. Meanwhile, a multi-platform genome analysis performed by the ACRG (Asian Cancer Research Group) on 300 primary gastric adenocarcinomas divided GC in 4 other subtypes that seemed to correlate with prognosis (Cristescu R *et al*., 2015; reviewed by Lin Y *et al*., 2015) (Figure 11).

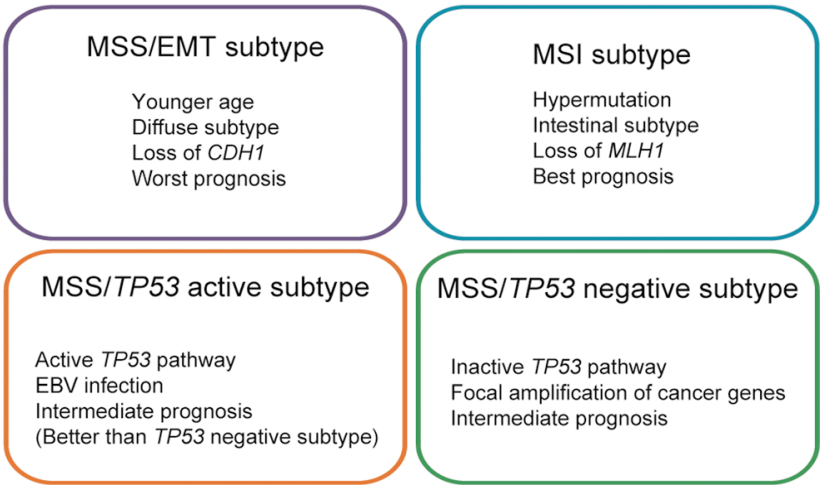

Figure 11. The ACRG classification of gastric cancer. MSI: microsatellite instability; MSS: microsatellite stable; EMT Epithelial-to-Mesenchymal Transition (from Lin Y *et al*., 2015).

Both the TCGA and ACRG classifiers had similar definitions for the MSI subtype, however in the case of ACRG the CIN and GS features were distributed among the different subtypes and did not form their own groups. Moreover, the TCGA GS subtype was not equivalent to the ACRG EMT subtype when considering the

frequencies of *CDH1* and *RHOA* mutations; nor did the TCGA EBV subtype completely overlap with the ACRG *TP53*-active group. This barely partial matching between the two classifiers can be explained by variability in the histological and anatomical locations of the analyzed tumors, the molecular approaches utilized, and the differential ethnic origins of the cohorts integrated in the two studies (reviewed by Corso S and Giordano S, 2016). Nonetheless, it is important to stress that the ACRG classifier was significantly associated with patient survival in three additional cohorts, including the one of TCGA, regardless of the aforementioned heterogeneity (reviewed by Lin Y *et al*., 2015).
The application of such stratification frameworks provides a stepping stone for advancing our knowledge on the molecular basis of gastric cancer and are crucial in building the proper rational for the treatment of GC patients, in the hope of improving their outcome.

6. Familial and hereditary gastric cancer

A family history of gastric cancer has been identified as an independent risk factor in developing this malignancy, especially if it is a first-degree relative who is affected (Yaghoobi M *et al*., 2010; reviewed by Monahan KJ and Hopkins L, 2016). The presence of familial clustering might be influenced by diet or *H. pylori* infection, however, studies in which those factors were adjusted for, found that neither had an impact on the relative risk associated with a positive family history. Such findings promote genetic susceptibility as being the instigating factor behind familial gastric cancer (Yaghoobi M *et al*., 2010; reviewed by Monahan KJ and Hopkins L, 2016).
As such, the recognition of familial gastric cancer cases is clinically useful as it is usually the first step in identifying a disease with a genetic component (reviewed by Oliveira C *et al*., 2015). Nonetheless, most familial cases lack detrimental germline alterations and are likely due to the co-involvement of multiple alleles with low/moderate penetrance, as well as environmental factors (reviewed by Wadhwa R *et al*., 2013; reviewed by McLean MH and El-Omar EM, 2014; Mocellin S *et al*., 2015). Such familial cases are recognized by the presence of at least two GC patients in a family and are classified based on whether the tumor histotype could be determined or not. In the case where the hisototype cannot be determined, diagnosed families are designated as Familial Gastric Cancer (FGC); meanwhile, those with an IGC or DGC index case are designated as Familial Intestinal Gastric Cancer (FIGC) and Familial Diffuse Gastric Cancer (FDGC), respectively (reviewed by Oliveira C *et al*., 2006; reviewed by Barber M *et al*., 2006). Indeed, germline transmission of highly penetrant alleles is responsible for a small percentage of GC cases that can be labeled as truly hereditary. In this scenario, predisposition to GC primarily occurs in the context of the Hereditary

Diffuse Gastric Cancer (HDGC) syndrome, which has originally been associated with germline mutations in the *CDH1* gene. Other hereditary cancer syndromes including Hereditary NonPolyposis Colorectal Cancer (HNPCC or Lynch Syndrome), Li-Fraumeni Syndrome (LFS), Peutz-Jeghers Syndrome (PJS), and Juvenile Polyposis Syndrome (JPS), can additionally increase the lifetime risk of developing GC, albeit to a much lesser extent than HDGC (reviewed by Tan RY and Ngeow J, 2015) (Table 3).

Condition	Genetic pathology	Lifetime risk of gastric cancer	Histological subtype
Hereditary diffuse gastric cancer	*CDH1* germline and other gene mutations	80%	Diffuse
Lynch syndrome	Mutations in mismatch repair genes	4.8% in *MLH1* carrier 9% in *MLH2* carrier	Mainly intestinal-type
Familial adenomatous polyposis	*APC* germline mutations	Population risk	No data
Li-Fraumeni syndrome	*TP53* mutations	4.9%	No predominant subtype
Peutz-Jegher's syndrome	*STK11* mutations	29%	No data
Juvenile polyposis syndrome	*SMAD4* or *BMPR1A* mutations	21%	No data

Table 3. Hereditary cancer syndromes associated with gastric cancer development. The corresponding susceptibility genes, GC risk and main histological subtype are also indicated (adapted from Tan RY and Ngeow J, 2015).

Aside from the aforementioned syndromes, Gastric Adenocarcinoma and Proximal Polyposis of the Stomach (GAPPS), has emerged as a novel autosomal-dominant syndrome predisposing to GC of the intestinal type (Worthley DL *et al*., 2012). GAPPS is characterized by the transmission of fundic gland polyps and has been identified in 9 families in the world thus far (Li J et *al*., 2016; Repak R *et al*., 2016). It is only diagnosed after dismissing all other types of gastric polyposis and hereditary gastrointestinal syndromes, as well as the use of proton-pump inhibitors that are known to induce a phenotype similar to GAPPS. Importantly, examined gastric polyps should be restricted to the body and fundus of the stomach without any involvement of the duodenal and colorectal regions (reviewed by Oliveira C *et al*., 2015; Li J *et al*., 2016; Repak R *et al*., 2016). This is especially relevant for

ruling out Familial Adenomatous Polyposis (FAP), in which more than 80% of affected individuals harbor fundic gland polyposis (Repak R *et al*., 2016).
The genetic basis of GAPPS has remained elusive up until recently, when three different point mutations (c.-191T>C, c.-192A>G and c.-195A>C) within promoter 1B of *APC* tumor suppressor gene were identified as being causative of the disease in 6 families (1 Australian and 5 North American of European descent) (Li J *et al*., 2016). These mutations mapped to the Ying Yang 1 (YY1) binding motif, and seemed to reduce allele-specific *APC* transcription by compromising the binding affinity of the YY1 transcription factor to promoter 1B. In specific, the c.-195A>C mutation cosegregated with GAPPS in the 27 affected members of the Australian family; c.-192A>G was found in both affected members in one North American family; and c.-191T>C was shared among the affected members of the 4 remaining North American families (Li J et *al*., 2016). This last mutation was subsequently reported in 3 patients diagnosed with GAPPS in a Czech family (Repak R *et al*., 2016). The implication of promoter 1B of the *APC* gene in the manifestation of fundic gland polyposis suggests that carriers of pathogenic mutations in this region are probably at a higher risk for developing gastric cancer, regardless of their phenotype, and must be managed accordingly.
Indeed, while only a small fraction of GC cases are hereditary (<3%), unraveling the underlying genetic lesion is highly advantageous, as it aids in recognizing at risk-individuals, by providing an opportunity for proper genetic counseling and genetic testing. This is especially relevant since the majority of GCs, DGCs in particular, show palpable symptoms only when the tumor has become incurable. In such setting, the effectivity of therapeutic strategies can be positively influenced by the identification and pre-symptomatic diagnosis of individuals at risk (Brooks-Wilson AR *et al*., 2004).

7. Hereditary Diffuse Gastric Cancer (HDGC) syndrome

Hereditary Diffuse Gastric Cancer (HDGC) syndrome is the highest lifetime risk factor for GC (reviewed by Monahan KJ and Hopkins L, 2016). It is a rare autosomal dominant gastric cancer predisposition that is associated with the development of diffuse-type gastric cancer, coupled with an increased risk of lobular breast cancer (LBC) in women. It was first documented in 1964 in 3 Maori families from New Zealand who segregated early-onset DGC in multiple generations (Jones EG, 1964). More than 30 years later, linkage analysis and mutation screening in these families identified germline mutations of the *CDH1*, E-cadherin encoding gene, as the genetic cause of HDGC (Guilford P *et al*., 1998). This association between *CDH1* mutations and DGC predisposition was confirmed soon after in families of the same and different ethnic origins (Guilford P *et al*., 1998; Gayther SA *et al.,* 1998; Guilford PJ *et al*., 1999; Keller G *et al*., 1999;

Richards FM *et al*., 1999; Yoon KA *et al*., 1999), and was followed by a global attempt to better define the HDGC syndrome.

7.1 Clinical presentation

The symptoms associated with HDGC normally appear only after the disease has reached an advanced stage, and they include weight loss, abdominal pain, nausea, anorexia, dysphagia, melaena and early satiety (reviewed by Tan RY and Ngeow J, 2015). In most countries, the average age of onset is estimated to be 38 years, and it is highly variable both within and among HDGC families, ranging from 14 to 82 years (reviewed by Tan RY and Ngeow J, 2015; reviewed by Gurzu S *et al*., 2015). For instance, in New Zealand the majority of cases occur at around 15-17 years of age, which makes the median age of diagnosis much lower in this country than that in the rest of the world (reviewed by Gurzu S *et al*., 2015). Such early-age diagnosis of DGC has also been reported in patients without family history and has been linked, in 4-8% of cases, to *de novo* mutations in *CDH1* gene (Shah MA *et al*., 2012; Sugimoto S *et al*., 2014). These cases have been designated as Early Onset Diffuse Gastric Cancer (EODGC).

In addition to DGC, females carrying a *CDH1* pathogenic germline mutation are predisposed to LBC. Interestingly, the lobular breast carcinomas present with a similar histological aspect to DGC, where, for example, signet ring cancer cells are dispersed in the stromal tissue (Schrader KA *et al*., 2008).

While the penetrance of HDGC is high, it is incomplete and, based on the latest and largest study to date (Hansford S *et al*., 2015), the cumulative risk of DGC for mutation carriers is estimated to be 70% for men (95% CI: 59-80%) and 56% for women (95% CI: 44-69%) by the age of 80. In addition, with respect to female carriers, the cumulative risk of LBC is 42% (95% CI: 23-68%) by 80 years (Hansford S *et al*., 2015). Apart from LBC, colorectal, prostate and ovarian carcinomas have also been reported in HDGC families (Caldas C *et al*., 1999; Brooks-Wilson AR *et al*., 2004). However, there is no concrete evidence, thus far, that *CDH1* mutation carriers have a significantly increased risk of other cancer types (Hansford S *et al*., 2015). Alternatively, cleft-lip (with or without cleft-palate) malformations have been reported in several HDGC families, some of which harbor *CDH1* truncating mutations (Frebourg T *et al*., 2006; reviewed by Oliveira C *et al*., 2013).

7.2. Selection criteria for HDGC families

In order to identify individuals at risk of HDGC and those who require genetic testing for the condition, the following criteria were established in 1999 by the International Gastric Cancer Linkage Consortium (IGCLC): (1) two or more documented cases of DGC in first/degree relatives, with at least one diagnosed

before the age of 50; or (2) three or more cases of documented DGC in first/degree relatives, independent of age of onset (reviewed by Fitzgerald RC *et al*., 2010; reviewed by Pinheiro H *et al*., 2014).
Later in 2004, an expansion of the stringent HDGC criteria was proposed by Brooks-Wilson AR and colleagues to increase patient eligibility for *CDH1* screening. This suggestion was implemented by IGCLC in 2010, when the selection criteria were broadened to include individuals with DGC diagnosed before the age of 40 years without a family history (*i.e.* EODGC), and cases diagnosed with lobular breast cancer, as well as those presenting *in situ* signet ring cells or signet-ring colon carcinomas (Brooks-Wilson AR *et al*., 2004; Suriano G *et al*., 2005; Fitzgerald RC *et al*., 2010). Van der Post RS and coworkers (2015) have recently reviewed the selection criteria and the updated guidelines are depicted in Table 4.

*Established criteria**
• 2 GC cases regardless of age, at least one confirmed DGC
• One case of DGC <40
• Personal or family history of DGC and LBC, one diagnosed <50
*Families in whom testing could be considered**
• Bilateral LBC or family history of 2 or more cases of LBC <50
• A personal or family history of cleft lip/palate in a patient with DGC
• *In situ* signet ring cells and/or pagetoid spread of signet ring cells

Table 4. Updated criteria for offering genetic testing to individuals at risk of HDGC (adapted from van der Post RS *et al*., 2015).

Of note, these new guidelines have already shown an improvement over the 2010 IGCLC criteria with respect to discriminating power and sensitivity in identifying *CDH1* mutation carriers (Benusiglio PR *et al*., 2015).

8. *CDH1* gene and E-cadherin protein

The human *CDH1* gene is annotated to the chromosome 16q22.1 and extends for over 90 kb of genomic DNA (Figure 12). *CDH1* comprises 16 canonical exons with a canonical transcription start site (TSS) currently mapping to the coordinate 68,737,225 bp (assembly GRCh38). The *CDH1* canonical promoter has been reported to span the region from 125 bp upstream of the TSS to 27 bp downstream of it. This region encompasses several regulatory elements, including E-boxes, GC boxes and a CCAAT-box, all of which are highly conserved across mammals (reviewed by Paredes J *et al*., 2012).

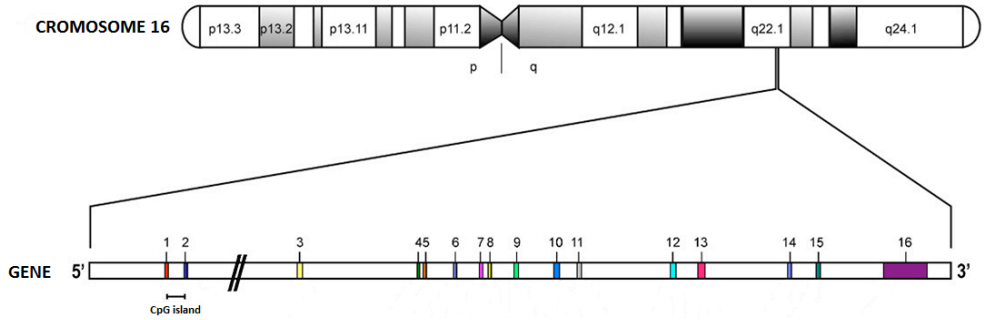

Figure 12. Chromosomal localization and structure of *CDH1* gene. Coding exons are represented as colored boxes (adapted from van Roy F and Berx G, 2008).

CDH1 is transcribed to an approximately 4.8 Kb mRNA (*NCBI RefSeq* NM_004360.4) with a translation start site (ATG) that is 192 bp downstream of the TSS. The coding sequence (CDS) measures 2,649 bp and is translated into a 120-kDa protein, named E-cadherin, which is composed of 882 amino acids (aa) (*CCDS* 10869.1; *RefSeq* NP_004351.1).

E-cadherin, the so-called Epithelial-cadherin, is predominantly expressed at the level of the adherens junctions (zonula adherens) in the basolateral membrane of epithelial cells. There it plays a fundamental role in the organization of epithelial architecture through the maintenance of cell polarity and differentiation. This function is pivotal both during development and in the fully-grown organism (reviewed by van Roy F and Berx G, 2008; reviewed by Lecuit T and Yap AS, 2015). E-cadherin belongs to the cadherin superfamily that is made up of more than 100 proteins, including cadherins, protocadherins, desmocollins and desmogleins. The members of this superfamily are glycoproteins involved in calcium-dependent cell-cell adhesion mediated by shared extracellular Ca^{2+}-binding domains. Despite this similarity, their functional specifications are attributed to a great structural diversity in terms of their ectodomains, composition of the cytoplasmic domains, presence of additional motif as well as a wide range of sizes (reviewed by van Roy F, 2014). E-cadherin, in particular, arises from a primary polypeptide precursor harboring a 27 aa signal sequence encoded by exons 1 and 2 that is required for transport towards the endoplasmic reticulum. The precursor also includes an approximately 130 aa sequence encoded by exons 2-4, which is post-translationally removed to form the mature protein of 728 residues (Figure 13).

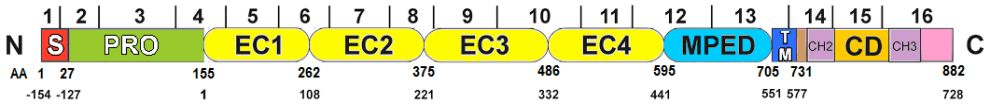

Figure 13. Schematic representation of E-cadherin protein (adapted from van Roy F and Berx G, 2008; courtesy of M. Marabelli). S: Signal peptide; PRO: Propeptide; EC1-4: Extracellular Cadherin repeat; MPED: Membrane Proximal Extracellular Domain.; TM: Transmembrane domain; CD: Cytoplasmic Domain. The exons coding for each portion of the protein are also indicated: signal peptide (exons 1-2); propeptide (exons 2-4); extracellular domain (exons 4-13); transmembrane domain (exons 13-14) and cytoplasmic domain (exons 14-16).

The resulting protein is structurally divided into three major parts: a highly conserved intracellular (cytoplasmic) domain, a transmembrane domain, and a large N-terminal ectodomain. This last is composed of five tandemly repeated domains, including four Extracellular Cadherin repeats (EC1 to EC4) and a Membrane Proximal Extracellular Domain (MPED), which is characterized by the presence of four conserved cysteine residues. The reduction of disulphide bridges formed among these residues is known to interfere with the stability of cell-cell contacts (reviewed by van Roy F and Berx G, 2008). In addition to that, the correct folding and dimerization of the protein, which is central to the adhesion mechanism itself, is greatly influenced by two highly conserved calcium-binding motifs present in all 5 domains (Nagar B *et al*., 1996; Shapiro L and Weis WI, 2009). Indeed, homotypic cell-cell adhesion is due to homophilic interactions between the ectodomains of cadherin molecules on adjacent cells. While the mechanism remains unclear, *cis*-homophilic dimers are known to form between E-cadherin molecules of the same cell and *trans*-homophilic dimers are known to occur between molecules on neighboring cells, each giving rise to zipper-like structures (reviewed by Mohamet L *et al*., 2011; Fichtner D *et al*., 2014) (Figure 14).

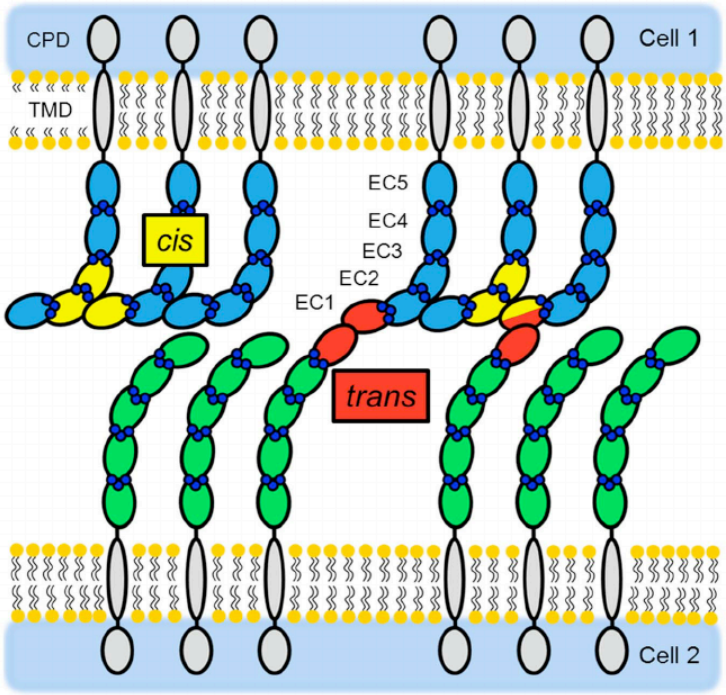

Figure 14. Schematic depiction of *cis*- and *trans*-interactions between Extracellular Cadherin Repeats (ECs) on the cell surface. Transinteractions (highlighted in red) are mediated by the EC1 and occur between cadherin molecules of neighboring cells. Lateral *cis*-interactions (indicated in yellow) occur between the EC1 of one cadherin molecule and the EC2, the EC2/EC3 linker region and part of the EC3 of an adjacent cadherin molecule exposed on the same cell. A combination of *cis*- and *trans*-interactions is depicted on the right hand side of the scheme. Here, the EC1 is simultaneously engaged in a *cis*-interaction with an adjacent cadherin molecule from the same cell, as well as in a *trans*-interaction with a cadherin molecule from an opposite cell (the EC2 involved in both *cis*- and *trans*- interactions is colored in yellow and red) (from Fichtner D *et al*., 2014).

While the ectodomain provides the necessary tools for cell-cell adhesion, the cytoplasmic portion of the protein mediates the interaction with the cytoskeleton. Two subdomains, a conserved juxtamembrane domain (JMD) and a β-catenin binding domain (CBD), can be distinguished within the intracellular region, and they are represented by short motifs known as CH2 and CH3, respectively (reviewed by van Roy F and Berx G, 2008). CBD is bound in a mutually exclusive manner by β-catenin and γ-catenin (also called plakoglobin), while the JMD is associated with the p120-catenin. Another catenin, namely α-catenin, joins the group to form the cadherin-catenin complex (CCC) (Aberle H *et al*., 1996) (Figure 15).

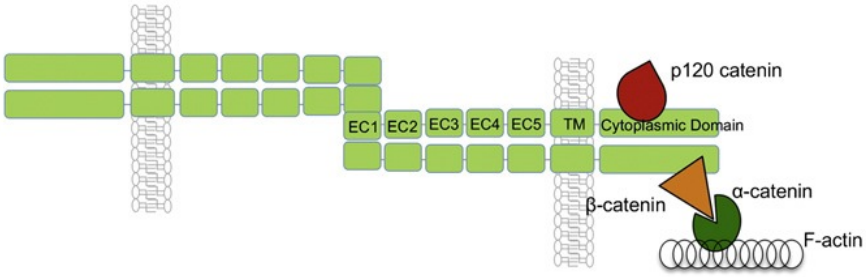

Figure 15. Schematic representation of the E-cadherin-catenin complex (from Paredes J *et al*., 2012).

Since the association of γ-catenin with CBD has been found to be less stable than that of β-catenin, the main adhesive complex consists of E-cadherin/β-catenin/α-catenin (reviewed by Aktary Z and Pasdar M, 2012). In this complex the primary role of the aforementioned p120-catenin is to control membrane trafficking and turnover rate by inhibiting clathrin-mediated endocytosis (Xiao K *et al*., 2007), as well as stabilizing the intercellular adhesion (Thoreson MA *et al*., 2000). Meanwhile, α-catenin plays a key role in linking the CCC with the actin cytoskeleton by directly binding the actin filaments (F-actin) and mediating less direct interactions via proteins such as Epithelial Protein Lost In Neoplasm (EPLIN) and vinculin (Watabe-Uchida M *et al*., 1998; Yonemura S *et al*., 2010; Huveneers S *et al*., 2012; Desai R *et al*., 2013; reviewed by Lecuit T and Yap AS, 2015). In addition to that, E-cadherin can associate with proteins such as myosin VI, independently of α-catenin (Mangold S *et al*., 2012; reviewed by Lecuit T and Yap AS, 2015). It is thought that the wide range of proteins linking cadherins to the cortical F-actin make distinct contributions to the junctions' mechanical properties (reviewed by Lecuit T and Yap AS, 2015). All these interactions are indispensible for preventing individual epithelial cell motility, thereby maintaining the homeostasis of tissue architecture.

The role of E-cadherin, however, goes beyond cell-cell adhesion and regulation of epithelial structure. Indeed, it is additionally an active participant in multiple signal transduction pathways (reviewed by Du W *et al*., 2014) (Figure 16). These pathways, including Wnt/β-catenin (Orsulic S *et al*., 1999), Rho GTPase (Noren NK *et al*., 2000), Receptor Tyrosine Kinase (RTK) (Qian X *et al*., 2004) and NF-κB (Van den Bossche J *et al*., 2012), are known to participate in cell-proliferation, -survival, -migration and inflammation. Moreover, cadherins might act as a force-sensor promoting mechanotransduction by converting fluctuations in intercellular tension into cortical signaling (reviewed by Leckband DE and de Rooij J, 2014; reviewed by Hoffman BD and Yap AS, 2015). However, a clear picture of how

cadherin-based mechanobiology promotes tissue architecture remains to be determined (reviewed by Hoffman BD and Yap AS, 2015).
On the whole, it seems like alterations in E-cadherin, of any kind, can affect the cell on multiple levels, including those related to aberrant signaling.

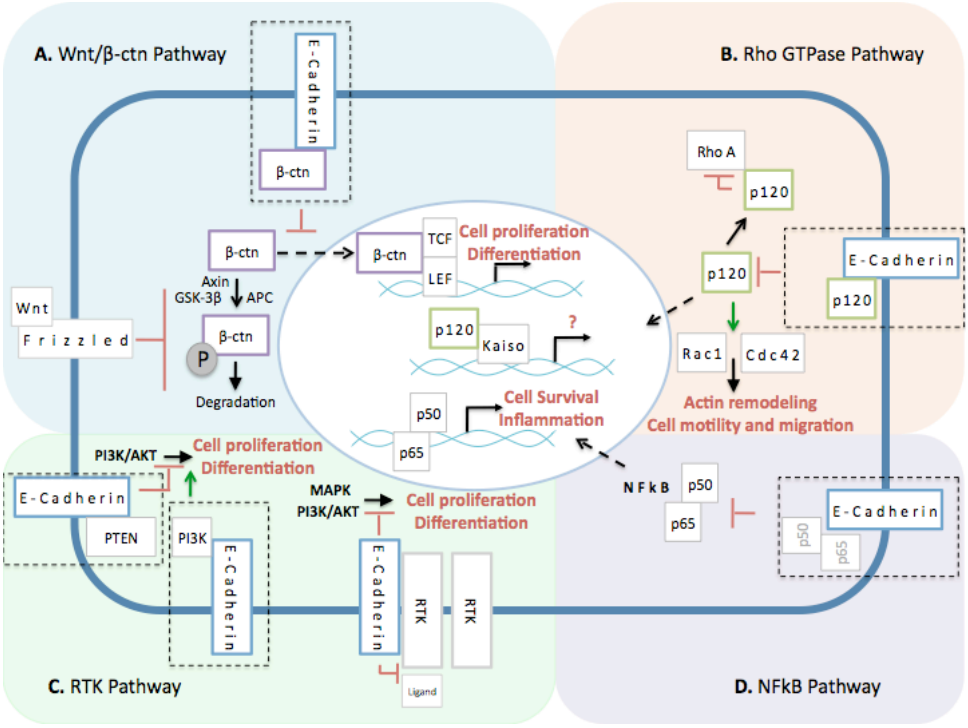

Figure 16. Signaling pathways involving E-cadherin/catenin complexes (adapted from Van den Bossche J *et al*., 2012; courtesy of V. Molinaro). **A) Wnt/β-catenin pathway**. In the absence of Wnt, cytoplasmic β-catenin is targeted for proteosomal degradation by a specific complex made up of GSK3β, APC and Axin proteins. When Wnt binds its receptor, Frizzled, β-catenin degradation is inhibited. Therefore, β-catenin can accumulate in the cytoplasm, enter into the nucleus and interact with transcription factors of the TCF/LEF1 family, activating the expression of target genes involved in cell proliferation and differentiation, including *CD44*, *c-MYC*, *cyclin D1*, and *MMP7*. E-cadherin might suppress Wnt/β-catenin signaling by sequestering β-catenin (β-ctn) at the cell membrane. **B) Rho GTPase pathway**. E-cadherin can regulate the action of Rho-family GTPases via p120-catenin. p120-catenin can repress RhoA and activate Rac1 and Cdc42, thus modulating the actin cytoskeleton and enhancing cell motility and migration. When p120 catenins accumulate in the cytoplasm, they can translocate to the nucleus and bind the Kaiso transcription factor. The functional implications of such transcriptional activation are still to be elucidated. **C) RTK pathway**. Growth factors are able to promote cell proliferation and prevent apoptosis through the binding to their receptors in the cell membrane, inducing dimerization of the receptors and concomitant activation of the intracellular tyrosine kinase domains. The activated Receptor Tyrosine Kinases (RTKs) can then phosphorylate their substrates, resulting in the activation of multiple downstream signaling pathways, including MAPK (mitogen-activated protein kinase) and PI3K/AKT signaling pathways. E-cadherin was found to interact through its extracellular domain with EGFR and other RTKs, thereby decreasing receptor mobility and ligand affinity. Moreover, depending on the context, the CCC (cadherin-catenin complex) could stimulate the PI3K/AKT pathway by recruiting and activating PI3K, or suppress this cascade by recruiting PTEN. **D) NF-κB pathway**. In mammals, the canonical NF-κB activation pathway mainly involves p65-p50 dimers. Following stimulation by inflammatory mediators (cytokines, microbial or endogenous danger-associated molecules), p65:p50 dimers translocate to the nucleus and activate the transcription of numerous target genes, including *Bcl-2*, *IL-6*, and *TNF*, consequently promoting cell survival and inflammation. The CCC may repress NF-κB pathway, by recruiting NF-κB transcription factors to the complex, possibly with p120-catenin as a docking site.

9. E-cadherin loss in tumor initiation and progression

E-cadherin-mediated cell-cell adhesion is a fundamental prerequisite for human development, growth and epithelial tissue homeostasis (reviewed by Schneider MR and Kolligs FT, 2015). In effect, the downregulation of E-cadherin signifies a hallmark of Epithelial-to-Mesenchymal Transition (EMT) (reviewed by Lamouille S *et al*., 2014). EMT, in itself, is a normal physiologically reversible process that is necessary for tissue remodeling and cell migration required for embryonic development; in adults, this process is also essential for wound healing. During EMT, the tightly connected epithelial cells are converted into more loosely adherent, fibroblast-like cells, which are endowed with migratory and invasive abilities. It is hence no surprise that the aberrant activation of this process by E-cadherin loss has been associated with metastasis of epithelial-based tumors and fibrosis (reviewed by Wirtz D *et al*., 2011; reviewed by Lee JY and Kong G, 2016). This indicates that E-cadherin loss generally occurs as a late event in tumorigenesis (Figure 17).

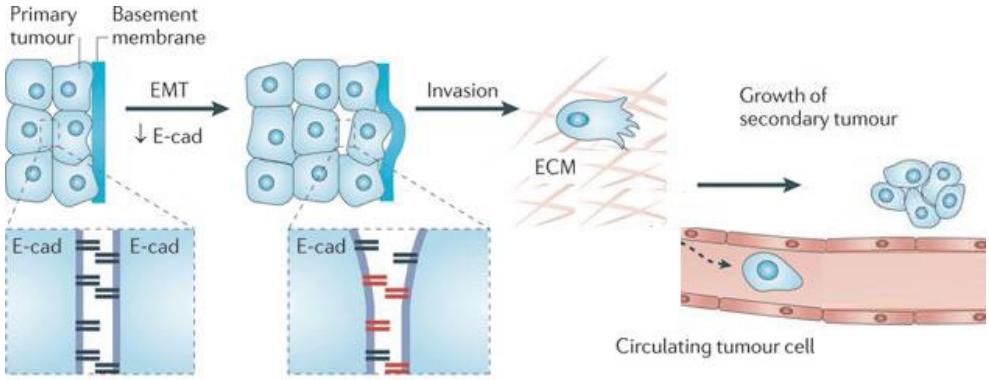

Figure 17. EMT, tumor invasion and metastasis. EMT entails the downregulation of E-cadherin that is coupled with the increased expression of mesenchymal neural cadherin (N-cadherin). Homotypic N-cadherin interactions (shown in red) are weaker than homotypic E-cadherin interactions (shown in blue) and facilitate cell migration and invasion. ECM: extracellular matrix (adapted from Wirtz D *et al*., 2011).

Nonetheless, the presence of *CDH1* germline mutations in HDGC patients, points towards a role of E-cadherin downregulation in the early stages of tumor development, independently of EMT. This was further substantiated by the work of Humar B and coworkers (2009); they found that the treatment of both wild-type and *cdh1*(+/-) mice with N-methyl-N-nitrosourea to promote gastric carcinogenesis, resulted in an 11 times higher frequency of intramucosal signet-ring cell carcinomas in the *cdh1*-knocked down mice than the wildtype. The authors proposed a model by which *CDH1* downregulation abrogates cell polarity,

which subsequently interferes with the positioning of the mitotic spindle and the segregation of cell fate determinants. Consequently, cell division occurs outside the epithelial plane, resulting in the accumulation of a portion of daughter cells in the lamina propria. The subsequent expansion and partial differentiation of this cell population, thus gives rise to signet ring cell foci (reviewed by Humar B and Guilford P, 2009) (Figure 18). While these foci can be considered as relatively indolent, they can progress to advanced disease in an unpredictable manner.

Figure 18. Model of the early development and differentiation of HDGC. Mis-orientation of the mitotic spindle can result in daughter cells dividing out of the epithelial plane and accumulating in the lamina propria, thereby forming foci of signet ring cells (from Humar B and Guilford P, 2009).

10. Mechanisms of E-cadherin deregulation in gastric cancer

E-cadherin is encoded by *CDH1*, which is a tumor suppressor gene (TSG), and, as for any classical TSG, effective dosage is finely regulated at multiple levels (Figure 19).

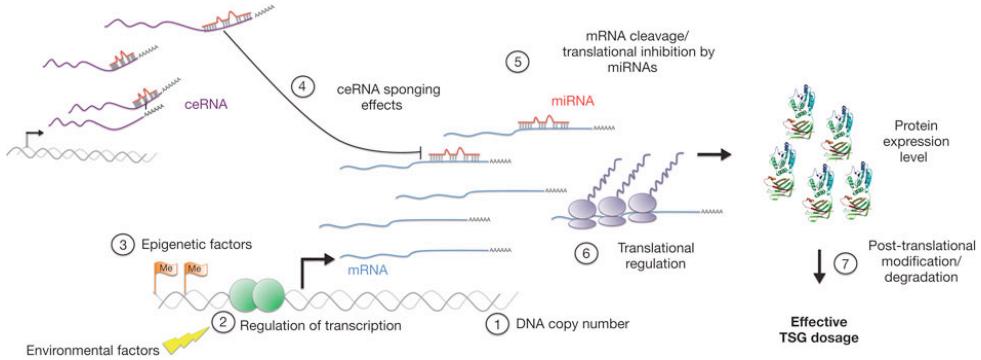

Figure 19. Mechanisms for regulation of tumor suppressor genes (from Berger AH *et al*., 2011).

For example, regarding transcriptional regulation, the binding of repressors including SNAI1 (Snail), SNAI2 (Slug), Twist, ZEB1 (δEF1), ZEB2 (SIP1) and

E47 (TCF3) to the E-box elements of the *CDH1* promoter leads to the downregulation of E-cadherin (reviewed by Peinado H *et al*., 2007; reviewed by Berx G and van Roy F, 2009). Moreover, Polycomb Repressive Complexes (PRCs) can regulate *CDH1* expression through reversible epigenetic silencing, which is mediated by the trimethylation of histone H3 at lysine 27 (H3K27me3) in nucleosomes surrounding the promoter region. This trimethylation is catalyzed by Enhancer of Zeste Homolog 2 (EZH2), which is the catalytic subunit of the Polycomb Repressive Complex 2 (PRC2) that functions alongside of the Suppressor of Zeste 12 homolog (SUZ12) (reviewed by Sparmann A and van Lohuizen M, 2006). A deregulation in any of these factors can disrupt the normal production levels of E-cadherin. In addition, *CDH1* germline and somatic alterations, microRNAs (miRNAs) and *H. pylori* infection, have been found to play distinct roles in *CDH1*-mediated gastric carcinogenesis.

10.1. *CDH1* genetic alterations

CDH1 genetic alterations have been reported in both sporadic and hereditary GC cases. As was previously mentioned, *CDH1* is a tumor suppressor gene, which indicates that the loss of E-cadherin expression necessitates the inactivation of both *CDH1* alleles following the "two hit model" (Knudson AG Jr, 1971). In hereditary cases, the *first hit* is already present as a constitutively mutated allele and a *second hit* is required at the somatic level to inactivate the wild-type allele. With respect to sporadic cases, the *first* and *second hits* must occur within the same somatic cell in the target tissue and lead to the reduction or complete absence of E-cadherin for cancer to develop. In accordance with that model, it is likely that malignancy occurs more frequently in heterozygous *CDH1* mutation carriers.

CDH1 germline mutations are common in hereditary cases of DGC; however, alterations encompassing this gene have also been found in both diffuse and intestinal sporadic cases (reviewed by Carneiro F *et al*., 2012). Interestingly, while somatic mutations in sporadic GCs cluster around exons 7 and 8 of *CDH1*, germline mutations tend to span the entire length of the gene (reviewed by Guilford P *et al*., 2010). The greater number of these mutations are nonsense, splice-site and frameshift that result in a truncated protein, while around 20-28% are missense leading to a single amino acid substitution (reviewed by Guilford P *et al*., 2010) (Figure 20). Up until today, more than 180 *CDH1* germline mutations have been uncovered in HDGC families of various ethnicities (reviewed by Monahan KJ and Hopkins L, 2016).

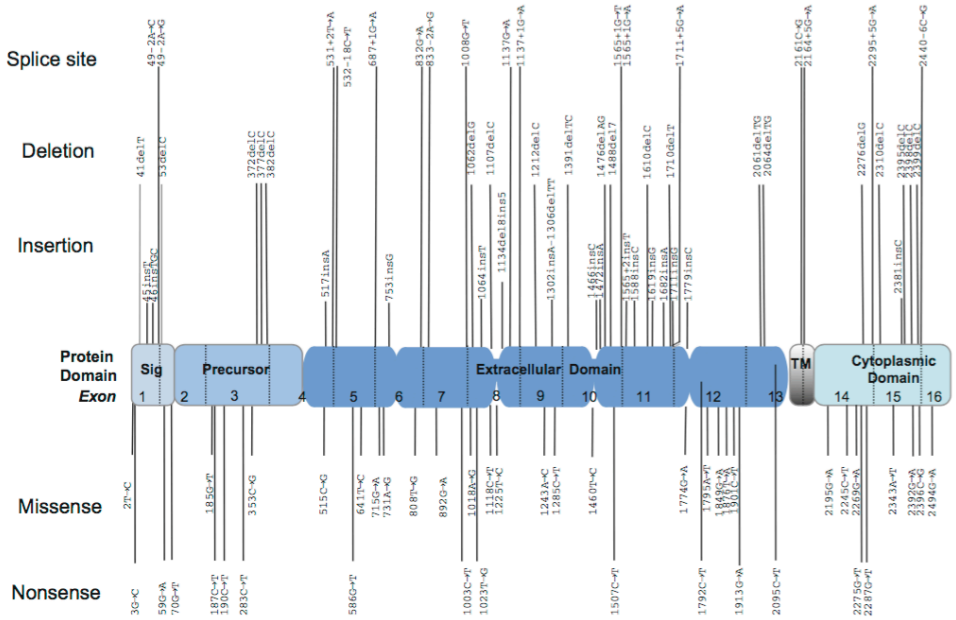

Figure 20. The location and subclassification of different *CDH1* germline mutations identified in HDGC families (from Guilford P *et al.*, 2010).

Apart from the clearly pathogenic mutations, a number of rare variants of unclear pathogenicity have been reported in selected individuals. These variants, termed as Variants of Uncertain Significance (VUSs), become a concern especially in the case of missense mutations and of both exonic and intronic base changes that may alter the splicing process or protein function. One way to assess the pathogenicity of VUSs would be by merging results from multiple sources including segregation analysis in affected families, *in silico* predictions, *in vitro* functional assays, screening of healthy controls, and RNA transcript analyses (reviewed by Oliveira C *et al*., 2013; reviewed by Sijmons RH *et al*., 2013).

The expansion of *CDH1* germline alteration testing beyond gene sequencing has identified novel germline defects in a fraction of HDGC families (Oliveira C *et al*., 2009a,b; reviewed by Sugimoto S *et al*., 2014; Molinaro V *et al*., 2014). Indeed, the use of Multiplex Ligation-dependent Probe Amplification (MLPA) assay showed that approximately 5% of families carried a gene deletion (Oliveira C *et al*., 2009a, reviewed by van der Post RS *et al*., 2015), while the SNuPE (Single-Nucleotide Primer Extension) assay, identified allelic imbalance (AI) in a number of cases, where the expression of one *CDH1* allele was reduced or completely lost. Note that *CDH1* promoter hypermethylation or germline deletions explained the AI only in a small number of cases, indicating the presence of so far undetected

mechanisms (Pinheiro H *et al*., 2010; Molinaro V *et al*., 2014).
With respect to the *second hit*, it most commonly occurs in the form of somatic hypermethylation of a CpG island found in the *CDH1* gene promoter (50-70% of cases). On the other hand, mutations and large deletions, the latter detectable as loss of heterozygosity (LOH), are less frequently found (Grady WM *et al*., 2000; Oliveira C *et al*., 2004, 2009b; reviewed by Carneiro F *et al*., 2012; Corso G *et al*., 2013).
Similarly, somatic *CDH1* sequence mutations are rare in the sporadic context of GC. In sporadic cancers, the most common alteration occurs in the form of epigenetic silencing by promoter hypermethylation, with varying frequencies between DGCs (50 to 83%) and IGCs (6.25 to 50%). Moreover, LOH has also been observed in 11-39% of sporadic DGCs and 36-46% of sporadic IGCs, respectively (Machado JC *et al*., 2001; Ascaño JJ *et al*., 2001; Corso G *et al*., 2013).
Importantly, sequence inactivating mutations, promoter hypermethylation and allelic deletions fail to explain a large fraction of HDGC and sporadic GCs where aberrant E-cadherin protein expression or downregulation is present without an unequivocal genetic determinant. This again highlights the existence of mechanisms for E-cadherin downregulation that are yet to be accounted for (Pinheiro H *et al*., 2012; Carvalho J *et al*., 2012).

10.2. miRNA-mediated E-cadherin deregulation

miRNAs are small (~22 bp) nucleic acids that function by modifying the expression of downstream target genes. These non-coding RNAs act post-transcriptionally by binding complementary sites within the 3'-UTR of target mRNAs, where a single miRNA can recognize more than one mRNA and *vice versa*. Based on the level of complementarity between a miRNA and its target, it can either degrade the mRNA (complete complementarity) or inhibit its translation (partial complementarity). They are, therefore, strong regulators of gene expression enabling cells to quickly respond to a novel fate or environment (reviewed by Bartel DP, 2004; reviewed by De Craene B and Berx G, 2013).
In cancer, miRNAs can be classified as oncogenic (oncomiRs) or tumor suppressors (tsmiRs) based on the outcome of the target mRNA; where the increased activity of oncomiRs inhibits tumor suppressor genes, and the decreased level of tsmiRs increases oncogene translation, thus promoting tumor progression (reviewed by Song S and Ajanirole JA, 2013). The deregulation of multiple oncomiRs and tsmiRs involved in cell growth, apoptosis, invasion and metastasis, have been reported in gastric cancer (reviewed by Song S and Ajanirole JA, 2013; reviewed by Jiang C *et al*., 2015; reviewed by Suárez-Arriaga MC *et al*., 2016) and

a difference in miRNA expression pattern has been detected in DGC compared to IGC (Ueda T *et al*., 2010; reviewed by Suárez-Arriaga MC *et al*., 2016).
Among the reported tsmiRs affecting *CDH1* expression are the members of the *miR-200* family and *miR-101*, which have been found to be significantly downregulated in GCs (Wang HJ *et al*., 2010; Zhou X *et al*., 2015; Chang L *et al*., 2015). In particular, *miR-101* (targeting the *CDH1* inhibitor, EZH2) was reported by Carvalho J and collaborators (2012) to be downregulated, due to deletions or microdeletions in the gene, in around 60% of tumors, which were mainly of the intestinal histotype. A concomitant overexpression of EZH2 was reported in 40% of cases and was associated with loss or aberrant expression of E-cadherin (Carvalho J *et al*., 2012).
With respect to members of the miR-200 family, they have been demonstrated to target the *CDH1* transcriptional repressors, ZEB1 and ZEB2 (Korpal M *et al*., 2008). More recently, it was shown that miR-200b also targets *SUZ12*, and the downregulation of this miRNA in gastric cancer cells results in the repression of E-cadherin as a result of the simultaneous increase in SUZ12 levels (Cui Y *et al*., 2013).
The sudden boom in investigating miRNA deregulation in cancer has uncovered other potential indirect regulators of E-cadherin, including miR-153 (targeting Snail) and miR-203 (targeting Slug), both of which were found to be downregulated in GC specimens compared to controls (Wang Z and Liu C, 2015; Shi Y *et al*., 2015). Further *in vitro* assays determined that the downregulation of either miRNA promoted metastasis by increasing the cell invasive capacity (Wang Z and Liu C, 2015; Shi Y *et al*., 2015).
Among the oncomiRs that are predicted to directly target E-cadherin expression, the miR-23 family members, miR-23a and miR-23b, have been found to be significantly upregulated in gastric cancer tissue compared to the normal mucosa. This upregulation was further correlated with poor prognosis in patients, and tumor progression in *in vitro* assays (Ma G *et al*., 2014). In addition to that, miR-199 (Zhao X *et al*., 2014) and miR-544a (Yanaka Y *et al*., 2015) have been found to modulate E-cadherin expression *in vitro* using GC cell lines, however, much like a number of other miRNAs, their relevance in gastric cancer remains to be elucidated.

10.3. *H. pylori*-mediated E-cadherin deregulation

CDH1 has been shown to be a target of *H. pylori*. The presence of this bacterium has been associated with epigenetic silencing of the *CDH1* promoter that would ultimately downregulate *CDH1* expression and consequently E-cadherin protein levels (Chan AO *et al*., 2003; Huang FY *et al*., 2012). In addition to that, *H. pylori* infection has also been attributed to E-cadherin ectodomain shedding, thereby

disrupting the adherens junctions (reviewed by Grabowska MM and Day ML, 2012). Indeed, IGC patients have been found to harbor high soluble E-cadherin serum levels (Juhasz M *et al*., 2003). The E-cadherin shedding mechanism was thought to be mainly due to *H. pylori*-induced expression of the host disintegrin and metalloprotease, ADAM-10. However, later it was demonstrated that siRNA-silencing of ADAM-10, in addition to the inhibition of other host metalloproteases, only slightly affected the ability of *H. pylori* to cleave E-cadherin's ectodomain. On the other hand, a major contribution of HtrA (high-temperature requirement A), a secreted bacterial serine protease, could be detected (Hoy B *et al*., 2010). Further work by Schmidt TP and collaborators (2016), pinned down the E-cadherin signature motifs that function as target sites for the HtrA-mediated proteolytic degradation and give rise to a defined fragmentation pattern of the protein *in vitro*. They also took an extra step in designing substrate-based peptide inhibitors that could bind HtrA and prevent it from degrading the E-cadherin ectodomain, thus impeding *H. pylori* transmigration. As such, it would seem that altered E-cadherin expression, whether at the transcriptional or the protein levels, is a crucial step in *H. pylori*-induced gastric carcinogenesis, further validating the significant contribution of the *CDH1* gene to this disease.

11. New mechanisms regulating E-cadherin expression

The complexity of *CDH1* gene regulation is governed by an interplay of multiple *cis*-regulatory elements dispersed throughout the locus (Alotaibi H *et al*., 2015). While most of the focus has been on the coding portion of the gene, intron 2 has emerged as a significant element contributing to the regulation of gene expression (Stemmler MP *et al*., 2005; reviewed by Paredes J *et al*., 2012). This large intron is more than 63 kb in length and is a structurally conserved feature in mammals (reviewed by Paredes J *et al*., 2012). Studies in mice have shown that intron 2 is indispensible for the activation of *Cdh1* gene transcription and its maintenance during development; moreover, the *Cdh1* promoter alone was demonstrated to be inadequate for conferring cell-type specific E-cadherin expression (Stemmler MP *et al*., 2003; 2005). Recently, Alotaibi H and collaborators (2015) showed the presence of 2 enhancers within the +7.8 kb and +11.5 kb regions of *Cdh1* intron 2. These enhancers bound the transcription factors Grhl3 and Hnf4α, respectively, in response to MET (Mesenchymal-to-Epithelial Transition) signals and their cooperation orchestrates E-cadherin re-expression (Figure 21).

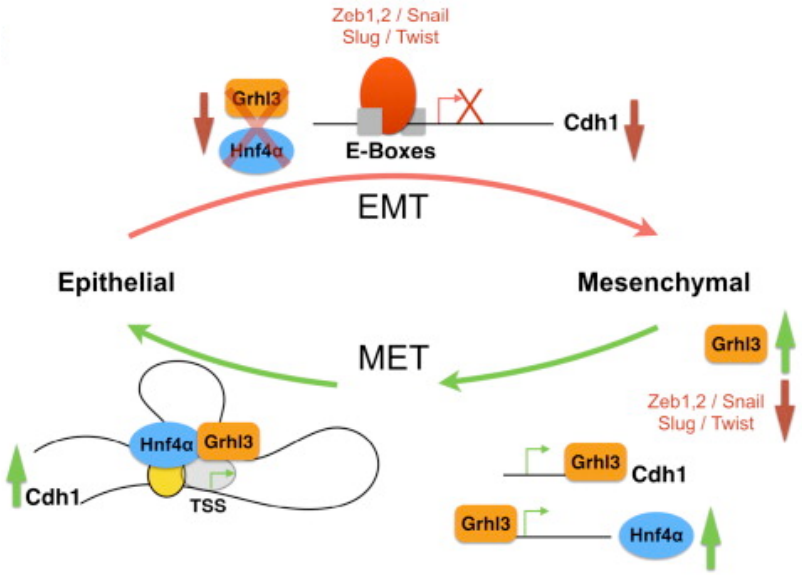

Figure 21. Proposed model of Grhl3 and Hnf4α mediated looping and activation of the *Cdh1* locus during MET. During EMT activators of E-cadherin expression are downregulated and transcription at the *Cdh1* locus is blocked by at least one of the EMT inducers Zeb1, Zeb2, Snail, Slug and Twist by binding to E-boxes at the promoter. Upon TGFβ3 withdrawal and MET induction, Grhl3 is binding to sites at *Cdh1* locus and to the Hnf4a protein. Subsequent expression of Hnf4α leads to recruitment of Grhl3 and Hnf4α to intronic enhancers that induces DNA looping by interaction of the two factors and interaction at the TSS that involves PolII (gray) and p300 (yellow) (enhancer cooperativity). The assembly and stabilization of the core transcription machinery leads to induction of E-cad expression. Up- and downregulation is indicated by vertical green and red arrows, respectively (from Alotaibi H *et al*., 2015)

In addition to MET, studies on the reverse process (*i.e.* EMT), provided novel insights into the molecular determinants controlling the transcriptional activity of *CDH1* locus and validated the existence of *cis*-regulatory elements within intron 2 of the gene (Alotaibi H *et al*., 2015).

Apart from enhancer elements, annotations produced by EST (Expressed Sequence Tags) and CAGE (5' cap analysis gene expression) support the existence of new areas of transcription within *CDH1* intron 2 (reviewed by Paredes J *et al*., 2012). In particular, a study by Pinheiro H and collaborators (2012) uncovered and characterized novel transcripts arising from within human *CDH1* intron 2. They first realized that upon performing quantitative PCR with proper primers, the ratio between PCR products containing *CDH1* exons 6-7 and products containing *CDH1* exons 1-2 was greater than 1. This suggested that somewhere downstream of exon 2, there is a region responsible for transcribing non-canonical transcripts. Further work guided by *in silico* databases reporting ESTs, led to the identification of two new exons within intron 2. Both exons were found to splice with exon 3 at its

canonical splice site: the two resulting transcripts, including all the remaining exons and terminating with the canonic sequence, were named *CDH1a* and *CDH1b*. Upon amplification by PCR, two other transcripts, labeled *CDH1j* and *CDH1b-10*, were found to arise from *CDH1a* and *CDH1b*, respectively; this was probably due to alternative splicing (Figure 22).

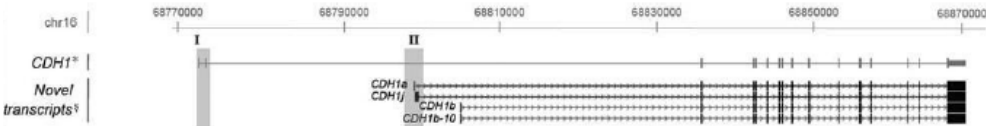

Figure 22. Transcripts arising from the transcription starting within *CDH1* intron 2 (adapted from Pinheiro H *et al*., 2012).

Among the reported transcripts, *CDH1a* was expressed in multiple normal tissues, including spleen, colon and breast, while being completely absent in the stomach. Furthermore, this transcript harbored properties that enabled its translation into a protein isoform that differed from the canonical E-cadherin protein only in its N-terminal domain. In particular, the amino acid sequences corresponding to exons 1 and 2 of E-cadherin were replaced by the sequence corresponding to exon 1a in CDH1a (Figure 23).

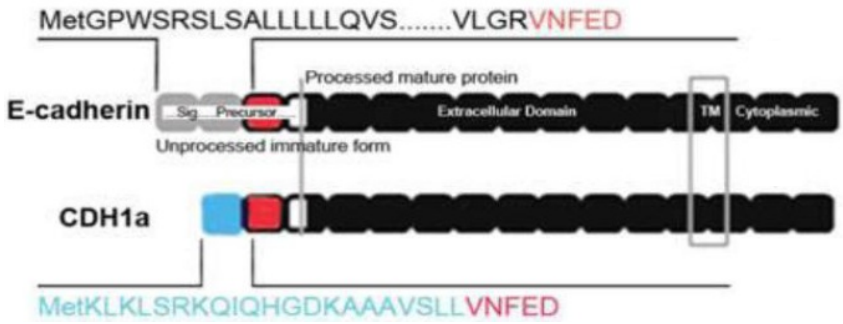

Figure 23. The canonical protein, E-cadherin, and the putative CDH1a isoform. In grey (exons 1-2) and blue (exon 1a) are the portions specific to E-cadherin and CDH1a, respectively; in red and white are the common regions that are cleaved during protein processing (cleavage site is marked by grey line); in black are the expected mature proteins (adapted from Pinheiro H *et al*., 2012).

These findings made *CDH1a* especially interesting and prompted further quantitative and functional assays; such assays determined that in the case where *CDH1a* was overexpressed in an E-cadherin negative context, using CHO (Chinese hamster ovary) cells, it mimicked the canonical protein function by inducing cell-cell aggregation and reducing cell invasion capacity. On the other hand, the co-expression of *CDH1a* with canonical E-cadherin in MNK28 gastric cancer cells

promoted cell invasion and angiogenesis. The pro-invasive capacity of *CDH1a* was predicted to occur through the induction of the interferon-induced genes *IFITM1* and *IFI27*, which were upregulated *in vitro* upon *CDH1a* overexpression. In addition to that, while *CDH1a* was never identified in the normal stomach, it was found to be *de novo* expressed in GC cell lines (Pinheiro H *et al*., 2012).
Taken together, these data point towards a potential role of the *CDH1a* non-canonical isoform in the manifestation of gastric cancer in general, and intestinal gastric cancer, in specific, where some level of E-cadherin is often retained.

12. Other genes implicated in HDGC predisposition

CDH1 germline alterations have been detected in around 25-50% of HDGC cases (reviewed by Fitzgerald RC *et al*., 2010). The detection rate has been shown to drop to 10-19% in countries with moderate to high GC incidence (reviewed by van der Post RS *et al*., 2015), including Italy (Molinaro V *et al*., 2014), and countries with mixed ethnic backgrounds (Hansford S *et al*., 2015). On the other hand, the detection rate increases considerably in the Maori population of New Zealand, and even more so in the Newfoundland population, in which a founder mutation has been reported (Kaurah P *et al*., 2007; reviewed by van der Post RS *et al*., 2015). This variability in the detection rate indicates that HDGC selection criteria can be less effective when applied in certain populations where genetic makeup and/or environmental factors contribute to the early onset of DGC and LBC and/or to familial clustering.
Moreover, other cancer susceptibility genes could be involved in a fraction of *CDH1* mutation-negative HDGC cases. In fact, in the last three years a number of studies have undertaken exome- and gene panel- sequencing strategies to uncover new genes predisposing to HDGC (Majewski IJ *et al*., 2013; Gaston D *et al*., 2014; Donner I *et al*., 2015; Hansford S *et al*., 2015).
The genes reported by exome sequencing are depicted in Table 5.

Gene	Mutation	Location	Mutation type	Ethnicity	Study type	Frequency
CTNNA1	c.76delGA	Chr 5: 138117693	Nonsense	No data	Family study	1/1 family
MAP3K6	c.598G>T	Chr 1: 27690792	Missense	Canada	Family study and case series	1/1 family 1/115 cases
MAP3K6	c.620T>G	Chr 1: 27690770	Missense	No data		No data
MAP3K6	c.2837C>T	Chr 1: 27684750	Silent	No data		No data
MAP3K6	c.2872C>A	Chr 1: 27684715	Missense	No data		No data
MAP3K6	c.2544delC	Chr 1: 27685238 - 27685239	Nonsense	Portugese		1/115 cases
INSR	c.3937 G>A	Chr 19: 7117279	Missense	Finland	Family study	1/1 family
FBXO24	c.242G>C	Chr 7: 100187900	Missense	Finland		1/1 family
DOT1L	c.3437C>T	Chr 19: 2223326	Missense	Finland		1/1 family

Table 5. Summary of non-*CDH1* germline mutations reported in HDGC patients (adapted from Tan RY and Ngeow J, 2015).

Among those genes, *MAP3K6* (Gaston D *et al*., 2014) and *CTNNA1* (Majewski IJ *et al*., 2013) seem more likely to be causative of DGC than *INSR*, *FBXO24* and *DOT1L*, in which missense variants have only been reported in single Finnish families (Donner I *et al*., 2015).

With respect to *MAP3K6*, a germline coding variant was reported in a large family from Maritime Canada presenting with *CDH1* mutation negative familial DGC. This variant was predicted to be damaging by several *in silico* programs and it was also identified, along with 4 other coding variants, upon screening of 115 unrelated probands. Interestingly, somatic second hits within *MAP3K6*, in the form of a single nucleotide variant (p.H506Y) and of DNA hypermethylation, were also reported in tumor specimens of two affected individuals. These findings, coupled with previous implication of *MAP3K6* as a tumor suppressor gene in mouse models, and reports of gene mutations in both gastric cancer cell lines and primary GC tumors, strengthen the involvement of *MAP3K6* in DGC susceptibility (Gaston D *et al*., 2014).

Nonetheless, probably the most relevant gene to arise from next generation sequencing is *CTNNA1*. This gene encodes alpha-E-catenin, which functions in the same complex as the *CDH1* encoded protein, E-cadherin. A germline truncating mutation in *CTNNA1* was first reported in a large Dutch family that presented with *CDH1* mutation negative HDGC (Majewski IJ *et al*., 2013). The second allele was found to be silenced in two cases for which tumor tissue could be screened (Majewski IJ *et al*., 2013). Four other carriers of the *CTNNA1* mutation did not develop DGC, indicating incomplete penetrance (Majewski IJ *et al*., 2013). Since no mutations in *CTNNA1* had been identified in a previous study on 22 *CDH1*

mutation-negative families from Canada, USA and UK (Schuetz JM *et al*., 2012), the finding of Majewski IJ and collaborators pointed towards the presence of a founder effect for the Dutch population. This, however, was countered by the recent detection of *CTNNA1* truncating mutations in two unrelated individuals of Italian and of "white" ethnicities, suggesting that mutations in this gene can act as a genocopy of *CDH1* in a subset of HDGC patients (Hansford S *et al*., 2015).
The finding of Hansford S and collaborators derived from sequencing of a gene panel representing 55 genes involved in upper gastrointestinal (GI) cancers or cancer susceptibility syndromes (Hansford S *et al*., 2015). By using this panel, the authors uncovered potentially pathogenic variants in 16 out of the 144 HDGC probands analyzed (11.1%) (Figure 24).

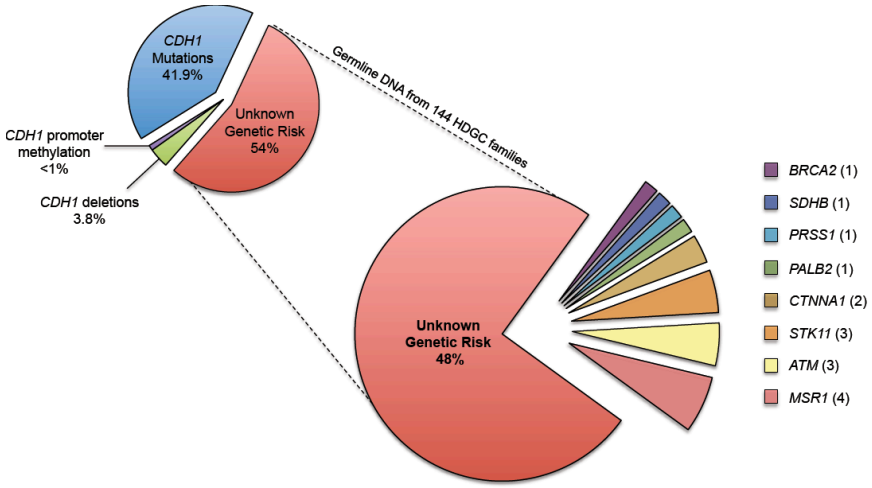

Figure 24. HDGC mutational profile, including novel potentially pathogenic mutations. Number of families found to carry these variants is indicated in parentheses (from Hansford S *et al*., 2015).

Other than *CTNNA1*, they reported pathogenic missense variants in *SDHB* (Peutz-Jeghers syndrome), *STK11* (Cowden-like syndrome) and *MSR1* (esophageal cancer) genes, as well as a truncating mutation in *PRSS1*, a gene associated with hereditary pancreatitis. Heterozygous truncating mutations were additionally uncovered within *ATM* and *PALB2* genes, which are predicted to have low to moderate penetrance in the upper GI tract. Finally, a truncating mutation in *BRCA2*, which commonly predisposes to breast and ovarian cancer, was also reported in one proband. This last finding is not totally unexpected, given that some *BRCA2*-positive breast cancer families have been shown to over-represent gastric cancer (Jakubowska A *et al*., 2002; Moran A *et al*., 2012).
In spite of recent advancements, the genetic underpinnings of HDGC remain

undetermined in at least 50% of cases. In this scenario, the routine application of next generation sequencing for HDGC genetic testing could be highly informative in terms of better defining the molecular basis of the disease and providing a step forward in genetic counseling of individuals at risk. Nonetheless, until a mutation within a newly identified susceptibility gene could be replicated in multiple families, it is not really feasible to attribute pathogenicity and confer management guidelines (van der Post RS *et al*., 2015). Pending such improvement, germline alterations in the *CDH1* gene remain the principal actionable drivers of HDGC.

13. Translating findings from research to clinical practice

To this day gastric cancer remains a prevalent disease with a highly poor prognosis, indicating that measures focused on prevention and early diagnoses are warranted (reviewed by den Hoed CM and Kuipers EJ, 2016). Among those measures could be the application of endoscopic screening, which, at a population-level, is only practical in high-risk countries due to high costs. Another strategy that has been considered is the eradication of *H. pylori*. While the consequence of population-based treatment against this bacterium remains to be investigated, recent data have shown that the eradication of *H. pylori* at any stage of gastritis decreases gastric cancer incidence (reviewed by den Hoed CM and Kuipers EJ, 2016). However, patients' response to *H. pylori* eradication therapy most often fails due to non-compliance or antibiotic resistance.

In the case of HDGC families, cancer-preventive measures have been implemented using an algorithm that takes into consideration family history, tumor histology, eligibility for genetic test (based on the established guidelines) and *CDH1* mutation status (Figure 25). Individuals at risk are then managed accordingly, with prophylactic gastrectomy ultimately being the only curative option. This procedure consists of surgically removing the entire stomach, and is preferentially postponed by certain at-risk patients who opt for annual endoscopic surveillance.

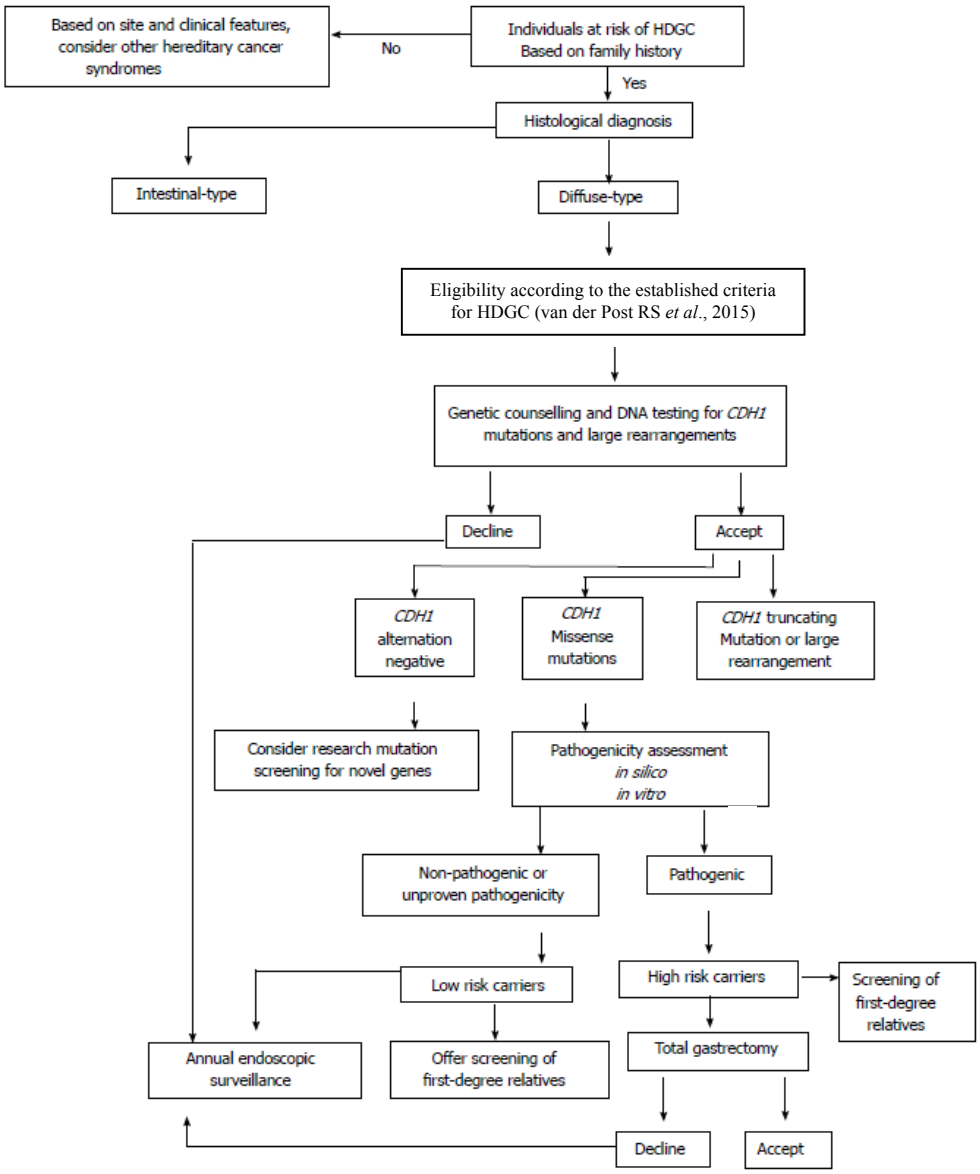

Figure 25. Algorithm for the clinical management individuals suspected of having HDGC (adapted from Tan RY and Ngeow J, 2015).

Conventional therapeutic strategies in gastric cancer consist of surgery, chemotherapy, and radiotherapy, all of which are applied homogeneously and with only modest success to all gastric cancer subtypes. The recent enrichment of our

knowledge on the molecular etiology and heterogeneity of gastric cancer allowed the development of targeted therapies (reviewed by Jou E and Rajdev L, 2016; reviewed by Corso S and Giordano S, 2016). In particular, criteria have been established for selecting HER2-positive gastric cancer patients (Figure 26), 20% of whom have been found to benefit from treatment with the anti-HER2 monoclonal antibody (mAb), Trastuzumab.

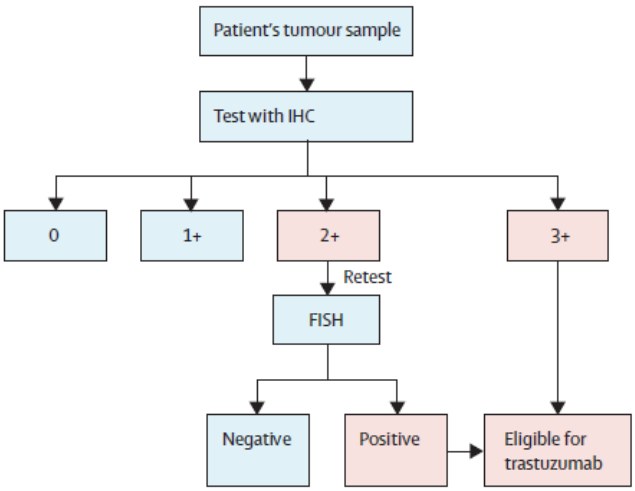

Figure 26. Testing algorithm for HER2 status in GC. IHC: immunohistochemistry; FISH: fluorescence in-situ hybridization (adapted from Van Cutsem E *et al*., 2016).

Apart from Trastuzumab, one other mAb, Ramucirumab, which targets the pro-angiogenic factor VEGFR-2, has been approved by the FDA (Food and Drug Administration) and the European Union for advanced gastric cancer. Nonetheless, the majority of patients have experienced little benefit from both drugs. Moreover, phase II and III clinical trials targeting a number of RTKs, which are genetically altered in GC, gave inadequate results (reviewed by Corso S and Giordano S, 2016). The lack of selection criteria for carriers of alterations in the concerned genes, coupled with the absence of harmonized testing protocols, might explain the disappointing failures of clinical trials in these cases (reviewed by Corso S and Giordano S, 2016). Therefore, further work is required to better predict drug response and delineate patient cohorts who are more likely to benefit from targeted therapy alone, or in combination with cytotoxic agents (reviewed by Jou E and Rajdev L, 2016). In parallel, a deeper investigation of the molecular mechanisms underlying GC is required to uncover novel putative therapeutic targets that provide an unmet need for patients who do not benefit from, or are unable to tolerate, current treatment options.

First Aim of the Research

1. Molecular characterization of *CDH1* mutation-negative patients with Hereditary Diffuse Gastric Cancer (HDGC) syndrome

In the first part of my research, the focus was on better characterizing the underlying mechanism for HDGC in a cohort of *CDH1* mutation-negative Italian patients. Thirty such cases were subjected to *CDH1* canonical and non-canonical transcripts' analysis to determine whether aberrant splicing could be involved. In addition, as part of a much larger study, the 7 most recently recruited cases underwent exome sequencing of a panel of 94 cancer-related genes. An *in silico* analysis was subsequently performed on all variants arising from sequencing to uncover whether other genes could be causative of HDGC in the investigated cohort.

Materials and Methods

1.1. Selection of HDGC patients

In total 35 HDGC patients were selected, in collaboration with IEO (Istituto Europeo di Oncologia) of Milano (MI-Italy), from Italian families that have been diagnosed based on the modified criteria of Brooks-Wilson and coworkers (2004): (i) family history with two or more first degree relatives with GC, with at least one documented DGC diagnosed before 50 years of age; (ii) family history with three or more first/second degree relatives with GC, diagnosed at any age, with at least one documented case of DGC; (iii) an individual diagnosed with DGC at <45 years of age, regardless of family history; (iv) family history with two or more first/second degree relatives with LBC, with or without DGC.

Among the 35 probands, 23 have been previously shown to be *CDH1* mutation-negative by our group (Molinaro V *et al*., 2014) using parallel complementary molecular approaches for DNA and RNA analysis, as well as bioinformatics analysis. These approaches were utilized to detect *CDH1* sequence variations (DNA sequencing), large gene deletions (Multiplex Ligation-dependent Probe Amplification, MLPA), allelic expression imbalance (Single Nucleotide Primer Extension, SNuPE), allelic silencing by promoter hypermethylation (bisulfite sequencing) and splicing defects (Reverse-Transcription PCR, RT-PCR).

In addition, 7 HDGC patients that have been recently recruited by IEO and shown to be *CDH1* mutation-negative (by sequencing the 16 *CDH1* canonical exons and exon-intron boundaries, and by MLPA analysis) were included in this study.

Blood samples from all patients were obtained upon approval by the Ethics Committee of IEO for *CDH1* transcripts' analysis and exome sequencing of the TruSight® Cancer Sequencing Panel. This study also included the blood of 5 healthy controls provided by the Department of Biology and Biotechnology, University of Pavia (PV-Italy).

1.2. DNA, RNA and cDNA preparations from blood

For exome sequencing, genomic DNA (gDNA) was required. This gDNA was extracted from blood samples of the 7 most recently recruited HDGC probands, by the QIAampVR DNA Blood Mini Kit (Qiagen) and QIAcubeTM purification system, according to the general protocols provided by the manufacturers.

With respect to *CDH1* transcripts' analysis, RNA samples from the 30 HDGC patients and those from 5 healthy controls were required. To reduce *in vitro* RNA degradation and to minimize gene induction, blood from all subjects was collected in PAXgeneTM Tubes (Qiagen). Total RNA was extracted within a few hours using the PAXgene Blood RNA Kit (Qiagen) according to manufacturer's

instructions. The quality and quantity of nucleic acids was then assessed with the Thermo Scientific NanoDropTM 1000 Spectrophotometer. Moreover, the integrity of the RNA was verified by evaluating the 2:1 ratio of 28S:18S rRNA (ribosomal RNA) bands on agarose gel electrophoresis.
The extracted RNA was subsequently reverse-transcribed with the SuperScript First-Strand Synthesis System for RT-PCR (Invitrogen™). In specific, cDNA was synthesized using about 800 ng of total RNA in a 20 µl reaction catalyzed by the SuperScript III Reverse Transcriptase and primed with random hexamers based on the manufacturer's protocol.

1.3. Bioinformatics analysis of the *CDH1* transcripts

The RNA coding sequences corresponding to *CDH1a* and *CDH1j* transcripts were obtained starting from the protein sequence and genomic coordinates reported by Pinheiro H and collaborators (2012); the human *CDH1* gene sequence present in the Ensembl database (http://www.ensembl.org) and the corresponding coding sequence reported in the CCDS database (http://www.ncbi.nlm.nih.gov/CCDS/CcdsBrowse.cgi) were used as references. As can be seen in Figure 5, both *CDH1a* and *CDH1j* splice with canonical exon 3 splice site and include all downstream *CDH1* canonical exons. As such, the noncanonical transcripts differ from *CDH1* only in the initial portion of the coding sequence, where exons 1 and 2 of *CDH1* are replaced by exon 1a and exon 1j of *CDH1a* and *CDH1j*, respectively. However, exon1a is not unique to the *CDH1a* transcript as it also forms the initial part of exon 1j (Figure 1).

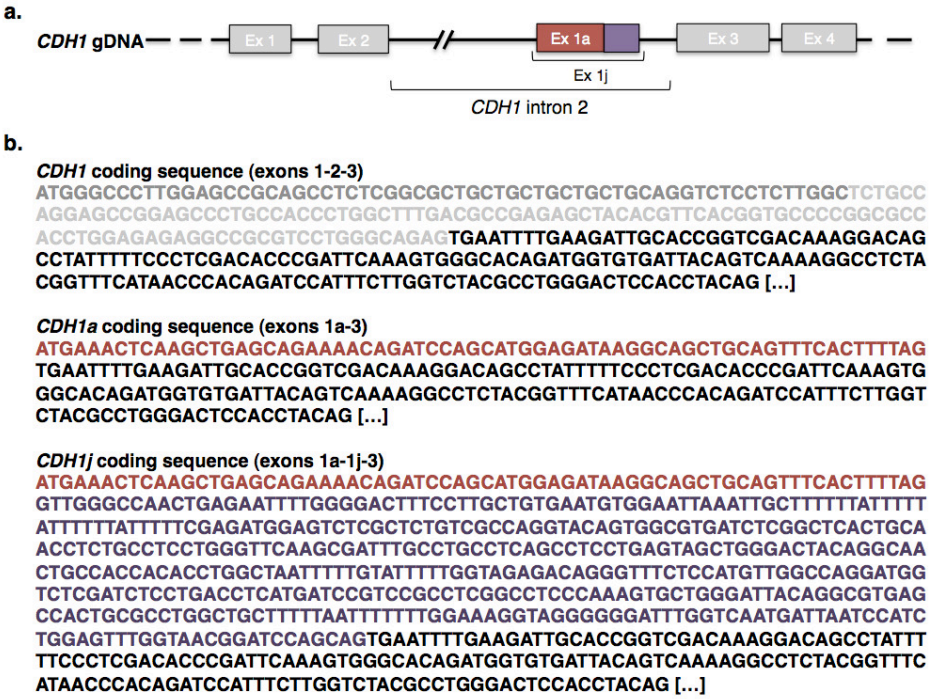

Figure 1. *CDH1* gene and related transcripts. a. Scheme of *CDH1* gDNA. In grey are canonical exons; in red is the *CDH1a* non-canonical exon; in purple is the *CDH1j*-specific portion of exon 1j. b. *CDH1*, *CDH1a* and *CDH1j* coding sequences. In grey are canonical exons 1 and 2; in red is the *CDH1a* non-canonical exon; in purple is the *CDH1j*-specific portion of exon 1j. Ex: exon.

1.4. Two-step RT-PCR amplification of *CDH1* transcripts

The cDNA obtained following the aforementioned reverse transcription was used for the amplification of the canonical *CDH1* and non-canonical *CDH1a* and *CDH1j* transcripts in the HDGC probands and controls. In order to selectively amplify the three targets, the forward primers were designed at the exon 1a-exon 3 junction for *CDH1a*, within the *CDH1j* specific portion of exon 1j for *CDH1j* and within exon 2 for *CDH1*. With respect to the reverse primers, all three were designed complementary to different regions of exon 3 (Figure 2).

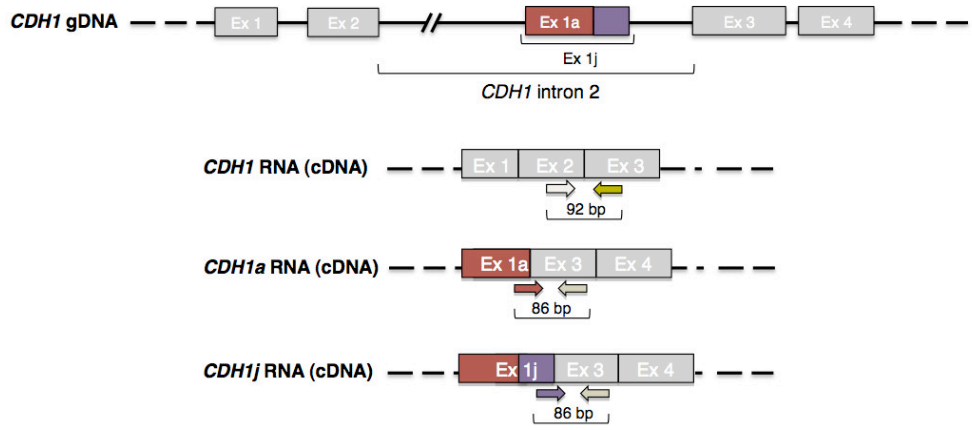

Figure 2. *CDH1* gene and transcripts with the primers used for their selective amplification. Primers are depicted by single arrows; the sizes of resulting amplicons are indicated underneath each transcript. Ex: exon.

In addition to that, PrimerQuest tool found on the Integrated DNA Technologies (IDT) website (http://www.idtdna.com/primerquest/home/index) was used to design the necessary primers for the amplification of *GAPDH*, which was the housekeeping gene included to control for cDNA quality. The specificity of all primers was further confirmed with the online BLAST (Basic Local Alignment Search Tool) (http://blast.ncbi.nlm.nih.gov/Blast.cgi) prior to being ordered from Metabion. All primers are depicted in Table 1.

Gene	Primers (5' to 3')	Amplicon size (bp)
CDH1	F: CTTTGACGCCGAGAGCTACACG R: TGAATTTTGAAGATTGCACCG	92
CDH1a	F: GCTGCAGTTTCACTTTTAGTG R: CTCGACACCCGATTCAAAGT	86
CDH1j	F: TTGGAAAGGTAGGGGGGAT R: TGAATTTTGAAGATTGCACCG	86
GAPDH	F: ATGTTCGTCATGGGTGTGAAC R: CAAGATCATCAGCAATGCCTC	65

Table 1. The sequences of primers amplifying *CDH1*, *CDH1a*, *CDH1j* and *GAPDH*, and sizes of the amplicons.

PCR reactions were subsequently performed on 2 µl cDNA of each patient and control samples using AccuPrime™ *Taq* DNA Polymerase System (Invitrogen™) coupled with the specific primers designed for the amplification of *CDH1*, *CDH1a*, *CDH1j* and *GAPDH* respectively (Table 2).

Reaction Component	Final concentration	Volume (µl)
AccuPrime™ PCR Buffer I (10X)	1X	1.5
Primer F (5 µM)	0,4 µM	1.2
Primer R (5µM)	0,4 µM	1.2
AccuPrime™ *Taq* DNA Polymerase	-	0.3
≈ 40 ng/µl cDNA template	≈ 80 ng	2
RNase-free water	-	to 15 µl

Table 2. Concentrations and volumes of the reagents used for two-step RT-PCR reactions.

The exact thermal cycling conditions are further depicted in Table 3.

Cycle steps	Temperature	Time
Initial denaturation	94°C	2 min
Denaturation	94°C	15 sec
Annealing	61°C	30 sec
Extension	68°C	30 sec

45 cycles

Table 3. Thermal cycling conditions for two-step RT-PCR reactions.

At the end of the reactions, PCR products were visualized on a 2.5% agarose gel electrophoresis.

1.5. Exome sequencing of the TruSight® Cancer Sequencing Panel

5 µg of extracted DNA from 7 HDGC probands was sent to our collaborators at IRST-IRCCS (Istituto Scientifico Romagnolo per lo Studio e la Cura dei Tumori Srl Istituto di Ricovero e Cura a Carattere Scientifico) of Meldola (FC-Italy). There, Illumina exome sequencing of the TruSight® Cancer Sequencing Panel (Illumina®) was performed on the DNA of each patient. The applied panel uses around 4,000 probes that target 94 genes and 284 SNPs associated with predisposition towards both common and rare cancers. The probes target over 1,700 exons of those genes, representing more than 95% exonic coverage. The properties of the probes and the genes being targeted are depicted in Figures 3 and

4, respectively.

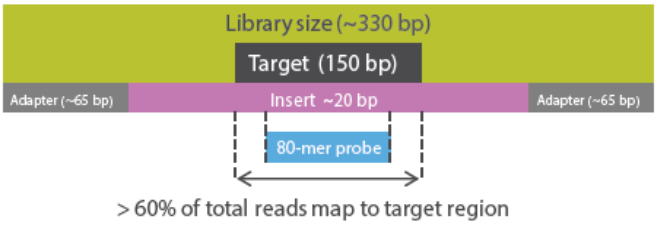

Figure 3. Probe footprint (from TruSight® Cancer Sequencing Panel data sheet, by Illumina).

AIP ALK APC ATM BAP1 BLM BMPR1A BRCA1 BRCA2 BRIP1 BUB1B CDC73 CDH1 CDK4 CDKN1C CDKN2A CEBPA CEP57 CHEK2 CYLD DDB2 DICER1 DIS3L2 EGFR EPCAM ERCC2 ERCC3 ERCC4 ERCC5 EXT1 EXT2 EZH2 FANCA FANCB FANCC FANCD2 FANCE FANCF FANCG FANCI FANCL FANCM FH FLCN GATA2 GPC3 HNF1A HRAS KIT MAX MEN1 MET MLH1 MSH2 MSH6 MUTYH NBN NF1 NF2 NSD1 PALB2 PHOX2B PMS1 PMS2 PRF1 PRKAR1A PTCH1 PTEN RAD51C RAD51D RB1 RECQL4 RET RHBDF2 RUNX1 SBDS SDHAF2 SDHB SDHC SDHD SLX4 SMAD4 SMARCB1 STK11 SUFU TMEM127 TP53 TSC1 TSC2 VHL WRN WT1 XPA XPC

Figure 4. List of genes targeted in the TruSight® Cancer Sequencing Panel (Illumina®).

The TruSight™ Rapid Capture protocol was applied for samples' preparation following manufacturer's instructions (Figure 5). Sequencing was subsequently performed using the MiSeq platform (Illumina®) with MiSeq Reagent Kit v2 configured 2x150 cycles.

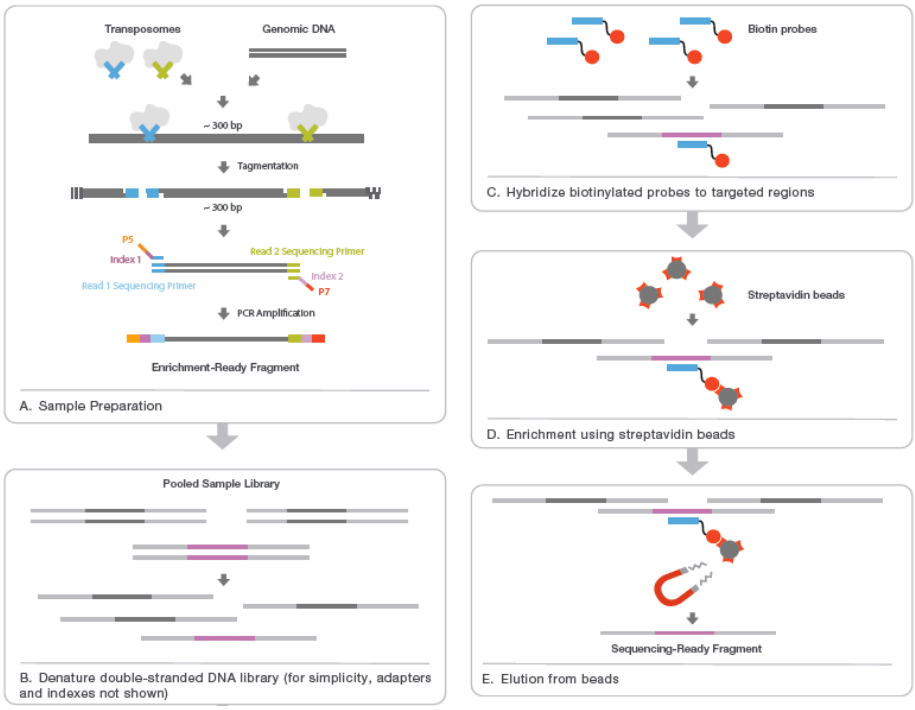

Figure 5. Schematic representation of the steps to obtain sequencing-ready DNA fragments from samples (from TruSight®Cancer Sequencing Panel data sheet, by Illumina).

1.5.1 Data analysis and variant calling

At the end of sequencing, the Burrows-Wheeler algorithm (Li H and Durbin R, 2009), running in paired-end mode, was used to align the raw de-multiplexed reads from Miseq sequencer to the reference human genome (UCSC-Build37/hg19). Base Quality Score Recalibration (BQSR) using the Genome Analysis Toolkit GATK, version 3.2.2 (McKenna A *et al*., 2010), was next applied on the samples to reduce the number of false positives and ensure a good call quality. Following BQSR, GATK was again utilized to locally realign sequences around regions with Indels (insertions, deletions). Duplicate read-pairs that have arisen as artifacts in the process of PCR amplification or sequencing were removed with MarkDuplicates (http://picard.sourceforge.net). SNVs (Single Nucleotide Variants) and Indels were then screened for with the Unified Genotyper of GATK and genomic and functional annotations of those variants were accomplished by ANNOVAR (Wang K *et al*., 2010).

Another tool of GATK that is DepthOfCoverage was utilized to perform coverage statistics, while BASH and R custom scripts were applied to acquire the list of low coverage (50X) regions per sample. Any regions below this threshold were dismissed. Finally, to eliminate strand bias and reduce false positive calls, variants were manually viewed on IGV (Integrative Genomics Viewer) (Robinson JT et *al*., 2011).

Polymorphic variants were then excluded by using the frequencies reported in the following databases: 1000 Genomes Project (http://www.1000genomes.org/), ESP6500 (Exome Sequencing Project) (http://evs.gs.washington.edu/EVS/) and ExAC v0.3 (Exome Aggregation Consortium) (http://exac.broadinstitute.org/).

1.5.2 *In silico* analysis of missense VUSs

Pathogenicity of the missense Variants of Uncertain Significance identified following exome sequencing was assessed by querying the following six different bioinformatics tools:

1. Poly-Phen2 v2.1.0 (http://genetics.bwh.harvard.edu/pph2/), a tool that combines data on human protein sequence-based and structural features to predict the potential impact of an amino acid substitution on the structure and function of a protein. In particular, the HumVar trained and tested model, which distinguishes between mutations with drastic effects from all the remaining human variation, was applied (Adzhubei IA *et al*., 2010).
2. SIFT v6.0.1 (http://sift.jcvi.org/), a sequence homology-based tool that distinguishes tolerant from intolerant amino acid substitutions, incurred by missense mutations, and predicts the phenotypic effect on the protein based on sequence alignments collected through PSI-BLAST (Ng PC and Henikoff S, 2001).
3. PROVEAN v1.1 (http://provean.jcvi.org/seq_submit.php), another sequence homology-based tool that predicts the potential impact of missense and Indel mutations on the biological function of a protein taking into account the conservation of the amino acid sequences. This conservation is based on sequence alignments derived from closely related sequences, collected through PSI-BLAST (Choi Y *et al*., 2012).
4. MUpro v1.0 (http://mupro.proteomics.ics.uci.edu/), a tool integrating two machine-learning methods to predict the impact of single-site amino acid mutations on protein stability (Cheng J *et al*., 2006).
5. SNAP2 (https://rostlab.org/services/snap2web/), a neural network based classifier that predicts the effect of single amino acid substitutions on protein function. The prediction is made by taking into account not only sequence-related information, such as evolutionary conservation, but also structural features, including secondary structure and solvent accessibility.

6. PhD-SNP v2.0.6 (http://snps.biofold.org/phd-snp/phd-snp.html), an SVM (support vector machine) based classifier, trained and tested on protein sequence and profile information (Capriotti E *et al*., 2006). It predicts the deleterious effects of a mutation based on conservation and sequence environment.

Combined, these tools are likely to give a more holistic view on the impact, if any, of a VUS on protein structure, stability and function.

Results

1.1. *CDH1* aberrant splicing analysis

In collaboration with IEO (Milano), blood was obtained from 30 Italian HDGC patients who were found to be *CDH1* mutation-negative by various molecular approaches (Table 1).

Patients	Diagnosis/age	Selection criteria [a]
1	DGC/41	i
2	DGC/42	iii
3	GC [b]/77	ii
4	GC [b]/40	ii
5	DGC/35	i
6	GC [b]/25	iv
7	GC [b]/52	ii
8	DGC/37	iii
9	DGC/39	iv
10	IGC [b]/60	ii
11	DGC/34	iii
12	DGC/44	iii
13	DGC/44	iii
14	DGC/42	iii
15	DGC/42	iii
16	IGC [b]/52	ii
17	DGC/36	iii
18	DGC/30	iii
19	MixGC/42	iii
20	DGC/79	ii
21	DGC/59	ii
22	DGC/37	iii
23	DGC/54	ii
24	DGC/42	iii
25	DGC+DBC/34+53	iii
26	DGC/39	iii
27	DGC/38	iii
28	DGC/30	iii
29	DGC/36	iii
30	DGC/41	iii

Table 1. Selected HDGC patients. [a] (i) family history with two or more first degree relatives with GC, with at least one documented DGC diagnosed before age 50; (ii) family history with three or more first-second degree relatives with GC, diagnosed at any age, with at least one documented case of DGC (iii) proband diagnosed with DGC at <45 years of age, regardless of family history; (iv) family history with two or more first-second degree relatives with LBC, with or without DGC. [b] In spite of cancer histotype, these six patients were recruited due to their strong family history of gastric or breast cancer. No samples were available from other family members. DGC: Diffuse Gastric Cancer; LBC: Lobular Breast Cancer; IGC: Intestinal Gastric Cancer; DBC: Ductal Breast Cancer; MixGC: Mixed diffuse/intestinal Gastric Cancer.

In order to assess whether aberrant splicing of the *CDH1* gene could alternatively be involved in the manifestation of DGC, the expression pattern of the *CDH1* canonical transcript was determined in the 30 patients. In addition, given the recent discovery of the non-canonical transcripts arising from within *CDH1* intron 2 (Pinheiro H *et al*., 2012), the non-canonical *CDH1a* and *CDH1j* transcripts were also evaluated in the examined cohort. The three transcripts differ only in the 5' end of the sequence, where exon 1 of *CDH1* is replaced by exon 1a of *CDH1a* and exon 1j of *CDH1j*, respectively. In addition to that, exon 1a forms the initial portion of exon 1j. As such, to distinctly amplify and compare the three transcripts the forward primers were designed in exon 1, at the exon1a-exon1j junction or the portion that is specific to exon 1j, respectively. Meanwhile, the reverse primers were designed in various regions of exon 3 (Figure 1).

Two-step RT-PCR was then carried out on RNA extracted from the blood of the 30 cases, as well as 5 healthy controls, for the amplification of the canonical *CDH1*, and non-canonical *CDH1a* and *CDH1j* transcripts. *GAPDH* was additionally amplified for the purpose of having an endogenous control.

The PCR amplified canonical and non-canonical *CDH1* transcripts in the cases were subsequently compared to those of controls (Figure 1).

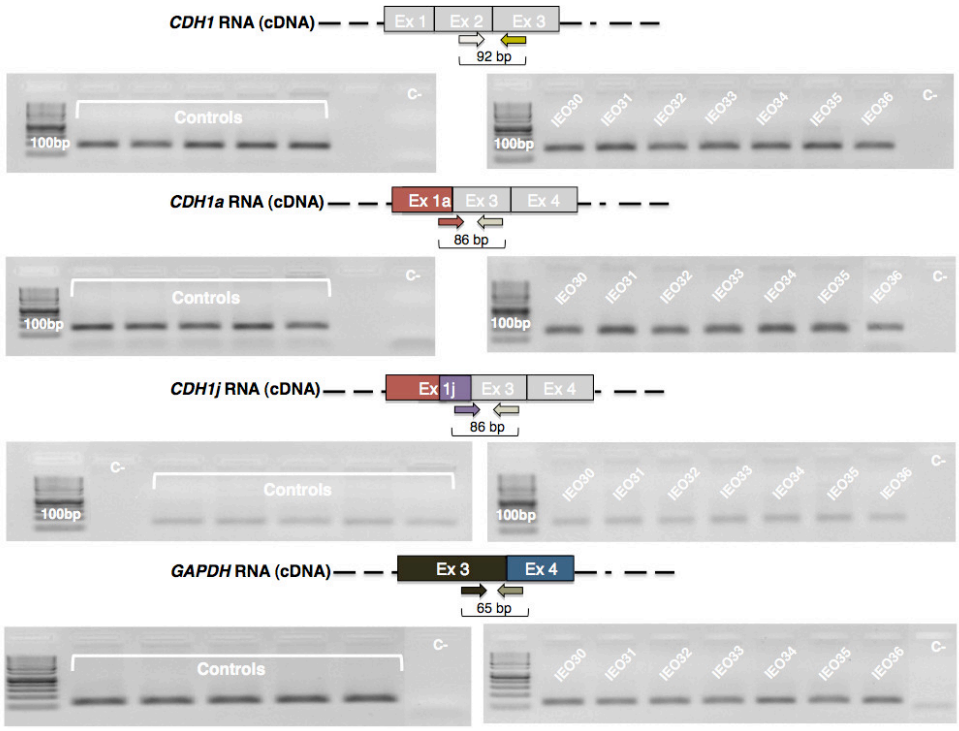

Figure 1. Expression patterns of *CDH-1*, *-1a*, *-1j* and *GAPDH* in the blood of 5 healthy controls and 7 HDGC patients. *CDH1* gene and transcripts with the primers used for their selective amplification. Primers used for the amplification of each transcript are depicted by single arrows; the sizes of resulting amplicons are indicated underneath each transcript. Ex: exon; c-: negative control.

This comparison revealed that the expression patterns of the three transcripts were the same in the patients and controls, which excluded the involvement of *CDH1* aberrant splicing in the investigated cohort.

1.2. Exome sequencing of the TruSight® Cancer Sequencing Panel

To test whether other cancer-related genes could explain the manifestation of gastric cancer in a fraction of cases, exome sequencing was performed on the 7 most recently recruited ones (1-7 in Table 1). In specific, DNA extracted from the blood of the 7 patients was subjected to Illumina exome sequencing of the TruSight® Cancer Sequencing Panel (Illumina®) by our collaborators at IRST-IRCCS, Meldola (FC). This panel targets 94 genes and 284 SNPs associated with predisposition towards both common and rare cancers. Following exome sequencing, bioinformatics analysis revealed the presence of 3,409 total variants, of which 965 were exonic. Polymorphic variants were then excluded by using the

frequencies reported in the 1000 Genomes Project, ESP6500 and ExAC databases. In this way, the list of 965 variants was narrowed down to 51 single nucleotide variants (SNVs), which have an undetermined frequency in the population or a frequency that is ≤0.01. Only insufficient or conflicting evidence exists on the functional significance of the 51 variants and hence they were all considered as Variants of Uncertain Significance (VUSs).

The 51 VUSs mapped to 35 of the 64 genes included in this panel. 19 VUSs (47%) were synonymous, while the majority, 32, were non-synonymous (63%), distributing across 17 and 27 genes, respectively. The most commonly affected genes were the tumor suppressor genes *SDHC* (2 synonymous and 2 non-synonymous VUSs), and *TSC2* (1 synonymous and 2 non-synonymous VUSs), as well as the proto-oncogene *RET* (2 synonymous and 1 non-synonymous VUSs). Meanwhile, the number of VUSs per patient ranged from a minimum of 4 (patient 7) to a maximum of 13 (patient 4). Details on the VUSs are reported in Table 2.

Patient	Chr	Gene	Accession number [a]	Exon	VUS	Classification
1	10	*RET*	NM_020975.4	12	c.2226G>A; p.Thr742Thr	Synonymous
2	3	*FANCD2*	NM_033084.4	15	c.1260A>C; p.Thr420Thr	Synonymous
2	14	*DICER1*	NM_001271282.2	21	c.3972G>A; p.Lys1324Lys	Synonymous
2	17	*BRIP1*	NM_032043.2	2	c.33T>C; p.Gly11Gly	Synonymous
2	18	*SMAD4*	NM_005359.5	10	c.1236C>T; p.Tyr412Tyr	Synonymous
3	4	*KIT*	NM_000222.2	21	c.2805T>A; p.Ile935Ile	Synonymous
3	10	*RET*	NM_020975.4	21	c.3972G>A; p.Lys1324lLys	Synonymous
3	15	*BUB1B*	NM_001211.5	5	c.573C>A; p.Ser191Ser	Synonymous
4	1	*SDHC*	NM_003001.3	5	c.273A>G; p.Leu91Leu	Synonymous
4	1	*SDHC*	NM_003001.3	5	c.306T>C; p.Leu102Leu	Synonymous
4	6	*FANCE*	NM_021922.2	10	c.1572G>A; p.Arg524Arg	Synonymous
4	15	*BLM*	NM_000057.3	10	c.2268A>G; p.Lys756Lys	Synonymous
4	16	*SLX4*	NM_032444.2	8	c.1824G>A; p.Glu608Glu	Synonymous
4	16	*CDH1*	NM_004360.4	3	c.345G>A; p.Thr115Thr	Synonymous
5	2	*ALK*	NM_004304.4	7	c.1464C>T; p.Gly488Gly	Synonymous

5	2	*PMS1*	NM_001289408.1	9	c.1755T>C; p.Leu585Leu	Synonymous
6	13	*BRCA2*	NM_000059.3	11	c.2817C>T; p.Thr939Thr	Synonymous
7	11	*WT1*	NM_024426.4	1	c.108G>C; p.Pro36Pro	Synonymous
7	16	*TSC2*	NM_000548.4	40	c.5130C>T; p.Phe1710Phe	Synonymous
1	9	*CDKN2A*	NM_000077.4	2	c.221A>C; p.Asp74Ala	Non-synonymous
1	11	*AIP*	NM_003977.3	6	c.911G>A; p.Arg304Gln	Non-synonymous
1	16	*TSC2*	NM_000548.4	26	c.2921C>T; p.Ala974Val	Non-synonymous
1	17	*RAD51C*	NM_058216.2	6	c.859A>G; p.Thr287Ala	Non-synonymous
2	2	*EPCAM*	NM_002354.2	6	c.605A>C; p.Lys202Thr	Non-synonymous
2	4	*KIT*	NM_000222.2	21	c.2867G>A; p.Arg956Gln	Non-synonymous
2	5	*APC*	NM_000038.5	16	c.7862C>G; p.Ser2621Cys	Non-synonymous
2	6	*FANCE*	NM_021922.2	8	c.1333C>T; p.Pro445Ser	Non-synonymous
2	16	*CDH1*	NM_004360.4	12	c.1774G>A; p.Ala592Thr	Non-synonymous
3	2	*ERCC3*	NM_000122.1	14	c.2111C>T; p.Ser704Leu	Non-synonymous
3	9	*TSC1*	NM_000368.4	17	c.2194C>T; p.His732Tyr	Non-synonymous
3	11	*AIP*	NM_003977.3	1	c.47G>A; p.Arg16His	Non-synonymous
3	11	*ATM*	NM_000051.3	11	c.1744T>C; p.Phe582Leu	Non-synonymous
4	1	*SDHC*	NM_003001.3	5	c.295T>A; p.Tyr99Asn	Non-synonymous
4	1	*SDHC*	NM_003001.3	5	c.307G>C; p.Val103Leu	Non-synonymous
4	10	*RET*	NM_020975.4	13	c.2371T>A; p.Tyr791Asn	Non-synonymous
4	13	*ERCC5*	NM_000123.3	14	c.2890C>T; p.Arg964Trp	Non-synonymous
4	16	*TSC2*	NM_000548.4	4	c.251C>T; p.Ala84Val	Non-synonymous
4	16	*SLX4*	NM_032444.2	14	c.4871C>T; p.Pro1624Leu	Non-synonymous
4	X	*FANCB*	NM_001018113.2	7	c.1349G>A; p.Cys450Tyr	Non-synonymous
5	8	*NBN*	NM_002485.4	5	c.506G>A; p.Arg169His	Non-synonymous
5	9	*PTCH1*	NM_001083603.1	1	c.157G>C; p.Glu53Gln	Non-synonymous
5	11	*SDHD*	NM_003002.3	1	c.34G>A; p.Gly12Ser	Non-synonymous

5	14	*FANCM*	NM_020937.3	2	c.527C>T; p.Thr176Ile	Non-synonymous
5	17	*NF1*	NM_001042492.2	26	c.3436G>A; p.Val1146Ile	Non-synonymous
6	3	*FANCD2*	NM_033084.4	3	c.195G>C; p.Gln65His	Non-synonymous
6	3	*XPC*	NM_004628.4	4	c.506C>T; p.Thr169Met	Non-synonymous
6	7	*EGFR*	NM_005228.3	22	c.2664C>G; p.His888Gln	Non-synonymous
6	11	*SDHD*	NM_003002.3	2	c.149A>G; p.His50Arg	Non-synonymous
6	13	*BRCA2*	NM_000059.3	26	c.9586A>G; p.Lys3196Glu	Non-synonymous
7	11	*EXT2*	NM_000401.3	4	c.809C>T; p.Ser270Leu	Non-synonymous
7	17	*NF1*	NM_001042492.2	34	c.4526G>A; p.Arg1509His	Non-synonymous

Table 2. Genomic localization of the VUSs generated in each patient and their classification. [a] Accession number is based on the NCBI-gene database (http://www.ncbi.nlm.nih.gov/gene). Chr: chromosome; VUS: Variant of Uncertain Significance.

While synonymous variants might affect the splicing process by introducing single base changes at the genomic level, they give rise to the same amino acid in the protein. Meanwhile, non-synonymous, or missense variants result in single amino acid substitutions that may directly alter a protein's function by affecting its stability, folding and structure. Note that the VUSs reported in the *CDH1* gene were both originally uncovered by direct sequencing and dismissed from being causative. Indeed, the synonymous variant (c.345G>A; p.Thr115Thr) in exon 3 did not affect the *CDH1* splicing process, while the non-synonymous variant (c.345G>A; p.Thr115Thr) has been previously reported as having no association with both HDGC and familial breast cancer (Salahshor S *et al*., 2001; Keller G *et al*., 2004; Schrader KA *et al*., 2011).

In order to determine the potential impact of all the reported missense variants on the proteins' phenotypes, an *in silico* analysis was performed using six different bioinformatics tools. These tools included sequence homology-based tools (SIFT and PROVEAN), sequence and function based predictive tools (PolyPhen-2, SNAP2), as well as machine-learning predictive tools based on sequence environment (PhD-SNP) and stability (MUpro). The results of the 32 VUSs are depicted in Table 3.

VUS	SIFT	PROVEAN	PolyPhen2	SNAP2	PhD-SNP	MUpro
AIP c.911G>A; p.Arg304Gln	Damaging	Neutral	Neutral	Damaging	Neutral	Damaging
AIP c.47G>A; p.Arg16His	Damaging	Damaging	Neutral	Damaging	Neutral	Neutral
APC c.7862C>G; p.Ser2621Cys	Neutral	Neutral	Neutral	Neutral	Neutral	Damaging
ATM c.1744T>C; p.Phe582Leu	Neutral	Neutral	Neutral	Neutral	Neutral	Damaging
BRCA2 c.9586A>G; p.Lys3196Glu	Neutral	Neutral	Neutral	Damaging	Neutral	Damaging
CDH1 c.1774G>A; p.Ala592Thr	Damaging	Damaging	Neutral	Damaging	Damaging	Damaging
***CDKN2A* c.221A>C; p.Asp74Ala**	Damaging	Damaging	Damaging	Damaging	Damaging	Damaging
EGFR c.2664C>G; p.His888Gln	Neutral	Damaging	Neutral	Neutral	Neutral	Damaging
EPCAM c.605A>C; p.Lys202Thr	Damaging	Damaging	Damaging	Damaging	Neutral	Damaging
ERCC3 c.2111C>T; p.Ser704Leu	Neutral	Damaging	Neutral	Damaging	Neutral	Neutral
ERCC5 c.2890C>T; p.Arg964Trp	Damaging	Damaging	Damaging	Damaging	Neutral	Damaging
EXT2 c.809C>T; p.Ser270Leu	Neutral	Neutral	Neutral	Neutral	Neutral	Neutral
FANCB c.1349G>A; p.Cys450Tyr	Neutral	Damaging	Damaging	Neutral	Damaging	Neutral
FANCD2 c.195G>C; p.Gln65His	Damaging	Damaging	Damaging	Damaging	Neutral	Neutral

FANCE c.1333C>T; p.Pro445Ser	Neutral	Neutral	Neutral	Neutral	Neutral	Damaging
FANCM c.527C>T; p.Thr176Ile	Neutral	Damaging	Neutral	Neutral	Neutral	Neutral
KIT c.2867G>A; p.Arg956Gln	Neutral	Neutral	Damaging	Neutral	Neutral	Damaging
NBN c.506G>A; p.Arg169His	Damaging	Damaging	Neutral	Neutral	Neutral	Damaging
NF1 c.3436G>A; p.Val1146Ile	Neutral	Neutral	Damaging	Neutral	Neutral	Damaging
NF1 c.4526G>A; p.Arg1509His	Damaging	Damaging	Damaging	Damaging	Damaging	Neutral
PTCH1 c.157G>C; p.Glu53Gln	Damaging	Neutral	Neutral	Neutral	Neutral	Damaging
RAD51C c.859A>G; p.Thr287Ala	Damaging	Damaging	Damaging	Damaging	Neutral	Damaging
RET **c.2371T>A; p.Tyr791Asn**	Damaging	Damaging	Damaging	Damaging	Damaging	Damaging
SDHC **c.295T>A; p.Tyr99Asn**	Damaging	Damaging	Damaging	Damaging	Damaging	Damaging
SDHC c.307G>C; p.Val103Leu	Damaging	Neutral	Neutral	Neutral	Neutral	Damaging
SDHD c.34G>A; p.Gly12Ser	Neutral	Neutral	Neutral	Damaging	Neutral	Neutral
SDHD c.149A>G; p.His50Arg	Neutral	Neutral	Damaging	Damaging	Neutral	Neutral
SLX4 c.4871C>T; p.Pro1624Leu	Neutral	Damaging	Damaging	Damaging	Neutral	Damaging
TSC1 c.2194C>T; p.His732Tyr	Neutral	Damaging	Damaging	Damaging	Neutral	Neutral
TSC2 c.2921C>T; p.Ala974Val	Damaging	Neutral	Damaging	Damaging	Neutral	Neutral

TSC2 c.251C>T; p.Ala84Val	Neutral	Neutral	Damaging	Neutral	Neutral	Neutral
XPC c.506C>T; p.Thr169Met	Neutral	Neutral	Damaging	Neutral	Neutral	Neutral

Table 3. *In silico* analysis of the 32 missense VUSs. In bold are the three variants found to be damaging by all six tools. VUS: Variant of Uncertain Significance

Based on this analysis, 3 variants (≈ 9%) were consistently found to be damaging, and hence pathogenic, by all utilized tools. These variants could be traced back to the *CDKN2A* gene in patient 1, as well as the *RET* and *SDHC* genes in patient 4, respectively. In addition, a common consensus could be reached on the neutrality of 1 missense variant present in the *EXT2* gene in patient 7. On the other hand, inconsistent results were reported by the six tools for 28 variants (≈ 88%), thus preventing an adequate conclusion to be drawn on their true impact on protein function. The VUS in *CDH1* was among the latter discordantly classified variants.
Taken together, these results show that underlying genetic variants might be associated with the manifestation of gastric cancer in 2 out of the 7 investigated patients.

Discussion

HDGC is an autosomal-dominant cancer predisposition syndrome that is associated with a high lifetime risk for diffuse-type gastric cancer, coupled with an increased risk of lobular breast cancer in women (reviewed by Monahan KJ and Hopkins L, 2016). Germline alterations in the *CDH1* tumor suppressor gene account for around 10 to 50% of HDGC cases (reviewed by Fitzgerald RC *et al*., 2010), with over 180 mutations identified in affected families thus far. Despite this number and despite the efforts made to integrate a multi-method approach for the detection of *CDH1* constitutive lesions such as large gene deletions and allelic imbalance, a large fraction of cases remains without a known genetic determinant (reviewed by Fitzgerald RC *et al*., 2010). This large fraction might be explained by mechanisms of *CDH1* inactivation that are yet to be accounted for; alternatively, the missing genetic heritability causative of HDGC in *CDH1* mutation-negative cases could be distributed along multiple loci. In this work we aimed to molecularly characterize a cohort of Italian HDGC patients by performing *CDH1* transcripts' analysis on 30 selected cases already proven to be *CDH1*-mutation negative by comprehensive molecular approaches. In addition, we sought to carry out exome sequencing of a panel of cancer-related genes in 7 of those cases, as part of a larger ongoing study.

Increasing evidence has implicated *CDH1* intron 2 as an important *cis*-modulator of E-cadherin gene and protein expression (Stemmler MP *et al*., 2005; reviewed by Paredes J *et al*., 2012). Moreover, it has recently been shown that four novel transcripts arise from within this intron, namely *CDH1a*, *CDH1b*, *CDH1j* and *CDH1b-10* (Pinheiro H *et al*., 2012). The *CDH1* gene, which maps to the long arm of chromosome 16 and comprises 16 exons, is transcribed to a 4.8 kb pre-mRNA. This pre-mRNA is processed by splicing to generate a mature, intron-free, mRNA that can then be translated into the E-cadherin protein. The splicing process itself is regulated by both *cis*- and *trans*- acting elements. Abnormal alterations in these elements could lead to aberrantly spliced transcripts that ultimately intervene with the production of mature and functional protein (reviewed by Liu X and Chu KM, 2014). Indeed, there have been reports on germline VUSs in HDGC affecting the splicing process of *CDH1* by creating preferential splice sites or activating cryptic ones (Oliveira C *et al*., 2004; Kaurah P *et al*., 2007; Li X *et al*., 2013; Molinaro V *et al*., 2014). Such defects are typically highlighted by performing a straightforward RNA analysis on suspected patients. In support of this notion, we set out to investigate the expression patterns of the *CDH1* canonical and *CDH-1a* and *-1j* non-canonical transcripts in the blood of 30 *CDH1* mutation-negative HDGC patients.

This is the first evaluation of the novel non-canonical transcripts in the blood of gastric cancer patients.
Upon amplifying the 3 transcripts by RT-PCR, we compared the resulting bands with those present in the blood of controls. No difference could be detected in any of the investigated cases, which excluded the possible involvement of *CDH1* aberrant splicing in these patients. Moreover, given that for the patient selection process we only sequenced the 16 canonical exons and known exon-intron boundaries, the absence of an abnormal expression of the non-canonical transcripts is a good indicator that we have not missed any potentially causative intronic variants in that region. In addition, since the aforementioned direct sequencing of the *CDH1* gene only uncovered harmless variants, we have no reason to believe that defective splicing incurred by exonic or intronic variants beyond exon 3 is contributory to disease manifestation in the investigated cohort.

While *CDH1* is the most consistently altered gene in HDGC, the application of exome sequencing and gene panel testing on *CDH1* mutation-negative cases have unraveled germline lesions in diverse unexpected genes that could predispose to the disease (Majewski IJ *et al*., 2013; Gaston D *et al*., 2014; Donner I *et al*., 2015; Hansford S *et al*., 2015). Such efforts are important as they would aid in the identification of at risk individuals in affected families who are negative for *CDH1* mutations, thereby improving their counseling and management. In keeping with this trend, we applied exome sequencing of a multi-gene panel (TruSight® Cancer Sequencing Panel) on a first group of 7 out of the 30 *CDH1* mutation-negative cases. The gene panel considered for this study comprises a compilation of 94 cancer-associated genes known to be involved in hereditary cancer syndromes. Among the genes are *ATM*, *BRCA2*, *PALB2*, *SDHB* and *STK11*, in which mutations have very recently been associated with HDGC (Hansford S *et al*., 2015).
At the end of exome sequencing and following the bioinformatics' pipeline for variant filtering, a total of 51 VUSs that have never been reported, or have been found to have sub-polymorphic frequencies in the population, were uncovered. Among those variants, 32 were missense, resulting in a single amino acid change that could be deleterious to the protein's proper folding, structure or function. Based on an *in silico*-driven analysis, 3 of these VUSs were unanimously classified as damaging (*CDKN2A* c.221A>C, p.Asp74Ala; *RET* c.2371T>A, p.Tyr791Asn; *SDHC* c.295T>A, p.Tyr99Asn), while 1 (*EXT2* c.809C>T, p.Ser270Leu) as benign, by the six utilized bioinformatics tools. On the other hand, the majority of the VUSs (88%) were discordantly classified. This is relatively unexpected since each of the tools used has its own algorithms and protocols for predicting the functional significance of a variant. For that very same reason, the predicted

pathogenicity of the 3 missense variants in the *CDKN2A*, *RET* and *SDHC* genes might even be more credible.

With respect to the missense variant in *CDKN2A*, it was detected in proband 1 who had been diagnosed with DGC at 41 years of age.
CDKN2A gene encodes two distinct onco-suppressors, p16INK4a and p14ARF, by alternative splicing of the first two exons. The two proteins play distinct, but key roles, in the regulation of the cell cycle and apoptosis (reviewed by Tsao H *et al*., 2012; Scaini MC *et al*., 2014). Germline mutations, the majority of which are missense and affecting p16INK4a, are associated with approximately 40% of melanoma-prone families (Scaini MC *et al*., 2014). Indeed, mutations in *CDKN2A* underlie cancer susceptibility to melanoma and pancreatic cancer in the familial atypical mole multiple melanoma (FAMMM)/melanoma-pancreatic cancer (melanoma-PC) syndrome (Wang W *et al*., 2007; Ghiorzo P *et al*., 2007; Scaini MC *et al*., 2014; reviewed by Petersen GM, 2015). Interestingly, there have been previous reports implicating *CDKN2A* germline mutations in breast cancer. In particular, it has been shown that families carrying a germline mutation that constitutes an in-frame 3-bp duplication at nucleotide 332 in exon 2 of *CDKN2A*, have an increased risk, not only of melanoma and pancreatic carcinoma, but also of breast cancer (Borg A *et al*., 2000). Indeed, among the melanoma-positive carriers of this mutation, one also developed breast cancer at age 61, while among the melanoma-free carriers two developed breast cancer at ages 23 and 62, respectively, without the manifestation of any other malignancy (Borg A *et al*., 2000). In addition, there was another mutation-carrier having a breast cancer diagnosis at 50 years of age, coupled with gastric cancer at 80 years of age (Borg A *et al*., 2000). Even though this mutation seems to represent a founder mutation in the Scandinavian countries, it being reported in melanoma-free breast cancer patients provides a proof-of-concept for the potential implication of *CDKN2A* in breast cancer. In fact, a recently massive parallel sequencing of 22 cancer susceptibility genes in 278 *BRCA1/2* negative patients with early-onset breast cancer, reported a missense mutation in *CDKN2A* that is known to be pathogenic. This mutation was found in a melanoma-free proband with a personal history of early-onset breast cancer and sarcoma diagnosed at 38 and 44 years of age, respectively (Maxwell KN *et al.,* 2015).
Moreover, while there has been no previous evidence of this gene in predisposition to HDGC, somatic *CDKN2A* gene alterations have been found to be frequent in GCs of both the intestinal and diffuse histotypes (Guimarães AC *et al*., 2007; Ali SM *et al*., 2015; Huang S *et al*., 2015), as well as in the EBV-associated GC subtype (Cancer Genome Atlas Research Network, 2014). In particular, one study that performed a prospective comprehensive genomic profiling of 116 gastric

cancer specimens, of which 20% were of the diffuse histotype, revealed a *CDKN2A* mutation rate of 14% (Ali SM *et al*., 2015). In addition, earlier studies on GC cell lines have shown that diffuse-type cell lines have a higher frequency of *CDKN2A* gene alterations (homozygous deletion and promoter hypermethylation), affecting p14ARF, than intestinal-type cell lines (Iida S *et al*., 2000).
Therefore, our finding of a non-synonymous, predicted-pathogenic mutation in the *CDKN2A* gene, coupled with the evidence of its implication in both gastric and breast cancer, strengthens its involvement as a novel gene predisposing to HDGC.

Regarding the missense VUSs reported in the *RET* and *SDHC* genes, they were both uncovered in case 4, who harbored the highest rate of both synonymous and non-synonymous variants among the 7 analyzed cases. This individual had been diagnosed with GC at 40 years of age, and was included in the study on the basis of a family history suggestive of HDGC. Based on the early age of diagnosis, it is likely that one or both of the variants have a contributory effect on the manifestation of the disease in this patient. Indeed, both *RET* and *SDHC* were two of the most commonly mutated genes in our cohort.
RET is a proto-oncogene alternatively spliced to yield three functional tyrosine kinase receptors involved in cell proliferation, differentiation, motility and survival (reviewed by Kouvaraki MA *et al*., 2005; reviewed by Elisei *et al*., 2014). Germline activating mutations in this gene account for 30-35% of hereditary medullary thyroid carcinoma (MTC) (Rohmer V *et al*., 2011). With respect to investigations of the gene in gastric cancer, the only report that comes close concerns the presence of a somatic missense variant in *RET* in one case of intestinal-type antral early gastric cancer (Fassan M *et al*., 2014). In addition, an unlikely causative homozygous germline variant in this gene has been reported in a female diagnosed with gastrointestinal stromal tumors (GIST) coupled with multiple pathologies, but without the expected medullary thyroid carcinoma (Bano G *et al*., 2013). To the best of our knowledge, there have been no other reports of either constitutive or somatic mutations in this gene in gastric cancer patients.
On the other hand, *SDHC* is a tumor suppressor gene encoded in the autosome, along with three other succinate dehydrogenases subunits (*SDHA*, *SDHB* and *SDHD*). The proteins encoded by these genes are assembled in the mitochondria to form the mitochondrial complex 2 that links the Krebs cycle and the electron transport chain, hence playing a key role in curbing oxidative stress (reviewed by Gill AJ, 2012). Although this complex is essential for life, haploinsufficiency is tolerated and compensated; meanwhile a two hit inactivation of any subunit of the complex renders it inactive (reviewed by Gill AJ, 2012; reviewed by Miettinen M and Lasota J, 2014). Germline mutations in the subunit genes underlie around 15% of familial phaeochromocytoma and paraganglioma (PHEO/PGL) cases (reviewed

by Gill AJ, 2012). In addition, germline mutations have been associated with Carney-Stratakis syndrome/PGL syndromes, predisposing to paraganglioma and GIST (reviewed by Gill AJ, 2012; reviewed by Miettinen M and Lasota J, 2014). SDH-deficient GISTs are a unique group of gastrointestinal stromal tumors in which an energy metabolism defect is the key oncogenic mechanism (reviewed by Miettinen M and Lasota J, 2014). This group of tumors most commonly manifests in children and young adults and is associated with the occurrence of other carcinomas, including not only paraganglioma, but also renal cell carcinomas that are normally asynchronous. Of relevance, a germline mutation of *SDHB* gene has recently been associated with LBC and GC in an Italian family (Hansford S *et al*., 2015). The same *SDHB* mutation had previously been reported to be associated with the cancer risk disorder Cowden-like syndrome (Ni Y *et al*., 2008).
Our exome sequencing results highlight the impressive aspect of applying next generation sequencing on patients selected for a specific disease. In particular, the emergence of unsuspected candidate genes that could somehow explain the disease in question opens the door to its better characterization. This process could ultimately lead to new therapeutic opportunities by providing novel targets. As an example, *CDKN2A* encodes p16INK4a that functions as an inhibitor of cyclin-dependent kinase (CDK4/6) (Scaini MC *et al*., 2014). Inactivating mutations in this gene could thereby result in the disruption of cell cycle control. Of relevance, GCs with somatic mutations in *CDKN2A* have recently been shown to be sensitive to a CDK4/6 inhibitor, PD-0332991 (Huang S *et al*., 2015).
At this point it should be emphasized that *in silico* predictions only give an approximation of pathogenicity and while they are unable to replace experimental validation, they do provide a fast screening facet that could better guide experimental efforts (Leonardi E *et al*., 2011; Scaini MC *et al*., 2014). Nonetheless, for a variant to be irrevocably defined as pathogenic, proof needs to be provided of its segregation with the disease, and its absence in normal controls; but even segregation analysis can be complicated by incomplete penetrance.

In our case, by combining results from the *in silico* analysis with supportive literature findings regarding the affected genes, we can put forth two novel candidates for HDGC, namely *CDKN2A* and *SDHC*. The implication of each gene will be further ascertained by first of all testing family members to determine whether the identified mutations co-segregate with the disease. Second, the tumor tissue of the affected individuals will be tested, when possible, in search of the second hits that would render each gene inactive. Moreover, as this work is part of a much larger study, more cases will be enrolled for the in-depth investigation of the molecular underpinning of HDGC; this might lead to the identification of additional families carrying either *CDKN2A* or *SDHC* mutations, thus confirming

these genes as being associated with gastric/breast cancer predisposition.

On the whole, we can safely dismiss the involvement of *CDH1* alterations, including those pertaining to aberrant splicing, in the 30 *CDH1* mutation-negative HDGC cases analyzed. In addition, the deeper investigation of 7 cases with exome sequencing indicate *CDKN2A* and *SDHC* as novel candidate genes predisposing to HDGC.

Second Aim of the Research

2. Evaluation of *CDH1* canonical/non-canonical transcripts and of *CDH1*-related miRNAs in intestinal-type gastric cancer

In the second part of my research, the aim was to evaluate, for the first time, the differential expression of both *CDH1* canonical and *CDH1a* non-canonical transcripts in tumor and normal tissue samples derived from 32 patients with intestinal-type gastric cancer (IGC). To quantify the transcripts, I set up the optimized conditions in digital-PCR (dPCR), a sensitive method for detection of gene expression in settings where the target RNA is limited or present in quantity that approaches the limits of quantitative-PCR sensitivity. In addition, I attempted to characterize the potential mechanisms by which *CDH1a* transcript is exerting its previously proposed tumorigenic effect by quantifying the expression of interferon-inducible genes in the tumor tissue samples. Moreover, to provide a more holistic picture of *CDH1* deregulation in IGC, I also performed an *in silico* analysis for the selection of suitable *CDH1*-regulating miRNAs and evaluated their expression in paired tissue samples.

Materials and Methods

2.1. Collection of fresh-frozen tissue samples

The retrospective study for the quantification of *CDH1* transcripts in IGC patients was conducted on 53 fresh-frozen specimens including 21 paired normal/cancer tissue samples and 11 additional tumor tissues. Samples were obtained from 32 patients diagnosed with gastric cancer of the intestinal histotype (Laurén's classification), recruited between 2007 and 2012 at the IRST-IRCCS (Istituto Scientifico Romagnolo per lo Studio e la Cura dei Tumori Srl Istituto di Ricovero e Cura a Carattere Scientifico) of Meldola (FC-Italy). Samples were macrodissected from blocks of tumor (containing at least 70% tumor cells) and normal tissue that had been cryopreserved immediately after surgical resection. The presence of *H. pylori* was further assessed by the examination of hematoxylin and eosin stained formalin-fixed paraffin-embedded (FFPE) tissue sections of each patient. This study was approved by the Local Ethics Committee (Comitato Etico Area Vasta Romagna e IRST) and informed consent was obtained from all patients (protocol number: IRSTB062).

2.2. RNA and cDNA preparations from fresh-frozen tissue samples

Total RNA was extracted from the 53 fresh-frozen tissue sections derived from the 32 IGC patients. This was done by the TRIzol®Reagent (Invitrogen), and was followed by a treatment with DNase (Qiagen), and purification with the RNeasy MinElute Cleanup Kit (Qiagen), all according to the standard protocols provided by the manufacturers. Purified RNA was eluted with RNase free water (Qiagen) and concentration and quality were assessed by the Spectrophotometer Nanodrop-ND-1000 (Celbio). RNA integrity was verified by using the Experion™ RNA StdSens Analysis Kit (Bio-Rad Laboratories) following manufacturer's instructions.

1 µg of RNA was reverse transcribed in a final volume of 20 µl using the 5X iScript buffer, containing a blend of oligo(dT) and random hexamer primers, as well as the iScript reverse transcriptase (Bio-Rad Laboratories). All reactions were carried out according to manufacturer's instructions and thermal cycling conditions were as follows: 25°C for 5 min, 42°C for 30 min, 85°C for 5 min. Reaction mixtures lacking the reverse transcriptase but containing RNA and no-template controls (NTC) using water instead of RNA were also included to monitor DNA contamination. In all cases no contamination of DNA was detected.

2.3. Setting up digital-PCR (dPCR) quantification of *CDH1* transcripts

All dPCR experiments were set up and carried out using the chip-based QuantStudio™ 3D Digital PCR system (Applied Biosystems) in accordance with

the "Minimum Information for Publication of Quantitative Real-Time dPCR Experiments" (dMIQE) guidelines (Huggett *et al*., 2013; see Appendix A). Chip based digital-PCR is an end-point reaction that relies on the calibrated partitioning of a sample into thousands of wells, such that each well contains zero or a single target molecule. Amplification then occurs only in the wells containing a copy of the target and is indicated by a florescent signal. The absolute number of target molecules in the original sample can then be calculated by determining the ratio of positive to total partitions using binomial Poisson statistics. The basic steps in dPCR can be seen in Figure 1.

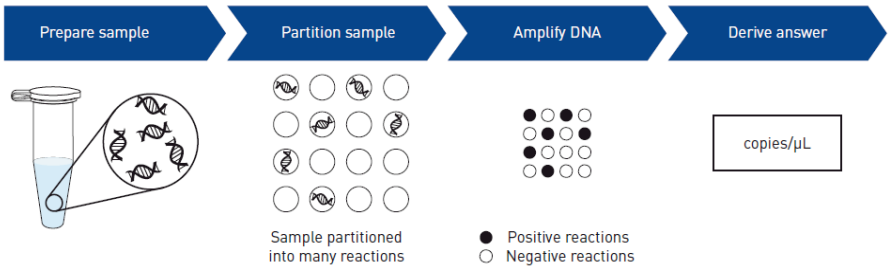

Figure 1. Basic steps in digital-PCR (from QuantStudio™ 3D Digital PCR System User Guide, by Applied Biosystems).

In this way, dPCR does not require the use of standards or internal controls; however, in the case of gene expression analysis, variability can be introduced in a number of steps required to attain cDNA from the original RNA. Therefore, for the quantification of the *CDH1* transcripts in the tumor and normal tissue samples of IGC patients, *GAPDH* was included as an endogenous expression control.

2.3.1. Design of gene-specific assays for dPCR

In order to selectively amplify the *CDH1* and *CDH1a* transcripts, assays were custom designed by using PrimerQuest tool found on the Integrated DNA Technologies (IDT) website (http://www.idtdna.com/primerquest/home/index). In particular, the *CDH1a* forward primer was designed at the exon 1a-exon 3 junction, and the probe and reverse primer in exon 3. With respect to *CDH1*, the forward primer was designed across the exon 2-3 junction while the probe and reverse primer were designed in exon 3 (Figure 2).

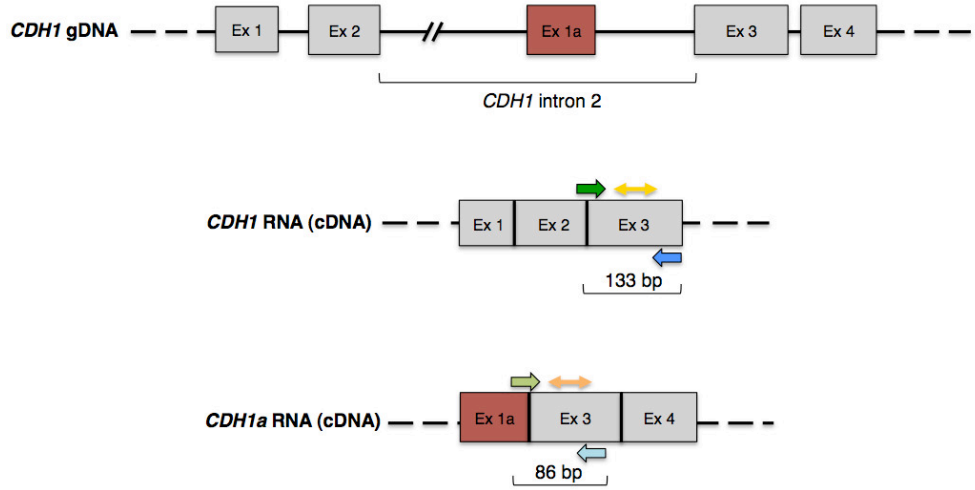

Figure 2. *CDH1* gene and related transcripts. In grey are canonical exons; in red is the *CDH1a* non-canonical exon; primers and probes designed to detect the specific transcripts are depicted by single and double arrows, respectively; the sizes of resulting amplicons are indicated underneath each transcript. Ex: exon.

In each case, the probe is only excised when the polymerase is actively amplifying the sequence; this requires the binding of the very specific forward primer. Therefore, a florescent signal is achieved only in the case of amplifying the target. This signal is blue in color and is perceived due to the presence of a FAM (Fluorescein) dye at the 5' end of the probes that is ordinarily quenched by a TAMRA (Carboxytetramethylrhodamine) dye present at the 3' end. The specificity of all designed primers and probes was confirmed with BLAST (Basic Local Alignment Search Tool) (http://blast.ncbi.nlm.nih.gov/Blast.cgi); they were subsequently ordered from Metabion. Details regarding the *CDH1* and *CDH1a* assays utilized in dPCR are presented in Table 1.

Assay	Target	Accession number [a]	Primers/probe sequences and modifications (5' to 3') [b]	Location [c]	Amplicon size (bp)
CDH1 IDT custom designed assay	*CDH1*	NM_004360	F: GTCCTGGGCAGAGTGAATTT	Exon 2-3	133
			R: GTGGGTTATGAAACCGTAGAGG	Exon 3	
			P:/FAM/TCAAAGTGGGCACAGATGGTGTGA/TAMRA/	Exon 3	
CDH1a IDT custom designed assay	*CDH1a*	N/A	F: GCTGCAGTTTCACTTTTAGTG	Intron 2-Exon 3	86
			R: ACTTTGAATCGGGTGTCGAG	Exon 3	
			P:/FAM/CGGTCGACAAAGGACAGCCTATT/TAMRA/	Exon 3	

Table 1. Gene-specific custom designed assays for the amplification of *CDH1* and *CDH1a* by dPCR. [a] Accession number is based on the NCBI gene (http://www.ncbi.nlm.nih.gov/gene). *CDH1a* is a novel non-canonical transcript without an accession number; [b] dPCR optimized assay concentrations: 900 nM (F), 900 nM (R), 250 nM (P). [c] location provided for *CDH1a* is based on the accession number of the canonical transcript (*CDH1*). IDT: Integrated DNA Technologies; N/A: Not available; F: forward primer; R: reverse primer; P: probe.

With respect to the amplification of *GAPDH*, a predesigned PrimeTime® Std qPCR Assay manufactured by IDT was ordered and utilized. Unlike the probes targeting the *CDH1* transcripts, the probe of the *GAPDH* assay emits a red signal upon cleavage. Note that the *GAPDH* Hs.PT.39a.22214836 PrimeTime® Std qPCR Assay is based on the accession number NM_002046; details are not depicted since they are the manufacturer's propriety.

2.3.2. Determination of the starting amount of cDNA for dPCR

The amount of cDNA used in dPCR is important since the most accurate copy number calculation can be achieved when a single copy of the target is present in each well. As such, an estimation of the amount of cDNA added in the experiment must first be made so as not to risk having multiple copies of the target in too many wells, which would give rise to spurious results.

With respect to *CDH1a*, the transcript proved to be rare and thus the maximum amount of cDNA was used to ensure amplification, which in our case was 300 ng. On the other hand, regarding *CDH1* and *GAPDH*, *GAPDH* was used to estimate the amount of cDNA that can safely be utilized in dPCR, since this transcript was expressed at slightly higher levels than *CDH1* in the investigated gastric tissue.

The estimation was done by first performing two-step reverse transcription quantitative PCR (RT-qPCR) experiments in which 10 ng of previously reverse transcribed cDNA from each paired normal and tumor tissue samples were loaded in triplicate in a 96-well plate, coupled with the *GAPDH* PrimeTime® Std qPCR assay (IDT) and the TaqMan® Universal Master Mix II (Applied Biosystems) (Table 2).

Reagent	Final concentration	Volume (µl)
TaqMan® Universal Master Mix II (2X)	1X	10
GAPDH PrimeTime® Std qPCR assay (20X)	1X	1
≈ 2.5 ng/µl cDNA	≈ 10 ng	4
RNase-free water	-	to 20

Table 2. Concentrations and volumes of the reagents used for two-step RT-qPCR amplification of *GAPDH* transcript.

Reactions were subsequently run on a 7500 Real-Time PCR (Applied Biosystems) by applying the thermal cycling conditions indicated in Table 3.

Cycle steps	Temperature	Time
UNG incubation	50°C	2 min
Polymerase activation	95°C	10 min
Denaturation	95°C	15 sec } 40 cycles
Annealing and extension	60°C	1 min } 40 cycles

Table 3. Thermal cycling conditions for the two-step RT-qPCR amplification of *GAPDH* transcript. UNG: Uracil-N glycosylase.

The qPCR experiments report back a Ct value for each sample that represents the

cycle in which a real amplification signal is first detected. It is inversely correlated with the concentration; where the higher the concentration of a target nucleic acid in the starting material, the sooner a significant increase in amplification signal will be observed, yielding a lower Ct. The resulting samples with different Ct values were then tested in dPCR and based on that a threshold was set by which the *GAPDH* Ct value of each sample in qPCR dictated the amount of cDNA that can safely be applied in dPCR. *CDH1* and *GAPDH* reactions were subsequently run in multiplex using 10 or 20 ng cDNA, while *CDH1a* reactions were run in singleplex using 300 ng cDNA.

2.4. Two-step singleplex and multiplex RT-dPCR reactions

dPCR experiments were carried out using the chip-based QuantStudio™ 3D Digital PCR system (Applied Biosystems). In each case reaction mixes containing either cDNA or water (no-template controls) were first prepared by adding 2X QuantStudio 3D™ Digital PCR Master Mix v2 (Applied Biosystems) and 20X gene-specific assay in a total volume of 15.5 µl (Tables 4 and 5).

Reagent	Final concentration	Volume (µl)
Quantstudio 3D MM2 (2X)	1X	7.25
CDH1a assay (20X)	1X	0.725
≈ 50 ng/µl cDNA	≈ 300 ng	6
RNase-free water	-	to 15.5

Table 4. Concentrations and volumes of the reagents used in the two-step RT-dPCR singleplex amplification of *CDH1a* transcript.

Reagent	Final concentration	Volume (µl)
Quantstudio 3D MM2 (2X)	1X	7.25
CDH1 assay (20X)	1X	0.725
GAPDH assay (20X)	1X	0.725
≈ 5 ng/µl cDNA	≈10 or 20 ng	2 or 4
RNase-free water	-	to 15.5

Table 5. Concentrations and volumes of the reagents used in the two-step RT-dPCR multiplex amplification of *CDH1* and *GAPDH* transcripts.

Fifteen µl of each sample were then loaded into a blade that is firmly clasped onto the arm of the chip loader and evenly distributed into the chip's 20,000 nano-sized reaction wells. Each chip was then coated with Immersion Fluid and sealed with a QuantStudio™ 3D Digital PCR Chip Lid v2. Up to 24 chips were run simultaneously using GeneAmp PCR System 9700 (Applied Biosystem) by applying the thermal cycling conditions reported in Table 6.

Cycle steps	Temperature	Time
Initial denaturation	96°C	10 min
Annealing and extension	60°C	2 min } 45 cycles
Denaturation	98°C	30 sec } 45 cycles
Final extension	60°C	2 min

Table 6. Thermal cycling conditions for the two-step RT-dPCR amplification of *CDH1*, *CDH1a* and *GAPDH* transcripts.

At the end of the reaction chips were processed using the QuantStudio™ 3D Digital PCR system (Applied Biosystems) and analyzed with QuantStudio™ 3D Analysis Suite™ software (version 3.0.3). Only chips with at least 13,000 analyzable data points were accepted. The data were exported as a comma-separated values (CSV) file and included the number of copies per µl of each target in the loaded reaction volume (15 µl). This volume harbored 10 or 20 ng of cDNA in the case of *CDH1* and *GAPDH*, and 300 ng of cDNA in the case of *CDH1a*. Therefore, to determine the initial absolute amounts of the transcripts in each sample, the reported values were multiplied by 15 µl and the result was divided by the added amount of cDNA (10, 20 or 300 ng). The absolute values of *GAPDH* were subsequently used to normalize both *CDH1* and *CDH1a* expression in the tested tissue, since all three transcripts were derived from the same reverse transcription of the same RNA, extracted from each sample.

2.5. Two-step RT-qPCR quantification of *IFI27* and *IFITM1*

For the quantification of the *IFI27* and *IFITM1* interferon-inducible genes, RT-qPCR experiments were performed in triplicate on 20 ng of IGC tumor cDNA obtained from the previously described RT. The qPCR reactions were carried out using PrimeTime® Std qPCR assays (IDT) and the TaqMan® Universal Master Mix II, no UNG (Applied Biosystems). As depicted in Tables 7 and 8, the two target genes *IFI27* and *IFITM1* were amplified in multiplex with the endogenous controls, *HPRT1*and *RPLP0*, respectively.

Reagent	Final concentration	Volume (µl)
TaqMan® Universal Master Mix II no UNG (2X)	1X	10
IFI27 PrimeTime® Std qPCR assay (20X)	1X	1
HPRT1 PrimeTime® Std qPCR assay (20X)	1X	1
≈ 2.5 ng/µl cDNA	≈ 10 ng	4
RNase-free water	-	to 20

Table 7. Concentrations and volumes of the reagents used in the two-step RT-qPCR multiplex amplification of *IFI27* and *HPRT1* transcripts.

Reagent	Final concentration	Volume (µl)
TaqMan® Universal Master Mix II no UNG (2X)	1X	10
IFITM1 PrimeTime® Std qPCR assay (20X)	1X	1
RPLP0 PrimeTime® Std qPCR assay (20X)	1X	1
≈ 2.5 ng/µl cDNA	≈ 10 ng	4
RNase-free water	-	to 20

Table 8. Concentrations and volumes of the reagents used in the two-step RT-qPCR multiplex amplification of *IFITM1* and *RPLP0* transcripts.

All reactions were run on a 7500 Real-Time PCR (Applied Biosystems) by applying the thermal cycling conditions indicated in Table 9.

Cycle steps	Temperature	Time	
Polymerase activation	95°C	10 min	
Denaturation	95°C	15 sec	40 cycles
Annealing and extension	60°C	1 min	

Table 9. Thermal cycling conditions for the two-step RT-qPCR amplification of *IFI27*, *IFITM1*, *HPRT1* and *RPLP0* transcripts.

With respect to data analysis, it was done by means of the comparative $2^{-\Delta Ct}$ method (Schmittgen TD and Livak KJ, 2008), and the mean of the two endogenous control genes (*HPRT1* and *RPLP0*) was used to normalize the *IFI27* and *IFITM1* gene-targets' expression levels.

2.6. Selection of *CDH1*-regualting miRNAs in gastric cancer

An *in silico* analysis was executed to determine miRNAs that can potentially target the 3'UTR of *CDH1* and its regulators *SNAI1, SNAI2, TWIST1, ZEBI*, *ZEB2*, *EZH2* and *SUZ12*. To that purpose, the following three databases and predictive bioinformatics tools were utilized:

1. miRTarBase v6.0 (http://miRTarBase.mbc.nctu.edu.tw/), which is the most comprehensively annotated online database reporting experimentally validated microRNA-target interactions (MTIs). The database has thus far accumulated 4,966 articles, 7,439 strongly validated MTIs (through reporter assays or western blots) and 348,007 MTIs from crosslinking and immunoprecipitation sequencing (Chou CH *et al*., 2016). In addition to that, in the current version of the database you can query miR-target in disease, and get as a result a list of experimentally tested miRNAs and their targets that have been reported in that specific disease.
2. TargetScan v7.1 (http://www.targetscan.org/) that is an online site in which biological targets of miRNAs are predicted based on the presence of conserved 8mer, 7mer and 6mer sites matching the seed region of each miRNA (Lewis BP *et al*., 2005). It provides accurate rankings of the predicted targets for miRNAs based on either the probability of evolutionarily conserved targeting (P_{CT} scores) (Friedman RC *et al*., 2009) or the predicted efficacy of targeting (context+++ scores) (Agarwal V *et al*., 2015). TargetScanHuman is based on matches to human 3' UTRs and their orthologs, as indicated by UCSC (University of California, Santa Cruz) whole-genome alignments.
3. miRDB v5 (http://mirdb.org/miRDB/), which is an online database for miRNA target prediction and functional annotations. Predictions on MTIs can be queried by miRNA names or by gene-target information. The targets reported in this database have been derived from the bioinformatics tool MirTarget v3. This tool was developed upon the analysis of thousands of MTIs from high-throughput sequencing experiments, which allowed the identification of common features associated with miRNA target binding. These features were then acquired by machine learning methods to predict miRNA targets (Wong N and Wang X, 2015; Wang X, 2016).

This extensive analysis was combined with a search of the literature regarding miRNAs' expression in gastric cancer to guide the selection of a list of miRNAs to be differentially quantified in paired normal and tumor tissue of the IGC patients.

2.7. Two-step RT-qPCR amplification of selected miRNAs

With respect to the quantification of the selected miRNAs, 20 ng total RNA from the IGC paired normal and tumor tissue samples was first reverse transcribed using RT-specific primers and components of the TaqMan® MicroRNA Reverse Transcription Kit (Applied Biosystems). All steps were carried out according to the manufacturer's instructions in a final volume of 12 µl.

Subsequently qPCR reactions were performed in triplicate for *miR-92a* and *miR-101* transcripts, as well as the internal reference gene, *RNU6B*. These reactions were prepared separately with TaqMan® Universal Master Mix II, no UNG (Applied Biosystems) and either one the two TaqMan™ MicroRNA assays (Applied Biosystems) hsa-miR-101 or hsa-miR-92a; or the MicroRNA Control Assay (Applied Biosystems), RNU6B (Table 10).

Reagent	Final concentration	Volume (µl)
TaqMan® Universal Master Mix II no UNG (2X)	1X	10
TaqMan™MicroRNA Assay or Control Assay (20X)	1X	1
≈ 1.67 ng/µl cDNA	≈ 2.2 ng	1.33
RNase-free water	-	to 20

Table 10. Concentrations and volumes of the reagents used in the two-step RT-qPCR amplification of *miR-101, miR-92a* and *RNU6B* transcripts.

All reactions were run on a 7500 Real-Time PCR (Applied Biosystems) following the previously indicated thermal cycling conditions (refer back to Table 9). miR quantification was then evaluated by means of the comparative $2^{-\Delta Ct}$ method (Schmittgen TD and Livak KG, 2008).

2.8. Statistical analysis

To compare dPCR and qPCR gene expression results either a two-tailed Student's *t*-test or a two-sided Wilcoxon rank-sum test was used, depending on the data distribution. These two tests were also used to determine associations between clinic-pathological parameters and *CDH1* or miRNA expression levels. On the other hand, Fisher's exact test was applied in the case of associations between clinic-pathological parameters and *CDH1a*. In addition, Pearson's product-moment correlation test was selected to check for correlations between miRNAs and *CDH1* expression levels. All analysis was done with R statistical software version 2.14.1 and *P*-values ≤ 0.05 were considered as statistically significant for each comparison.

Results

2.1. Setting up dPCR for *CDH1* transcripts' analysis in IGC patients

In collaboration with IRST-IRCCS, Meldola (FC), fresh-frozen tissue samples were first collected from 32 patients with intestinal-type gastric cancer (Table 1).

Parameter [a, b]	Total	
	n (32)	%
Sex		
F	14	43.8
M	18	56.2
Age		
≤65	9	28.1
66-75	10	31.3
>75	13	40.6
T		
1	2	6.3
2	17	53.1
3	12	37.5
4	1	3.1
N		
0	12	37.5
1	10	31.3
2	6	18.7
3	4	12.5
M		
0	21	65.6
1	2	6.3
NA	9	28.1
Grade		
1	1	3.1
2	9	28.1
3	21	65.7
NA	1	3.1
Tumor site		
Distal third (L)	7	21.9
Middle third (M)	24	75
Proximal third (U)	1	3.1
Tumor size (cm)		
2-5	15	46.9
5-10	15	46.9
>10	1	3.1
NA	1	3.1
Helicobacter Pylori		
Positive	15	46.9
Negative	15	46.9
NA	2	6.2
Adjuvant chemotherapy		
Yes	10	31.3
No	16	50
NA	6	18.7

Table 1. Clinic-pathological parameters of IGC patients. [a] The mean age at diagnosis was 72.8. [b] Tumor staging was done based on the tumor (T), lymph node (N) and metastasis (M) system. NA: not available.

For 21 of these patients paired tumor and normal tissue could be obtained, while for the remaining 11, only tumor tissue was available. RNA was extracted from all tissue sections and reverse transcribed to cDNA. The cDNA was applied in the Quantstudio™ 3D dPCR system for quantifying the *CDH1* canonical and *CDH1a* non-canonical transcripts, as well as the *GAPDH* transcript, which was included as an internal control.

As per the essential requirements of the dMIQE checklist (Huggett JF *et al*., 2013; see Appendix A), only samples with good-quality RNA and pure cDNA could be included in the analysis. To that purpose, the Experion assay was applied to assess RNA integrity (Figure 1); an inhibition testing experiment on a randomly selected sample ascertained the purity of the cDNA (Figure 2).

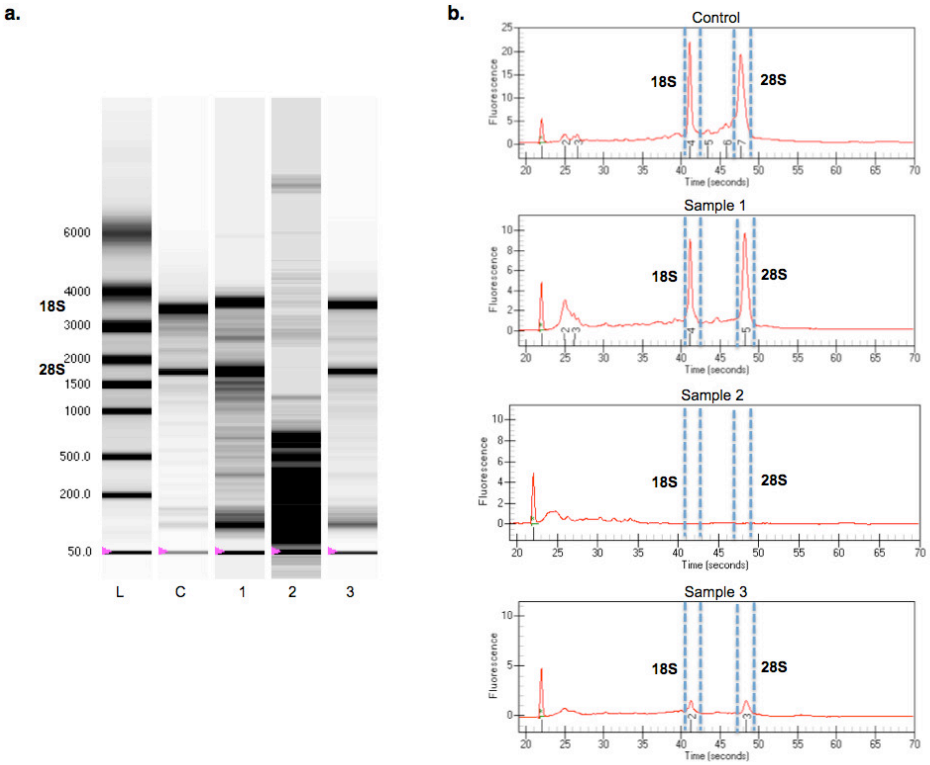

Figure 1. Experion results. a. Gel electrophoresis of Experion RNA ladder (L), Experion total RNA control (C) and RNA derived from 3 IGC patients (1, 2, 3). b. Electropherogram of the Experion total RNA control and of the RNAs from the 3 patients. The relative positions of the 18S rRNA and 28S rRNA are indicated. Only samples showing clear peaks for the two rRNA comparable with the control RNA were selected. In this example, sample 2 was excluded.

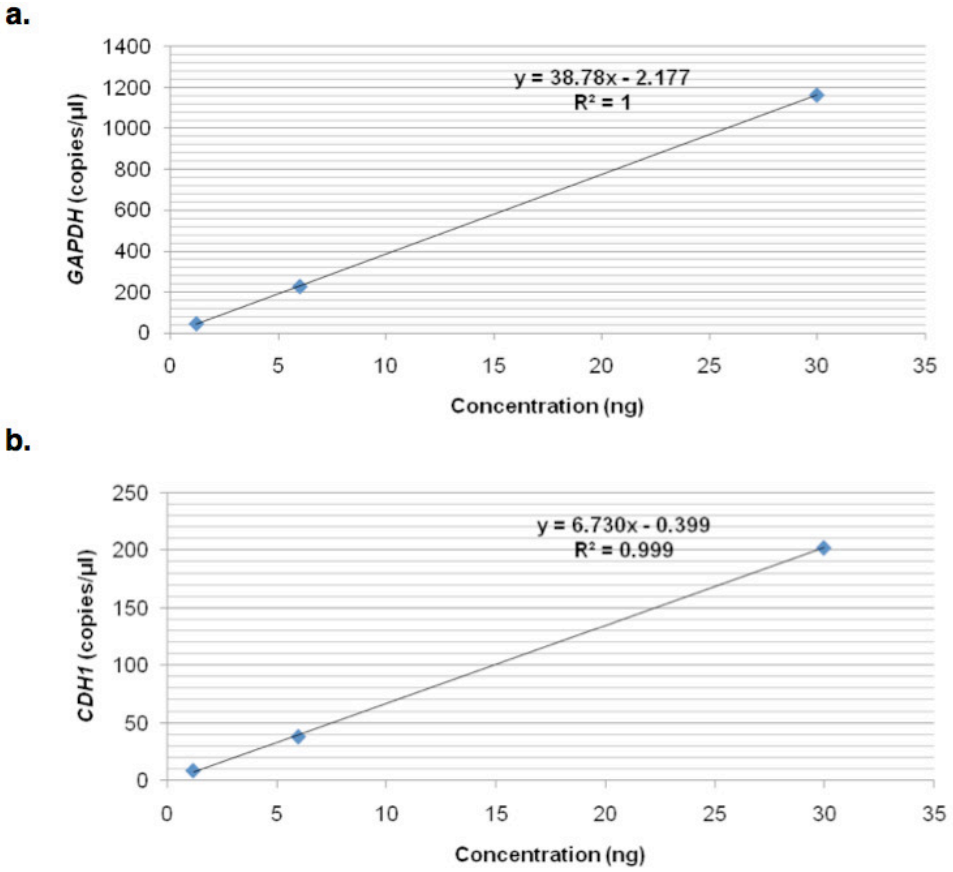

Figure 2. dPCR inhibition testing. 1:5 serial dilutions were performed on the cDNA of a randomly selected sample to test for the effect of inhibitors on dPCR. The number of copies per µl was obtained from the dPCR reactions for each dilution for *GAPDH* (a) and *CDH1* (b). In this case, there is an almost perfect correlation between the added amount of cDNA (ng) and the reported number of copies per µl (coefficient of determination R^2 is almost 1), indicating that there are no inhibitors affecting the reaction.

Not only is the quality of the integrated biological material important, but also its amount. To ensure the most accurate copy number calculation, various amounts of cDNA were tested in the Quantstudio™ 3D dPCR system for *CDH1* and *CDH1a* expression, either alone (singleplex dPCR) or in combination with *GAPDH* (multiplex dPCR) using the relevant assays. Furthermore, since *GAPDH* was the most highly expressed transcript among the 3, it was also used to dictate the starting amount of cDNA. This was accomplished by first applying a straightforward two-step RT-qPCR experiment to acquire an estimate of *GAPDH* levels in each sample on the basis of the reported Ct values. As a result, *CDH1* and

GAPDH could be successfully multiplexed using 10 or 20 ng of cDNA, depending on the sample. Conforming to the dMIQE checklist once more, the feasibility of the multiplex dPCR reaction was confirmed when the comparison of normalized *CDH1* expression levels in a couple of samples, in singleplex and multiplex settings, gave very similar values (coefficient of variation (CV) = 5%).
On the other hand, being a rare transcript, *CDH1a* necessitated the use of the maximum amount of cDNA possible, which in our case was 300 ng. This amount was up to 30 times more than that required for accurate copy number quantification of *GAPDH*, which led to *CDH1a* dPCR reactions being set up in singleplex.
In addition, again following the dMIQE guidelines, we tested the repeatability of our method by performing intra-assay dPCR technical replicates of 4 randomly selected samples (starting from the same RT), and the average CV was found to be 10%.
The final general applied workflow for all dPCR experiments is depicted in Figure 3.

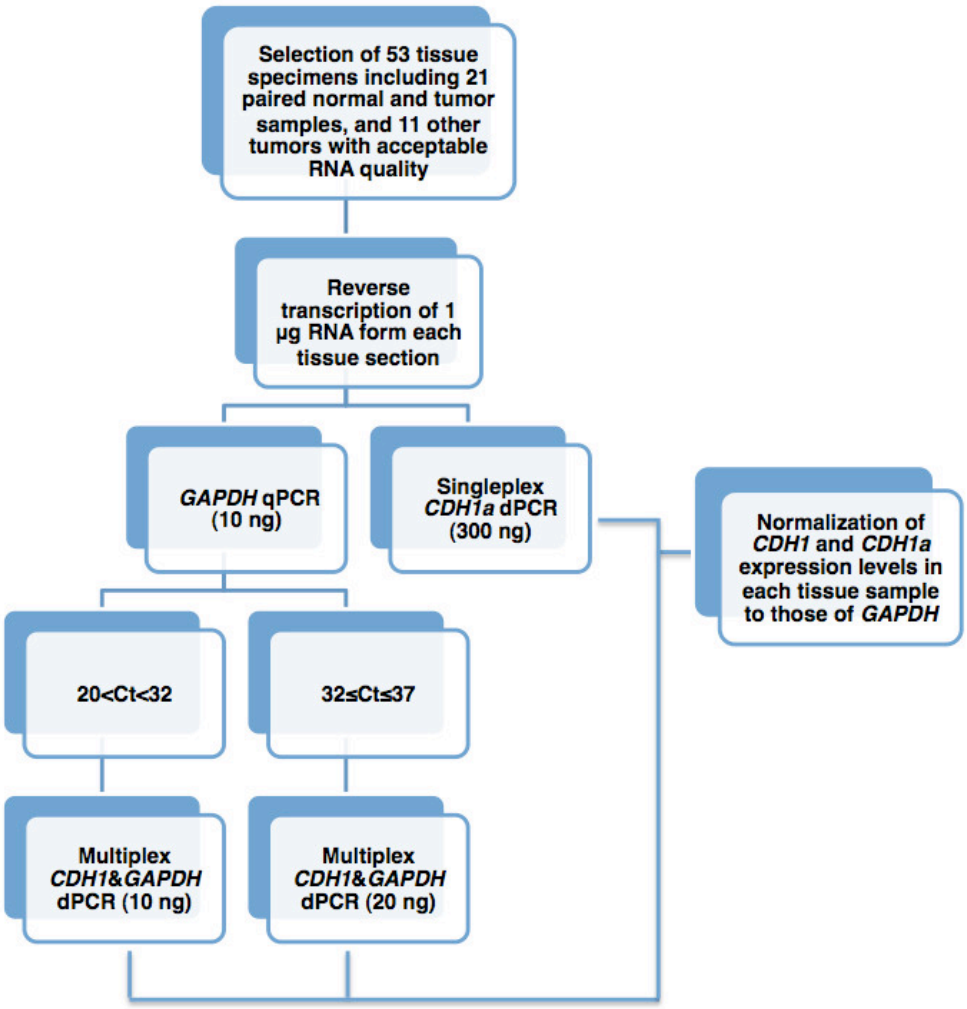

Figure 3. General workflow for dPCR analysis of the *CDH1* transcripts.

Moreover, examples of dPCR chips and the corresponding output scatter plots of negative controls and samples that are either negative or positive for *CDH1* and *CDH1a* expression, are provided in Figure 4.

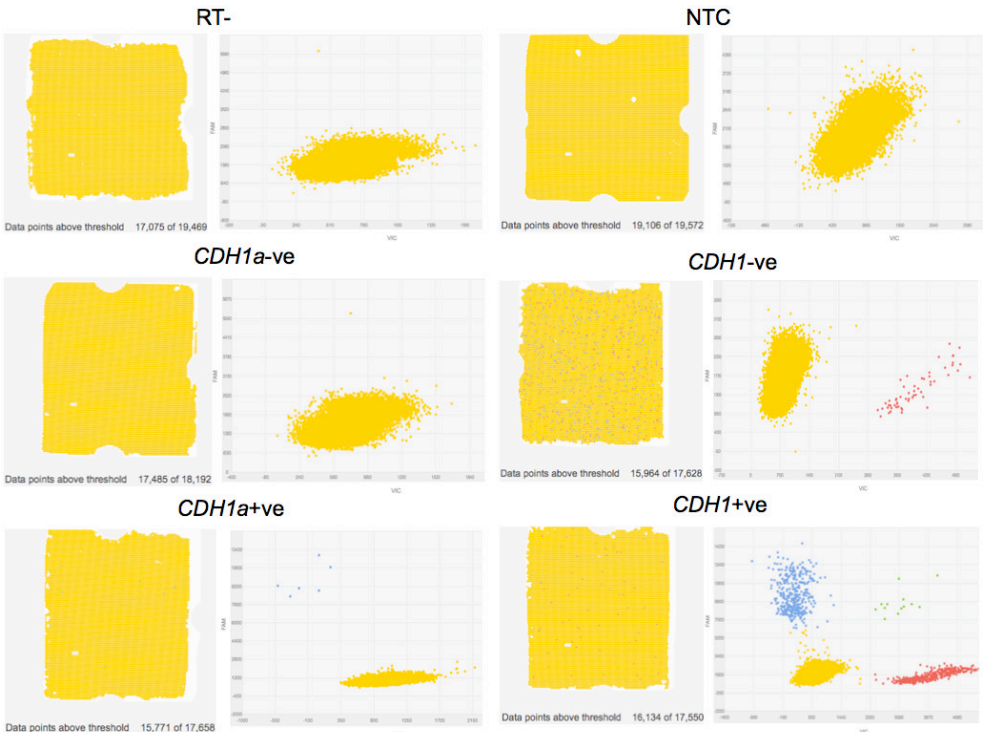

Figure 4. Typical dPCR results. For each sample, on the left side is the chip view, which shows the wells and the number of analyzable data points. In our analysis we only accepted samples with at least 13,000 data points above threshold. On the right side is the scatter plot showing the distribution of the data points based on the dyes used (VIC and FAM). Yellow refers to "No Amplification", red to the VIC amplified *GAPDH*, blue to the FAM amplified *CDH1* or *CDH1a*, and green to co-amplified *CDH1* and *GAPDH*. RT- refers to a sample reverse transcribed without including the reverse transcriptase, serving as a negative control for the RT reaction. NTC refers to a no-template control in which water is added instead of cDNA, serving as a negative control for the dPCR reaction. *CDH1a*-ve and *CDH1*-ve are examples of samples negative for the indicated transcript. *CDH1a*+ve and *CDH1*+ve are examples of samples positive for the indicated transcript.

2.2. Canonical *CDH1* expression levels in IGC

By multiplex dPCR, we compared the expression levels of *CDH1* in tumor and corresponding normal tissue samples from 21 IGC patients using a custom designed *CDH1* assay and a pre-designed *GAPDH* assay. Figure 5a depicts an example of dPCR output scatter plots obtained for paired samples from the same subject. The analysis of the distribution of *CDH1* expression levels in normal and cancer tissue samples, following normalization to the *GAPDH* endogenous control, revealed a significantly lower level of *CDH1* in tumors compared to normal samples (P=0.001) (Figure 5b). In particular, reduced *CDH1* expression by at least 1.5 times was found in 16 out of 21 cases (76%).

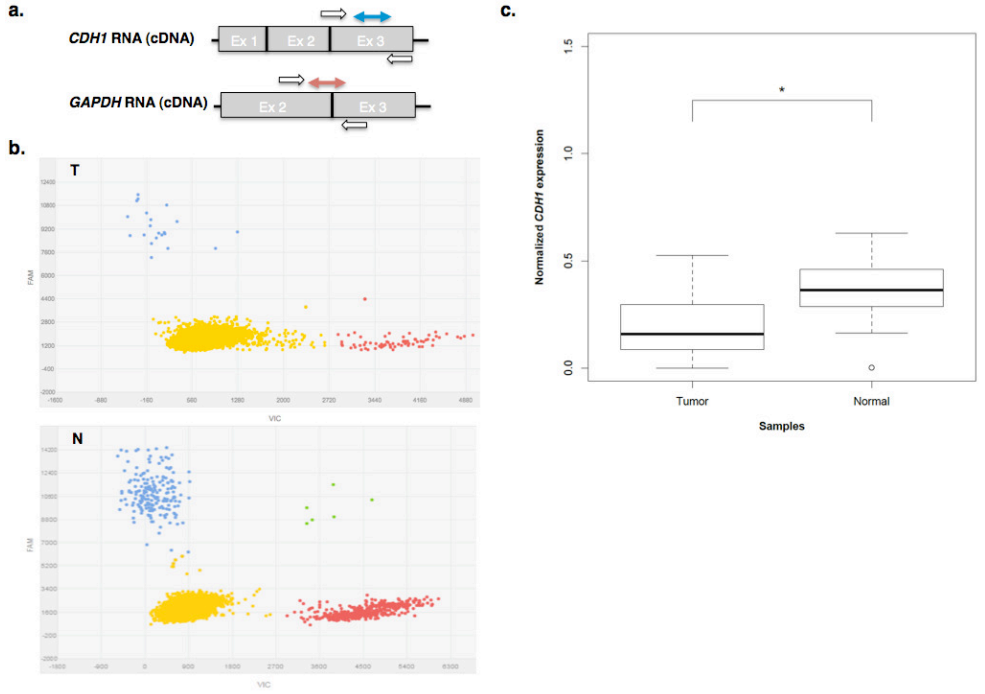

Figure 5. *CDH1* expression analysis in multiplex dPCR. a. A scheme of the assays applied; primers and probes are depicted by single and double arrows, respectively. b. Typical dPCR output scatter plots of tumor (T) and normal (N) samples showing the distribution of the data points based on the dyes used (VIC and FAM). Yellow refers to "No Amplification"; red to VIC amplified *GAPDH*; blue to FAM amplified *CDH1*; and green to co-amplified *CDH1* and *GAPDH*. c. Box plots of normalized *CDH1* expression levels in 21 tumors compared to the paired normal tissue. "*" refers to a statistically significant difference with a *P*-value=0.001 as calculated by Student's *t*-test. Ex: exon.

2.3. Non-canonical *CDH1a* expression levels in IGC

By singleplex dPCR, we then determined the differential expression of *CDH1a* in the 21-paired samples, as well as in 11 additional tumor samples for which the corresponding normal tissue was not available, using a custom designed *CDH1a* assay. We could detect *CDH1a* at a very low level in 15 out of 32 (47%) tumors. Under the same experimental conditions, *CDH1a* transcript proved to be undetectable in normal tissue samples, including those corresponding to *CDH1a*-positive tumors. In these tumors, the amount of *CDH1a* was too low to provide for accurate numerical dPCR quantitation; accordingly, we grouped cases as simply being *CDH1a* positive or negative. Figure 6 shows an example of dPCR output scatter plots obtained for paired samples from the same subject, with *CDH1a* being barely detectable in tumor and undetectable in normal cDNA

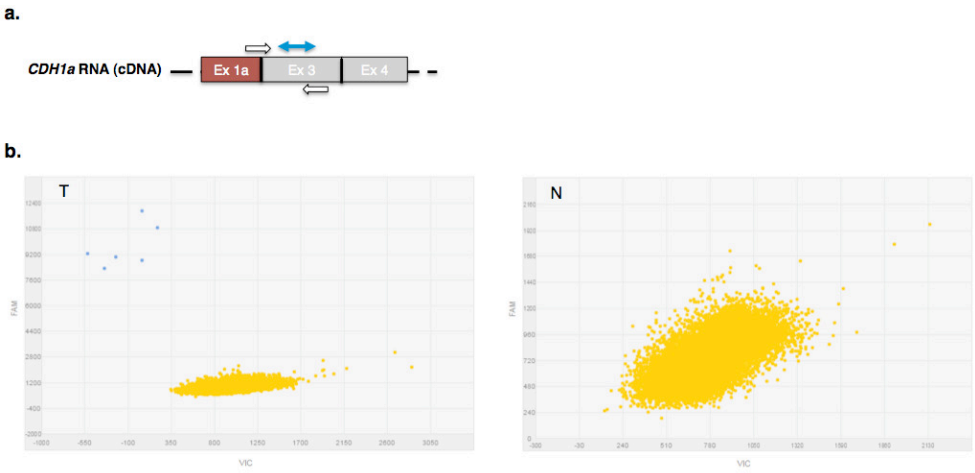

Figure 6. *CDH1a* expression analysis in singleplex dPCR. . a. A scheme of the assay applied; primers and probes are depicted by single and double arrows, respectively. b Typical dPCR output scatter plots of tumor (T) and normal (N) samples showing the distribution of data points based on the dye used (FAM). Yellow refers to "No Amplification" and blue to FAM amplified *CDH1a*. Ex: exon.

2.4. Effect of the presence of *CDH1a* on *CDH1* expression

In order to determine whether the expression of the non-canonical transcript, *CDH1a*, was affecting that of *CDH1*, we compared *CDH1* expression levels in the presence and absence of *CDH1a* in the 34 IGC tumors. A non-statistically significant trend was observed of having more *CDH1* in tumors lacking *CDH1a* (*P*=0.455) (Figure 7).

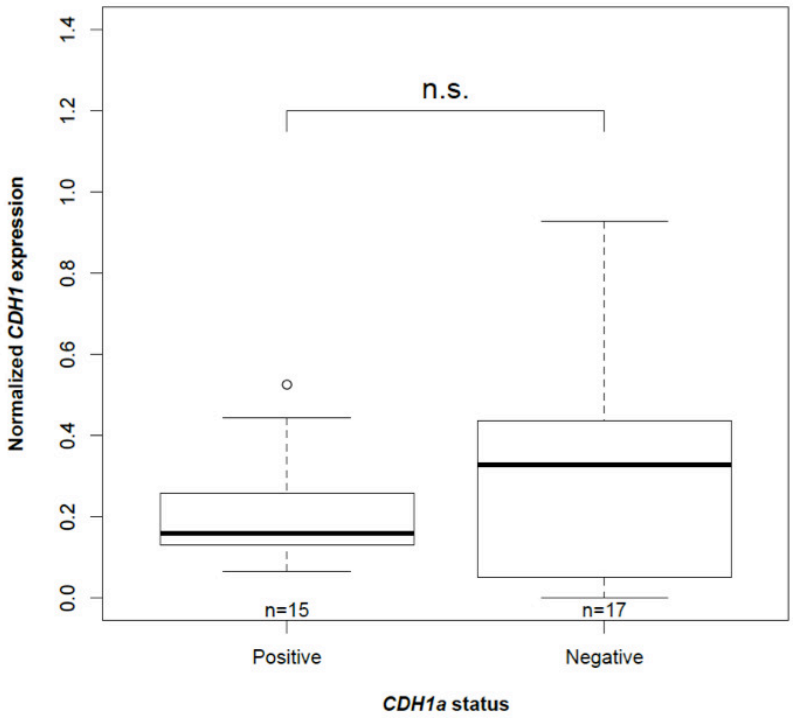

Figure 7. Association between *CDH1* expression and *CDH1a* status. "n" is the number of IGC tumors in each category. "n.s." refers to a statistically non-significant difference with a *P*-value=0.455 as calculated by Wilcoxon rank sum test.

2.5. Effect of the presence of *CDH1a* on interferon-inducible genes

It has previously been reported (Pinheiro *et al*., 2012) that in GC transfected cells the induced overexpression of *CDH1a* increases the expression levels of the interferon-inducible *IFITM1* and *IFI27* genes. To evaluate whether a similar effect occurs in IGC patients, we applied RT-qPCR and compared *IFITM1* and *IFI27* expression in 15 *CDH1a* -positive and 15 -negative tumors. In doing so, we report no significant difference in the expression of either *IFITM1* (*P*=0.486) or *IFI27* (*P*=0.683) between the *CDH1a* positive and negative tumors (Figure 8).

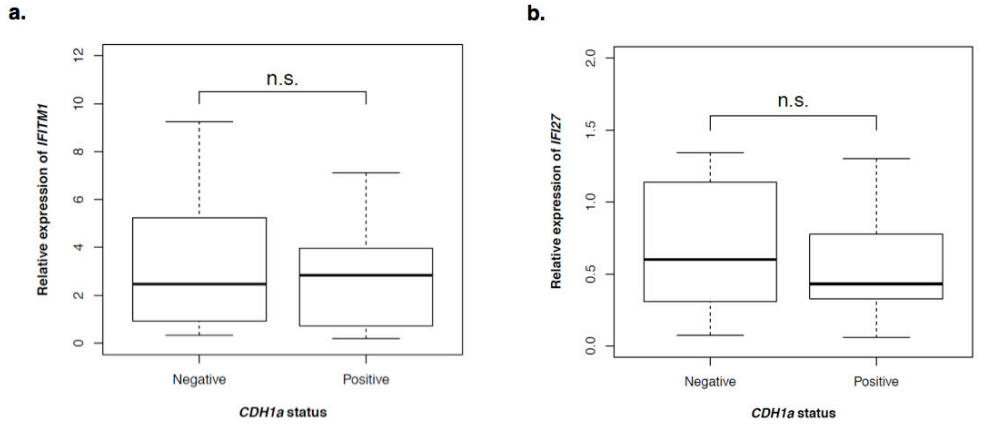

Figure 8. Association between the relative expression of interferon-inducible genes and *CDH1a* status. Box plots of the relative expression of *IFITM1* (a) and *IFI27* (b) to the mean of the internal controls genes *HPRT1* and *RPLP0*, in 15 *CDH1a* negative and 15 *CDH1a* positive tumors. "n.s." refers to a statistically non-significant difference with *P*-value=0.486 (a) and *P*-value=0.683 (b) as calculated by Wilcoxon rank sum test.

2.6. Selection of *CDH1*-regulating miRNAs

In order to select miRNAs that potentially, directly or indirectly, regulate *CDH1* expression in intestinal gastric cancer, an *in silico* analysis coupled with a literature search was executed.

The *in silico* analysis consisted of first of all querying the *CDH1* gene and its regulators (*SNAI1, SNAI2, TWIST1, ZEBI*, *ZEB2*, *EZH2* and *SUZ12*) in three different bioinformatics databases and predictive tools, namely miRTarBase, TargetScan and miRDB. At the end of this work, 62 miRNAs were simultaneously reported by all three tools to modulate the queried genes.

To better constrain this list, miRTarBase was additionally used to view the miR-target interactions that have been reported in gastric cancer. The miRNAs targeting our genes of interest were then fished out and cross-matched with the original list of 62 miRNAs. This approach led to the selection of 10 potentially interesting miRNAs. Meanwhile, 6 other miRNAs were selected based on findings in the literature implicating them in gastric cancer.

Table 2 reports all 16 miRNAs to be investigated and the means by which they have been selected.

miRNA	*In silico prediction*				*Experimental evidence*
	Target gene	**miRTarBase**	**miRDB**	**TargetScan**	**miR-target interaction (+) or miR expression deregulation (ref.) in GC**
miR-506	*SNAI2*	+	+	+	+
miR-141	*ZEB1* *ZEB2*	+ +	+ +	+ +	+
miR-217	*EZH2*	+	+	+	+
miR-429	*EZH2*	+	+	+	+
miR-199a	*CDH1*	+	+	+	+
miR-200a	*ZEB1* *ZEB2*	+ +	+ +	+ +	+
miR-200b	*ZEB1* *ZEB2*	+ +	+ +	+ +	+
miR-200c	*ZEB1* *ZEB2* *SUZ12*	+ + +	+ + +	+ + +	+
miR-101	*EZH2* *ZEB1*	+ +	+ +	+ +	+
miR-153	*SNAI1*	+	+	+	+

miR-26b	*EZH2*	+	+	+	Juzėnas S *et al*., 2015; Li FQ *et al*., 2015
miR-23a	*CDH1*	+	+	+	+
miR-544a	*CDH1*	+	+	+	+
miR-34c	*SNAI1*	+	+	+	Jiang C *et al*., 2015
miR-203	*SNAI2* *ZEB1*	+ +	+ +	+ +	+
miR-92a	*CDH1*	+		+	Shrestha S *et al*., 2014

Table 2. List of 16 selected miRNAs, their target genes and means of selection. GC: gastric cancer.

Among the selected miRNAs: 4 target *CDH1* (*miR-199a*, *miR-23a*, *miR-544a*, *miR-92a*); 2 target *SNAI1* (*miR-153*, *miR-34c*); 1 targets *SNAI2* (*miR-506*); 1 targets both *SNAI2* and *ZEB1* (*miR-203*); 3 target concomitantly *ZEB1* and *ZEB2* (*miR-141*, *miR-200a*, *miR-200b*); 1 targets *SUZ12* as well as *ZEB-1* and *-2* (*miR-200c*); 3 target *EZH2* (*miR-217*, *miR-429*, *miR-26b*); and finally 1 targets both *EZH2* and *ZEB1* (*miR-101*). In this way we would be covering all genes of interest by selecting at least one potential regulatory miRNA for each.

2.7. Expression analysis of *CDH1*-regulating miRNAs in IGC

Once the list of miRNAs has been established, the next step was to determine the differential expression of these miRNAs in the paired normal and tumor tissue sections of the 21 IGC patients. This was done starting with the two miRNAs, *miR-92a* (targeting *CDH1*) and *miR-101* (targeting *EZH2* and *ZEB1*), by performing two-step RT-qPCR using the pertinent TaqMan assays. While the relative expression levels of *miR-92a* were found to be quite similar in the normal and tumor tissue (P=0.850; Figure 9a), the levels of *miR-101* were found to significantly decrease in tumors compared to the normal gastric mucosa (P=1.565x10^{-05}; Figure 9b).

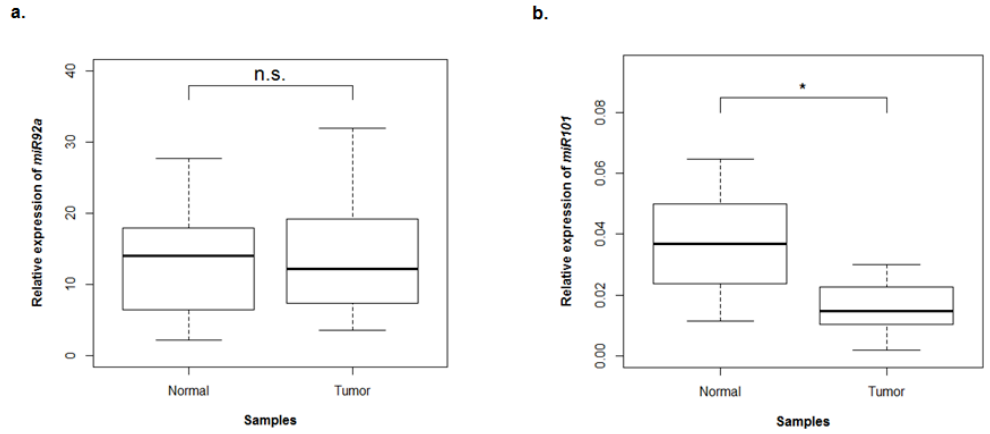

Figure 9. *miR-92a* and *miR-101* expression analysis in real time PCR. Box plots of the relative expression of *miR-92a* (a) and *miR-101* (b) to the internal control gene *RNU6B* in 21 tumors compared to the paired normal tissue. "n.s" refers to a statistically non-significant difference with a *P*-value=0.850 as calculated by Wilcoxon rank sum test; "*" refers to a statistically significant difference with a *P*-value=$1.565x10^{-05}$ as calculated by Wilcoxon rank sum test.

2.8. Correlations of *miR-92a* and *miR-101* expression levels with *CDH1*

miR-92a is a predicted direct regulator of *CDH1*, while *miR-101* is an indirect regulator that affects *CDH1* expression by repressing its inhibitor, EZH2. We therefore determined whether a correlation existed between the variation of *CDH1* expression and the two miRNAs in the paired normal and tumor tissue samples of the investigated IGC patients. In the optimal scenario, a perfect positive correlation corresponds to 1 (r=1), while a perfect negative correlation is represented by -1 (r=-1). In our analysis, r turned out to be 0.12 between *miR-92a* and *CDH1*, (Figure 10a), and -0.18 between *miR-101* and *CDH1* (Figure 10b). Therefore, there was no correlation between the expression of either miRNA and *CDH1* in the examined case series.

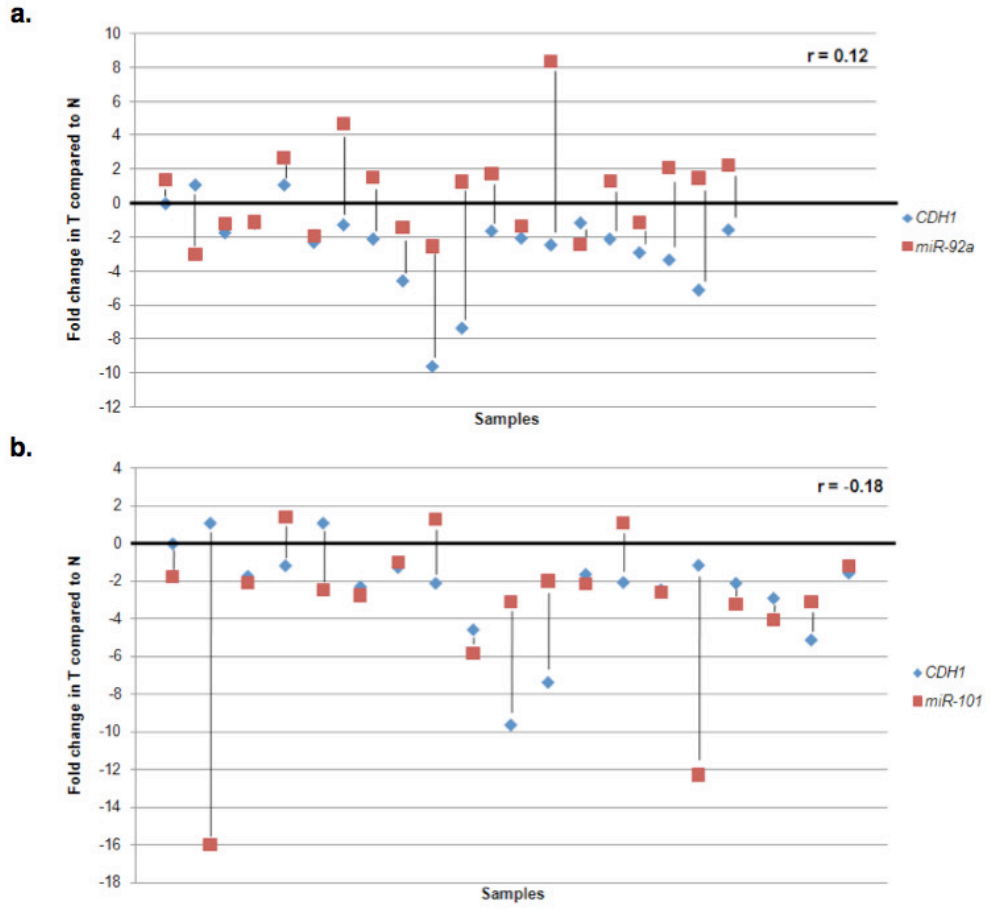

Figure 10. Scatter plots of the fold change of miRNAs and canonical *CDH1* expression in tumors compared to paired normal tissue samples. a. Fold change of *miR-92a* and *CDH1* expression levels. b. Fold change of *miR-101* and *CDH1* expression levels. "r" refers to the Pearson's correlation coefficient as calculated by the Pearson's product-moment correlation test. T: tumor; N: normal.

2.9. Associations with clinic-pathological parameters

Possible associations between patients' clinical parameters and *CDH1* expression levels, *CDH1a* presence/absence, as well as miRNA expression profiles, were also evaluated. This was done by dividing the parameters of interest into two different groups and subsequently testing for potential associations by applying the appropriate statistical test (Table 3).

Clinic-pathological parameter	Compared groups	*P*-value			
		CDH1	***CDH1a***	***miR-92a***	***miR-101***
Grade	G1+G2 *vs.* G3	0.224ǂ	0.600	0.535	0.424ǂ
Tumor size (T)	T = 1&2 *vs.* T = 3&4	0.404	1	0.553	0.717ǂ
Lymph node (N)	N = 0 *vs.* N = 1,2&3	0.430	0.328	0.349	0.393
H. pylori infection status	*H. pylori* negative *vs. H. pylori* positive	0.870	1	1	0.348ǂ

Table 3. Associations between clinic-pathological parameters, *CDH1* transcripts and miRNAs. *CDH1*, *miR-92a* and *miR-101 P*-values are calculated by Wilcoxon rank sum test; *CDH1a P*-values are calculated by Fisher's Exact Test. "ǂ" refers to *P*-values calculated by Student's *t*-test.

None of the tested associations showed any clear-cut significance. However, a trend of lower *CDH1* gene expression in tumors with a higher grade (G3 *vs.* G2+G1) and in those positive for *H. pylori* infection could be observed (Figure 11).

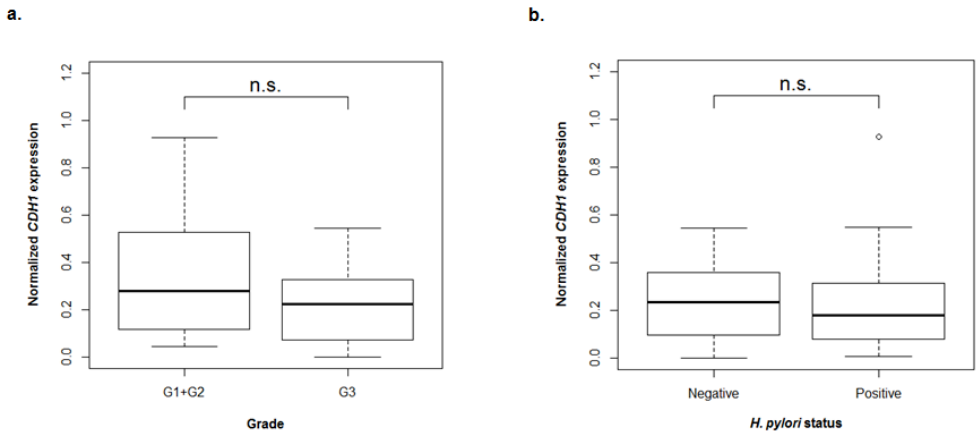

Figure 11. Association between *CDH1* expression and clinic-pathological parameters. a. Box plots of the normalized expression levels of *CDH1* in tumors of G1 and G2 grade in comparison with those of G3 grade. b. Box plots of the normalized expression levels of *CDH1* in tumors that are *H. pylori* negative in comparison with those that are *H. pylori* positive. "n.s." refers to a statistically non-significant difference with *P*-value=0.224 as calculated by Student's *t*-test (a) and *P*-value=0.870 as calculated by Wilcoxon rank sum test (b).

Discussion

Different genetic and epigenetic lesions underlie the carcinogenic processes involved in intestinal- and diffuse- types gastric cancer (Mclean MH and El-Omar EM, 2014). In particular, whole-genome sequencing and gene expression and methylation profiling have identified subtype-specific molecular signatures (Wang K *et al*., 2011; Zang ZJ *et al*., 2012; Wang K *et al*., 2014). Such comprehensive molecular approaches also highlighted that adherens junctions and focal adhesions are driver pathways in gastric carcinogenesis, with alterations in genes associated with these pathways occurring in most GC cases (Wang K *et al*., 2011; Zang ZJ *et al*., 2012; Wang K *et al*., 2014). One adhesion-related gene family largely implicated in carcinogenesis of both the intestinal and diffuse GC histotypes is the cadherin family, with the best-known member being *CDH1*. *CDH1* encodes E-cadherin, a calcium-dependent transmembrane adhesion protein. The majority of GCs show an immunohistochemical aberrant pattern (mislocalization) of E-cadherin expression, while complete protein loss is highly predominant in GCs of the diffuse histotype (Stănculescu D *et al*., 2011; Carvalho J *et al*., 2012; Corso G *et al*., 2013). Alternative mechanisms have been implicated in the negative regulation of *CDH1*, indicating the existence of factors acting at different levels that can subtly modulate E-cadherin expression in gastric cancer of the intestinal type (Carneiro P *et al*., 2012; Carvalho J *et al*., 2012). Among those factors, are small regulatory RNAs (micro-RNAs), which have been associated with a decrease of E-cadherin in IGC (Carvalho J *et al*., 2012). In addition, an intron-mediated mechanism of *CDH1* regulation has also been identified (Stemmler MP *et al*., 2005; Carvalho J *et al*., 2012; Pinheiro H *et al*., 2012). In particular, intron 2, harboring an exceptionally high number of repetitive elements involved in exonization, can act as a *cis*-modulator of E-cadherin gene and protein expression (Stemmler MP *et al*., 2005; Oliveira P *et al*., 2012; Alotaibi H *et al*., 2015). In certain cell lines, intron 2 has been shown to give rise to a number of non-canonical transcripts, one of which, *CDH1a*, harbors properties that enable its translation into a protein isoform differing from the canonical E-cadherin in its N-terminal domain (Pinheiro H *et al*., 2012). Functional assays performed on gastric cancer transfected cells associated the overexpression of *CDH1a* with increased angiogenesis and invasion in the presence of the canonical transcript (Pinheiro H *et al*., 2012). These findings make *CDH1* gene transcripts likely players in gastric carcinogenesis of the intestinal type, where some level of E-cadherin expression is often retained.

In this study we evaluated *CDH1* gene expression in intestinal-type gastric cancer by quantifying *CDH1* canonical and *CDH1a* non-canonical transcripts in RNA

from normal and cancer tissue samples by means of digital-PCR (dPCR). In addition, we attempted to characterize the potential mechanisms by which *CDH1a* is exerting its proposed tumorigenic effect by quantifying the expression of interferon-inducible genes in the tumor tissue samples. We then performed an *in silico* analysis for the selection of suitable *CDH1*-regulating miRNAs and initiated the evaluation of their expression in the paired tissue samples. Finally, clinic-pathological associations with the aforementioned factors were tested.

dPCR is a method for sensitive measurement and quantification of nucleic acids. It improves precision and reproducibility with respect to real-time quantitative PCR (qPCR) and provides an alternative approach for detection of gene expression in settings where the target RNA is limited or present in quantity that approaches the limits of qPCR sensitivity (Huggett JF *et al*., 2013; Takahashi K *et al*., 2014). The reason behind this sensitivity is the partitioning of the same sample into thousands of reaction chambers, such that each has on average a single copy of target nucleic acid or none at all. Subsequently, Poisson statistics are utilized to convert the count of positive signals into an absolute number. This "divide and conquer" strategy applied by dPCR can be performed on two different platforms, droplet digital PCR (ddPCR) and chip-based dPCR. The main difference between the two is that the first partitions the sample into thousands of droplets, while the second partitions them into minuscule wells. The advantage of applying a chip-based digital-PCR approach is that, first of all, it requires less pipetting steps, which reduces errors and contamination risks; second, the generation of droplets for sample compartmentalization in ddPCR might be extremely variable, especially compared to the partitioning into standard-sized wells (Conte D *et al*., 2015). Both ddPCR and chip-based dPCR, which will be referred to as simply dPCR, have already been successfully applied in multiple experimental settings, including the quantification of extracellular RNA (Takahashi K *et al*., 2014), water-born RNA viruses (Rački N *et al*., 2014), circulating miRNAs (Conte *et al*., 2015) and RNA transcripts derived from bone marrow cells (Albano F *et al*., 2015), as well as formalin-fixed paraffin-embedded tissue (Heredia NJ *et al*., 2013). With the advent and increased application of this technology, Huggett JF and coworkers (2013), took it upon themselves to report the Minimum Information for Publication of Quantitative Digital PCR Experiments (dMIQE) guidelines. These guidelines are aimed at standardizing experimental protocols, maximizing efficient utilization of resources and enhancing the impact of this promising new technology on the scientific community (Huggett JF *et al*., 2013).
In keeping with these guidelines, for the amplification of *CDH1* and *CDH1a* using Quantstudio™ 3D dPCR system, we made sure to accomplish the following: include good quality RNA and inhibitors-free cDNA; report the details of the

protocols and assays used; utilize an internal reference gene, *GAPDH*, for normalizing the results; prove the feasibility and repeatability of our experimental setup.
Following the fine-tuning of dPCR, quantification of the canonical transcript demonstrated that *CDH1* was significantly less expressed in cancer tissue compared to normal mucosa of the same patients (P=0.001); this downregulation was especially evident (at least 1.5 times less) in 76% of IGC tumors. Previous evaluations obtained by qPCR on intestinal GCs gave variable results with reports ranging from a clear decrease (Rosivatz E *et al*., 2002; Ferraz MA *et al*., 2016) to a non-significant difference (Yang M *et al*., 2014; Ibarrola-Villava M *et al*., 2015) in *CDH1* expression in cancer compared to the normal counterpart. Therefore, through the implementation of the more precise dPCR in our quantification we were able to provide evidence on the downregulation of *CDH1* expression in IGC.

By applying the same experimental approach, we determined for the first time the expression of *CDH1a* in IGC and could detect *CDH1a* in a fraction of tumors (47%), while no dPCR amplification signal was observed in the tested normal tissue samples. This finding is in line with data reported by Pinheiro H and coworkers (2012) who, by Quantitative-SnapShot method, found *CDH1a* to be expressed in gastric cancer cell lines but not in commercially available RNA from the normal stomach. However, while most of their tested cell lines expressed high levels of *CDH1a*, this transcript proved to be barely detectable in the cancer samples we analyzed. A possible explanation is intratumor cell heterogeneity or "dilution" of the *CDH1a* dPCR amplification signal by the presence of normal cells. What is more likely, however, is that *CDH1a* is indeed expressed at a very low level in tumors, while at a high level in cell lines due to positive selection during cell line stabilization.

It has been reported that in transfected GC cell lines, the induced overexpression of *CDH1a* leads to an increase of *IFITM1* and *IFI27* interferon-inducible genes (Pinheiro H *et al*., 2012). We hence evaluated whether the presence of *CDH1a* in tumors could be stimulating the expression of these genes in tissue samples derived from IGC patients. RT-qPCR quantification of the *IFITM1* and *IFI27* gene transcripts, however, showed no detectable association with *CDH1a* presence/absence. This result does not necessarily contradict the notion that the aforementioned factors are downstream effectors of *CDH1a*; it just shows that in gastric cancer tumors, the amount of *CDH1a* present is not sufficient to exert any substantial action on them.

The fact that *CDH1a* is expressed only in a fraction of tumors raised the question

of whether its presence could be interfering with that of the canonical *CDH1* transcript. Upon investigating the interplay between *CDH1* and *CDH1a*, we found that *CDH1a* was present in tumors with lower levels of *CDH1*, suggesting a shift of transcription factors and/or other components of the transcription initiation machinery in favor of one transcript over the other in cancer cells (reviewed by Pal S *et al*., 2012). These types of events, together with disturbance of alternative splicing programs have frequently been linked to the carcinogenic process (reviewed by Biamonti G *et al*., 2014).
In addition, intron 2, from which *CDH1a* arises, contains multiple transcription initiation sites and evolutionary conserved elements, as well as enhancers and repeated sequences (Stemmler MP *et al*., 2005; Oliveira P *et al*., 2012; Pinheiro H *et al*., 2012; Alotaibi H *et al*., 2015). In particular, surrounding exon 1a different regulatory sequences have been identified, including CpG islands, DNaseI hypersensitive sites and CTCF (CCCTC-binding factor) insulator elements (Pinheiro H *et al.,* 2012). A dysfunction of any of these elements can result in *CDH1* gene transcripts' imbalance.

miRNAs can additionally affect *CDH1* regulation either directly, by targeting *CDH1* itself, or indirectly by targeting members of its regulatory network, including transcription factors and trimethylation complexes. Therefore, we applied a concomitant *in silico* and literature-based approach to determine the most likely miRNAs to be implicated in the modulation of these factors in IGC. We queried *CDH1*, its transcriptional repressors *SNAI1*, *SNAI2*, *TWIST*, *ZEB1*, *ZEB2* and epigenetic inhibitors *SUZ12* and *EZH2* in three different databases and predictive bioinformatics tools. By crosschecking the results of the *in silico* analysis among each other, and against the literature, we could select a promising list of 16 miRNAs. Among the list of miRNAs, were the potential oncomiR *miR-92a*, and the tsmiR *miR-101*. Regarding *miR-92a*, it has been reported to be one of the most consistently upregulated miRNAs in gastric cancer (reviewed by Shrestha S *et al*., 2014). It was also shown to directly target *CDH1* expression in esophageal squamous cell carcinoma (ESCC), where it inversely correlated with the expression of *CDH1* in both ESCC tumor tissues and cell lines (Chen ZL *et al*., 2011). On the other hand, *miR-101* has been shown to be significantly downregulated in gastric cancer (Wang HJ *et al*., 2010; Carvalho J *et al*., 2012), where it was found to target the *CDH1* inhibitor *EZH2*, resulting in a decrease in E-cadherin protein (Carvalho J *et al*., 2012). In addition, *miR-101* has been reported to target another *CDH1* inhibitor, *ZEB1*, in other cancer types, including ovarian carcinoma (Guo F *et al*., 2014), and oral squamous-cell carcinoma (Wu B *et al*., 2016).
We thereby performed a differential expression analysis starting with these two miRNAs in the 21 paired normal and tumor tissue samples of the IGC patients by

RT-qPCR. While we found no difference in the expression levels of *miR-92a* in tumors compared to the normal gastric mucosa, we did observe a significant decrease of *miR-101* (P=1.565x10^{-05}), which is in agreement with the aforementioned reports. Nonetheless, when we attempted to independently correlate the expression levels of the two miRNAs with those of *CDH1*, it seemed that both had a negligible effect, if any, on its expression.
Based on these results, it would seem like in our IGC samples, *miR-92a* does not play a significant role, neither in regulating *CDH1*, nor in contributing to gastric carcinogenesis. Meanwhile, the decrease of *miR-101* in tumors might be promoting the carcinogenic process in ways that surpass *CDH1* regulation. Indeed, *miR-101* was reported to target both COX-2 (He XP *et al*., 2012) and mTOR (Riquelme I *et al*., 2016), and to inversely correlate with their expression in gastric cancer. COX-2 (Cyclooxygenase-2) has been found to be overexpressed in GC, where it plays a role in tumor progression by promoting cellular proliferation and apoptosis-evasion (He XP *et al*., 2012). Meanwhile, multiple studies have implicated the activation of the PI3K/AKT/mTOR pathway in the deregulation of critical cell processes such as cell growth, proliferation and angiogenesis in GC (Tapia O *et al*., 2014; Ying J *et al*., 2015).

Finally, we attempted to associate the expression levels of *CDH1*, and the 2 miRNAs, as well as the presence/absence of *CDH1a*, with the IGC patients' clinical parameters. We found a suggestive association between a more effective decrease in *CDH1* expression and a higher tumor grade. In keeping with this finding, a decrease of E-cadherin protein levels has been associated with higher tumor grading in IGC (Stănculescu D *et al*., 2011). In addition, we observed a trend of expressing less *CDH1* in tumors positive for *H. pylori* infection. In this case, however, the scenario appears to be more complex. On one hand, some studies implicated *H. pylori* infection with the epigenetic silencing of the *CDH1* promoter that would ultimately downregulate *CDH1* expression, and consequently E-cadherin protein levels (Chan AO *et al*., 2003; Huang FY *et al*., 2012); on the other hand, others have reported that E-cadherin protein levels are decreased through a mechanism which is independent from transcription. Indeed, very recently, it was shown that *H. pylori* itself secretes a protease (HtrA: high-temperature requirement A) that targets E-cadherin by directly cleaving its extracellular domain, thus opening cell-to-cell junctions (Hoy B *et al*., 2010; Schmidt TP *et al*., 2016). This E-cadherin ectodomain shedding also results in high serum levels of soluble peptides in IGC patients (Juhasz M *et al*., 2003; Schmidt TP *et al*., 2016).

On the whole, a series of "strong" genetic and epigenetic mechanisms are known to

underlie E-cadherin loss or impairment in gastric carcinogenesis, particularly of the diffuse type. However, our results support the notion that abnormal isoforms and transcripts' imbalance resulting from cryptic abnormalities along the *CDH1* locus, although subtly modulating E-cadherin expression, can still contribute to the carcinogenic process of the intestinal type.

Appendix A

Checklist for the fulfilled essential requirements of the "Minimum Information for Publication of Quantitative Real-Time dPCR Experiments" (dMIQE) guidelines

ITEM TO CHECK	COMMENTS/ WHERE?
EXPERIMENTAL DESIGN	
Definition of experimental and control groups	Materials and Methods
Number within each group	Materials and Methods
SAMPLE	
Description	Materials and Methods
Microdissection or macrodissection	Materials and Methods
Processing procedure	Materials and Methods
If frozen - how and how quickly?	Materials and Methods
Sample storage conditions and duration (especially for FFPE samples)	Materials and Methods
NUCLEIC ACID EXTRACTION	
Procedure and/or instrumentation	Materials and Methods
Name of kit and details of any modifications	Materials and Methods
Details of DNase or RNase treatment	Materials and Methods
Contamination assessment (DNA)	DNase treatment
Nucleic acid quantification	Materials and Methods
Instrument and method	Materials and Methods
RNA integrity method/instrument	Materials and Methods; Results-Figure 1
Inhibition testing (Ct dilutions, spike or other)	Results-Figure 2
REVERSE TRANSCRIPTION	
Complete reaction conditions	Materials and Methods
Amount of RNA and reaction volume	Materials and Methods
Reverse transcriptase and concentration	Materials and Methods; Manufacturer's proprietary
Temperature and time	Materials and Methods

dPCR TARGET INFORMATION	
Sequence accession number	Materials and Methods-Table 1
Amplicon length	Materials and Methods-Table 1
In silico specificity screen (BLAST, etc.)	Materials and Methods
Location of each primer by exon or intron	Materials and Methods-Table 1
What splice variants are targeted?	Materials and Methods-Table 1
dPCR OLIGONUCLEOTIDES	
Primer sequences	Materials and Methods-Table 1
Probe sequences	Materials and Methods-Table 1
Location and identity of any modifications	Materials and Methods-Table 1
dPCR PROTOCOL	
Complete reaction conditions	Materials and Methods
Reaction volume and amount of cDNA/DNA	Materials and Methods
Primer, (probe), Mg++ and dNTP concentrations	Materials and Methods; Manufacturers' proprietary
Polymerase identity and concentration	Materials and Methods
Buffer/kit identity and manufacturer	Materials and Methods
Additives (SYBR Green I, DMSO, etc.)	None
Complete thermocycling parameters	Materials and Methods
Partition number	20,000
Individual partition volume	755 pL
Total volume of the partitions measured (effective reaction size)	Number of partitions multiplied by 755 pL
Comprehensive details and appropriate use of controls	Materials and Methods; Results-Figure 4
Manufacturer of dPCR instrument	Materials and Methods
dPCR VALIDATION	
If multiplexing, comparison with singleplex assays	Results

DATA ANALYSIS	
dPCR analysis program (source, version)	Materials and Methods
Outlier identification and disposition	No outliers identified
Results of no template controls (NTCs)	Results-Figure 4
Examples of positive(s) and negative experimental results as supplemental data	Results-Figure 4
Justification of number and choice of reference genes	Materials and Methods
Description of normalization method	Materials and Methods
Number and stage (RT or dPCR) of technical replicates	Results
Repeatability (intra-assay variation)	Results
Statistical methods for result significance	Materials and Methods
Software (source, version)	Materials and Methods

References

Aberle H, Schwartz H, Kemler R. Cadherin-catenin complex: protein interactions and their implications for cadherin function. J Cell Biochem. 1996;61:514-523.

Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR. A method and server for predicting damaging missense mutations. Nat Methods. 2010;7:248-249.

Agarwal V, Bell GW, Nam J, Bartel DP. Predicting effective microRNA target sites in mammalian mRNAs. Elife. 2015;4.

Aktary Z, Pasdar M. Plakoglobin: role in tumorigenesis and metastasis. Int J Cell Biol. 2012;2012:189521.

Albano F, Zagaria A, Anelli L, Coccaro N, Tota G, Brunetti C, Minervini CF, Impera L, Minervini A, Cellamare A, Orsini P, Cumbo C, Casieri P, Specchia G. Absolute quantification of the pretreatment *PML-RARA* transcript defines the relapse risk in acute promyelocytic leukemia. Oncotarget. 2015;6:13269-13277.

Ali SM, Sanford EM, Klempner SJ, Rubinson DA, Wang K, Palma NA, Chmielecki J, Yelensky R, Palmer GA, Morosini D, Lipson D, Catenacci DV, Braiteh F, Erlich R, Stephens PJ, *et al.* Prospective comprehensive genomic profiling of advanced gastric carcinoma cases reveals frequent clinically relevant genomic alterations and new routes for targeted therapies. Oncologist. 2015;20:499-507.

Alotaibi H, Basilicata MF, Shehwana H, Kosowan T, Schreck I, Braeutigam C, Konu O, Brabletz T, Stemmler MP. Enhancer cooperativity as a novel mechanism underlying the transcriptional regulation of E-cadherin during mesenchymal to epithelial transition. Biochim Biophys Acta. 2015;1849:731-742.

Amieva M, Peek RM Jr. Pathobiology of Helicobacter pylori-Induced Gastric Cancer. Gastroenterology. 2016;150:64-78.

Amieva MR, El-Omar EM. Host-bacterial interactions in Helicobacter pylori infection. Gastroenterology. 2008;134:306-323.

Ang TL, Fock KM. Clinical epidemiology of gastric cancer Singapore Med J. 2014;55:621-8.

Ascaño JJ, Frierson H Jr, Moskaluk CA, Harper JC, Roviello F, Jackson CE, El-Rifai W, Vindigni C, Tosi P, Powell SM. Inactivation of the E-cadherin gene in sporadic diffuse-type gastric cancer. Mod Pathol. 2001;14:942-949.

Baccarelli A, Hou L, Chen J, Lissowska J, El-Omar EM, Grillo P, Giacomini SM, Yaeger M, Bernig T, Zatonski W, Fraumeni JF Jr, Chanock SJ, Chow WH. Mannose-binding lectin-2 genetic variation and stomach cancer risk. Int J Cancer. 2006;119:1970-1975.

Bano G, Siedel V, Beharry N, Wilson P, Cranston T, Hodgson S. A complex endocrine conundrum. Fam Cancer. 2013;12:577-580.

Barber M, Fitzgerald RC, Caldas C. Familial gastric cancer - aetiology and pathogenesis. Best Pract Res Clin Gastroenterol. 2006;20:721-734.

Bartel DP. MicroRNAs: genomics, biogenesis, mechanism, and function. Cell. 2004;116:281-297.

Benusiglio PR, Colas C, Rouleau E, Uhrhammer N, Romero P, Remenieras A, Moretta J, Wang Q, De Pauw A, Buecher B, Stoppa-Lyonnet D, Mouret-Fourme E, Noguès C, Di Maria M, Tlemsani C, *et al.* Hereditary diffuse gastric cancer syndrome: improved performances of the 2015 testing criteria for the identification of probands with a *CDH1* germline mutation. J Med Genet. 2015;52:563-565.

Berger AH, Knudson AG, Pandolfi PP. A continuum model for tumour suppression. Nature. 2011;476:163-169.

Berlth F, Bollschweiler E, Drebber U, Hoelscher AH, Moenig S. Pathohistological classification systems in gastric cancer: diagnostic relevance and prognostic value. World J Gastroenterol. 2014;20:5679-84.

Bertuccio P, Rosato V, Andreano A, Ferraroni M, Decarli A, Edefonti V, La Vecchia C. Dietary patterns and gastric cancer risk: a systematic review and meta-analysis. Cancer. Ann Oncol. 2013;24:1450-1458.

Berx G, van Roy F. Involvement of members of the cadherin superfamily in cancer. Cold Spring Harb Perspect Biol. 2009;1:a003129.

Biamonti G, Catillo M, Pignataro D, Montecucco A, Ghigna C. The alternative splicing side of cancer. Semin Cell Dev Biol. 2014;32:30-36.

Borg A, Sandberg T, Nilsson K, Johannsson O, Klinker M, Måsbäck A, Westerdahl J, Olsson H, Ingvar C. High frequency of multiple melanomas and breast and pancreas carcinomas in *CDKN2A* mutation-positive melanoma families. J Natl Cancer Inst. 2000;92:1260-1266.

Brooks-Wilson AR, Kaurah P, Suriano G, Leach S, Senz J, Grehan N, Butterfield YS, Jeyes J, Schinas J, Bacani J, Kelsey M, Ferreira P, MacGillivray B, MacLeod P,

Micek M, *et al.* Germline E-cadherin mutations in hereditary diffuse gastric cancer: assessment of 42 new families and review of genetic screening criteria. J Med Genet. 2004;41:508-517.

Caldas C, Carneiro F, Lynch HT, Yokota J, Wiesner GL, Powell SM, Lewis FR, Huntsman DG, Pharoah PD, Jankowski JA, MacLeod P, Vogelsang H, Keller G, Park KG, Richards FM, *et al*. Familial gastric cancer: Overview and guidelines for management. J Med Genet. 1999;36:873-880.

Cancer Genome Atlas Research Network. Comprehensive molecular characterization of gastric adenocarcinoma. Nature. 2014;513:202-209.

Cao Z, Song J, Wang J, Guo X, Yu S, Dong W. Association between polymorphisms in *XRCC1* gene and treatment outcomes of patients with advanced gastric cancer: a systematic review and meta-analysis. PLoS One. 2014;9:e85357.

Capriotti E, Calabrese R, Casadio R. Predicting the insurgence of human genetic diseases associated to single point protein mutations with support vector machines and evolutionary information. Bioinformatics. 2006;22:2729-2734.

Carneiro F. Hereditary gastric cancer. Pathologe. 2012;33:231-234.

Carneiro P, Fernandes MS, Figueiredo J, Caldeira J, Carvalho J, Pinheiro H, Leite M, Melo S, Oliveira P, Simões-Correia J, Oliveira MJ, Carneiro F, Figueiredo C, Paredes J, Oliveira C, Seruca R. E-cadherin dysfunction in gastric cancer--cellular consequences, clinical applications and open questions. FEBS Lett. 2012;586:2981-2989.

Carvalho J, van Grieken NC, Pereira PM, Sousa S, Tijssen M, Buffart TE, Diosdado B, Grabsch H, Santos MA, Meijer G, Seruca R, Carvalho B, Oliveira C. Lack of microRNA-101 causes E-cadherin functional deregulation through EZH2 up-regulation in intestinal gastric cancer. J Pathol. 2012;228:31-44.

Chan AO, Lam SK, Wong BC, Wong WM, Yuen MF, Yeung YH, Hui WM, Rashid A, Kwong YL. Promoter methylation of E-cadherin gene in gastric mucosa associated with Helicobacter pylori infection and in gastric cancer. Gut. 2003;52:502-506.

Chang L, Guo F, Huo B, Lv Y, Wang Y, Liu W. Expression and clinical significance of the microRNA-200 family in gastric cancer. Oncol Lett. 2015;9:2317-2324.

Chen XZ, Chen H, Castro FA, Hu JK, Brenner H. Epstein-Barr virus infection and gastric cancer: a systematic review. Medicine (Baltimore). 2015;94:e792.

Chen ZL, Zhao XH, Wang JW, Li BZ, Wang Z, Sun J, Tan FW, Ding DP, Xu XH,

Zhou F, Tan XG, Hang J, Shi SS, Feng XL, He J. microRNA-92a promotes lymph node metastasis of human esophageal squamous cell carcinoma via E-cadherin. J Biol Chem. 2011;286:10725-10734.

Cheng J, Randall A, Baldi P. Prediction of Protein Stability Changes for Single-Site Mutations Using Support Vector Machines. Proteins. 2006;62:1125-1132.

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

Chou CH, Chang NW, Shrestha S, Hsu SD, Lin YL, Lee WH, Yang CD, Hong HC, Wei TY, Tu SJ, Tsai TR, Ho SY, Jian TY, Wu HY, Chen PR *et al.* miRTarBase 2016: updates to the experimentally validated miRNA-target interactions database. Nucleic Acids Res. 2016;44:D239-247.

Conte D, Verri C, Borzi C, Suatoni P, Pastorino U, Sozzi G, Fortunato O. Novel method to detect microRNAs using chip-based QuantStudio 3D digital PCR. BMC Genomics. 2015;16:849.

Corso G, Carvalho J, Marrelli D, Vindigni C, Carvalho B, Seruca R, Roviello F, Oliveira C. Somatic mutations and deletions of the E-cadherin gene predict poor survival of patients with gastric cancer. J Clin Oncol. 2013;31:868-875.

Corso S, Giordano S. How Can Gastric Cancer Molecular Profiling Guide Future Therapies? Trends Mol Med. 2016;22:534-544.

Cristescu R, Lee J, Nebozhyn M, Kim KM, Ting JC, Wong SS, Liu J, Yue YG, Wang J, Yu K, Ye XS, Do IG, Liu S, Gong L, Fu J, *et al.* Molecular analysis of gastric cancer identifies subtypes associated with distinct clinical outcomes. Nat Med. 2015;21:449-456.

Cui Y, Chen J, He Z, Xiao Y. SUZ12 depletion suppresses the proliferation of gastric cancer cells. Cell Physiol Biochem. 2013;31:778-784.

De Craene B, Berx G. Regulatory networks defining EMT during cancer initiation and progression. Nat Rev Cancer. 2013;13:97-110.

D'Elia L, Rossi G, Ippolito R, Cappuccio FP, Strazzullo P. Habitual salt intake and risk of gastric cancer: a meta-analysis of prospective studies. Clin Nutr. 2012;31:489-498.

den Hoed CM, Kuipers EJ. Gastric Cancer: How Can We Reduce the Incidence of this Disease? Curr Gastroenterol Rep. 2016;18:34.

Desai R, Sarpal R, Ishiyama N, Pellikka M, Ikura M, Tepass U. Monomeric α-catenin

links cadherin to the actin cytoskeleton. Nat Cell Biol. 2013;15:261-273.

Donner I, Kiviluoto T, Ristimäki A, Aaltonen LA, Vahteristo P. Exome sequencing reveals three novel candidate predisposition genes for diffuse gastric cancer. Fam Cancer. 2015;14:241-246.

Du W, Liu X, Fan G, Zhao X, Sun Y, Wang T, Zhao R, Wang G, Zhao C, Zhu Y, Ye F, Jin X, Zhang F, Zhong Z, Li X. From cell membrane to the nucleus: an emerging role of E-cadherin in gene transcriptional regulation. J Cell Mol Med. 2014;18:1712-1719.

Edge SB, Byrd DR, Compton CC, Fritz AG, Greene FL, Trotti A, editors. AJCC cancer staging manual. 7th ed. New York, NY: Springer, 2010.

Elisei R, Molinaro E, Agate L, Bottici V, Viola D, Biagini A, Matrone A, Tacito A, Ciampi R, Vivaldi A and Romei C. Ret Oncogene and Thyroid Carcinoma. J Genet Syndr Gene Ther 2014;5:214.

El-Omar EM, Rabkin CS, Gammon MD, Vaughan TL, Risch HA, Schoenberg JB, Stanford JL, Mayne ST, Goedert J, Blot WJ, Fraumeni JF Jr, Chow WH. Increased risk of noncardia gastric cancer associated with proinflammatory cytokine gene polymorphisms. Gastroenterology. 2003;124:1193-1201.

Fassan M, Simbolo M, Bria E, Mafficini A, Pilotto S, Capelli P, Bencivenga M, Pecori S, Luchini C, Neves D, Turri G, Vicentini C, Montagna L, Tomezzoli A, Tortora G, *et al.* High-throughput mutation profiling identifies novel molecular dysregulation in high-grade intraepithelial neoplasia and early gastric cancers. Gastric Cancer. 2014;17:442-449.

Feng YB, Fan DQ, Yu J, Bie YK. Association between *XPG* gene polymorphisms and development of gastric cancer risk in a Chinese population. Genet Mol Res. 2016;15.

Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, Parkin DM, Forman D, Bray F. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. Int J Cancer. 2015;136:E359-386.

Ferlay J, Steliarova-Foucher E, Lortet-Tieulent J, Rosso S, Coebergh JW, Comber H, Forman D, Bray F. Cancer incidence and mortality patterns in Europe: estimates for 40 countries in 2012. Eur J Cancer. 2013;49:1374-1403.

Ferraz MA, Zabaglia LM, Pereira WN, Orcini WA, de Labio RW, Caleman Neto A, Wisnieski F, Calcagno DQ, Santos LC, Assumpção PP, Burbano RR, Artigiani R, Smith Mde A, Payão SL, Rasmussen LT. Downregulated Expression of E-cadherin and TP53 in Patients with Gastric Diseases: the Involvement of *H. pylori*

Infection and Its Virulence Markers. J Gastrointest Cancer. 2016;47:20-26.

Fichtner D, Lorenz B, Engin S, Deichmann C, Oelkers M, Janshoff A, Menke A, Wedlich D, Franz CM. Covalent and density-controlled surface immobilization of E-cadherin for adhesion force spectroscopy. PLoS One. 2014;9:e93123.

Figueiredo C, Machado JC, Pharoah P, Seruca R, Sousa S, Carvalho R, Capelinha AF, Quint W, Caldas C, van Doorn LJ, Carneiro F, Sobrinho-Simões M. Helicobacter pylori and interleukin 1 genotyping: an opportunity to identify high-risk individuals for gastric carcinoma. J Natl Cancer Inst. 2002;94:1680-1687.

Fitzgerald RC, Hardwick R, Huntsman D, Carneiro F, Guilford P, Blair V, Chung DC, Norton J, Ragunath K, Van Krieken JH, Dwerryhouse S, Caldas C; International Gastric Cancer Linkage Consortium. Hereditary diffuse gastric cancer: updated consensus guidelines for clinical management and directions for future research. J Med Genet. 2010;47:436-444.

Frebourg T, Oliveira C, Hochain P, Karam R, Manouvrier S, Graziadio C, Vekemans M, Hartmann A, Baert-Desurmont S, Alexandre C, Lejeune Dumoulin S, Marroni C, Martin C, Castedo S, Lovett M, *et al.* Cleft lip/palate and *CDH1*/E-cadherin mutations in families with hereditary diffuse gastric cancer. J Med Genet. 2006;43:138-142.

Friedman RC, Farh KK, Burge CB, Bartel DP. Most Mammalian mRNAs Are Conserved Targets of MicroRNAs. Genome Res. 2009;19:92-105.

Gaston D, Hansford S, Oliveira C, Nightingale M, Pinheiro H, Macgillivray C, Kaurah P, Rideout AL, Steele P, Soares G, Huang WY, Whitehouse S, Blowers S, LeBlanc MA, Jiang H, *et al.* Germline mutations in *MAP3K6* are associated with familial gastric cancer. PLoS Genet. 2014;10:e1004669.

Gayther SA, Gorringe KL, Ramus SJ, Huntsman D, Roviello F, Grehan N, Machado JC, Pinto E, Seruca R, Halling K, MacLeod P, Powell SM, Jackson CE, Ponder BA, Caldas C. Identification of germ-line E-cadherin mutations in gastric cancer families of European origin. Cancer Res. 1998;58:4086-4089.

Ghiorzo P, Gargiulo S, Nasti S, Pastorino L, Battistuzzi L, Bruno W, Bonelli L, Taveggia P, Pugliese V, Borgonovo G, Mastracci L, Fornarini G, Romagnoli P, Iiritano E, Savarino V, Bianchi-Scarrà G. Predicting the risk of pancreatic cancer: on *CDKN2A* mutations in the melanoma-pancreatic cancer syndrome in Italy. J Clin Oncol. 2007;25:5336-5337.

Gill AJ. Succinate dehydrogenase (SDH) and mitochondrial driven neoplasia. Pathology. 2012;44:285-292.

Gomceli I, Demiriz B, Tez M. Gastric carcinogenesis. World J Gastroenterol. 2012;18:5164-5170.

Grabowska MM, Day ML. Soluble E-cadherin: more than a symptom of disease. Front Biosci. (Landmark Ed) 2012;17:1948-1964.

Grady WM, Willis J, Guilford PJ, Dunbier AK, Toro TT, Lynch H, Wiesner G, Ferguson K, Eng C, Park JG, Kim SJ, Markowitz S. Methylation of the *CDH1* promoter as the second genetic hit in hereditary diffuse gastric cancer. Nat Genet. 2000;26:16-17.

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

Guilford P, Humar B, Blair V. Hereditary diffuse gastric cancer: translation of *CDH1* germline mutations into clinical practice. Gastric Cancer. 2010;13:1-10.

Guilford PJ, Hopkins JB, Grady WM, Markowitz SD, Willis J, Lynch H, Rajput A, Wiesner GL, Lindor NM, Burgart LJ, Toro TT, Lee D, Limacher JM, Shaw DW, Findlay MP, *et al*. E-cadherin germline mutations define an inherited cancer syndrome dominated by diffuse gastric cancer. Hum Mutat. 1999;14:249-255.

Guimarães AC, Lima EM, Khayat AS, Girão Faria MH, Barem Rabenhorst SH, Pitombeira MV, Assumpção PP, de Oliveira Bahia M, Lima de Lima PD, de Arruda Cardoso Smith M, Burbano RR. Interrelationships among chromosome aneuploidy, promoter hypermethylation, and protein expression of the *CDKN2A* gene in individuals from northern Brazil with gastric adenocarcinoma. Cancer Genet Cytogenet. 2007;179:45-51.

Guo F, Cogdell D, Hu L, Yang D, Sood AK, Xue F, Zhang W. MiR-101 suppresses the epithelial-to-mesenchymal transition by targeting ZEB1 and ZEB2 in ovarian carcinoma. Oncol Rep. 2014;31:2021-2028.

Gurzu S, Jung I, Orlowska J, Sugimura H, Kadar Z, Turdean S, Bara T Jr. Hereditary diffuse gastric cancer--An overview. Pathol Res Pract. 2015;211:629-632.

Hansford S, Kaurah P, Li-Chang H, Woo M, Senz J, Pinheiro H, Schrader KA, Schaeffer DF, Shumansky K, Zogopoulos G, Almeida Santos T, Claro I, Carvalho J, Nielsen C, Padilla S, *et al.* Hereditary diffuse gastric cancer syndrome: *CDH1* mutations and beyond. JAMA Oncol. 2015;1:23-32.

Hayakawa Y, Sethi N, Sepulveda AR, Bass AJ, Wang TC. Oesophageal adenocarcinoma and gastric cancer: should we mind the gap? Nat Rev Cancer.

2016;16:305-318.

He XP, Shao Y, Li XL, Xu W, Chen GS, Sun HH, Xu HC, Xu X, Tang D, Zheng XF, Xue YP, Huang GC, Sun WH. Downregulation of miR-101 in gastric cancer correlates with cyclooxygenase-2 overexpression and tumor growth. FEBS J. 2012;279:4201-4212.

Heredia NJ, Belgrader P, Wang S, Koehler R, Regan J, Cosman AM, Saxonov S, Hindson B, Tanner SC, Brown AS, Karlin-Neumann G. Droplet Digital™ PCR quantitation of *HER2* expression in FFPE breast cancer samples. Methods. 2013;59:S20-23

Herrera V, Parsonnet J. Helicobacter pylori and gastric adenocarcinoma. Clin Microbiol Infect. 2009;15:971-976.

Hoffman BD, Yap AS. Towards a Dynamic Understanding of Cadherin-Based Mechanobiology. Trends Cell Biol. 2015;25:803-814.

Hoy B, Löwer M, Weydig C, Carra G, Tegtmeyer N, Geppert T, Schröder P, Sewald N, Backert S, Schneider G, Wessler S. Helicobacter pylori HtrA is a new secreted virulence factor that cleaves E-cadherin to disrupt intercellular adhesion. EMBO Rep. 2010;11:798-804.

Hu B, El Hajj N, Sittler S, Lammert N, Barnes R, Meloni-Ehrig A. Gastric cancer: Classification, histology and application of molecular pathology. J Gastrointest Oncol. 2012;3:251-261.

Huang FY, Chan AO, Rashid A, Wong DK, Cho CH, Yuen MF. Helicobacter pylori induces promoter methylation of E-cadherin via interleukin-1β activation of nitric oxide production in gastric cancer cells. Cancer. 2012;118:4969-4980.

Huang S, Ye H, Guo W, Dong X, Wu N, Zhang X, Huang Z. CDK4/6 inhibitor suppresses gastric cancer with *CDKN2A* mutation. Int J Clin Exp Med. 2015;8:11692-11700.

Hudler P. Challenges of deciphering gastric cancer heterogeneity. World J Gastroenterol. 2015;21:10510-10527.

Huggett JF, Foy CA, Benes V, Emslie K, Garson JA, Haynes R, Hellemans J, Kubista M, Mueller RD, Nolan T, Pfaffl MW, Shipley GL, Vandesompele J, Wittwer CT, Bustin SA. The digital MIQE guidelines: Minimum Information for Publication of Quantitative Digital PCR Experiments. Clin Chem. 2013;59:892-902.

Humar B, Blair V, Charlton A, More H, Martin I, Guilford P. E-cadherin deficiency

initiates gastric signet-ring cell carcinoma in mice and man. Cancer Res. 2009;69:2050-2056.

Humar B, Guilford P. Hereditary diffuse gastric cancer: a manifestation of lost cell polarity. Cancer Sci. 2009;100:1151-1157.

Huveneers S, Oldenburg J, Spanjaard E, van der Krogt G, Grigoriev I, Akhmanova A, Rehmann H, de Rooij J. Vinculin associates with endothelial VE-cadherin junctions to control force-dependent remodeling. J Cell Biol. 2012;196:641-652.

Ibarrola-Villava M, Llorca-Cardeñosa MJ, Tarazona N, Mongort C, Fleitas T, Perez-Fidalgo JA, Roselló S, Navarro S, Ribas G, Cervantes A. Deregulation of *ARID1A*, *CDH1*, *cMET* and *PIK3CA* and target-related microRNA expression in gastric cancer. Oncotarget. 2015;6:26935-26945.

Iida S, Akiyama Y, Nakajima T, Ichikawa W, Nihei Z, Sugihara K, Yuasa Y. Alterations and hypermethylation of the *p14(ARF)* gene in gastric cancer. Int J Cancer. 2000;87:654-658.

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;87:888-891.

Japanese Gastric Cancer Association. Japanese classification of gastric carcinoma: 3rd English edition. Gastric Cancer. 2011;14:101-112.

Jia ZF, Wang LZ, Cao XY, Wang C, Cao DH, Wu X, You LL, Jin MS, Wang YP, Zhou BS, Jiang J. *CD24* genetic variants contribute to overall survival in patients with gastric cancer. World J Gastroenterol. 2016;22:2373-2382.

Jiang C, Chen X, Alattar M, Wei J, Liu H. MicroRNAs in tumorigenesis, metastasis, diagnosis and prognosis of gastric cancer. Cancer Gene Ther. 2015;22:291-301.

Jones EG. Familial gastric cancer. N Z Med J. 1964;63:287-296.

Jou E, Rajdev L. Current and emerging therapies in unresectable and recurrent gastric cancer. World J Gastroenterol. 2016;22:4812-4823.

Juhasz M, Ebert MP, Schulz HU, Röcken C, Molnar B, Tulassay Z, Malfertheiner P. Dual role of serum soluble E-cadherin as a biological marker of metastatic development in gastric cancer. Scand J Gastroenterol. 2003;38:850-855.

Juzėnas S, Saltenienė V, Kupcinskas J, Link A, Kiudelis G, Jonaitis L, Jarmalaite S, Kupcinskas L, Malfertheiner P, Skieceviciene J. Analysis of Deregulated

microRNAs and Their Target Genes in Gastric Cancer. PLoS One. 2015;10:e0132327. Erratum in: PLoS One. 2015;10:e0135762.

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, *et al.* Founder and recurrent *CDH1* mutations in families with hereditary diffuse gastric cancer. JAMA. 2007;297:2360-2372.

Keller G, Vogelsang H, Becker I, Hutter J, Ott K, Candidus S, Grundei T, Becker KF, Mueller J, Siewert JR, Höfler H. Diffuse type gastric and lobular breast carcinoma in a familial gastric cancer patient with an E-cadherin germline mutation. Am J Pathol. 1999;155:337-342.

Keller G, Vogelsang H, Becker I, Plaschke S, Ott K, Suriano G, Mateus AR, Seruca R, Biedermann K, Huntsman D, Döring C, Holinski-Feder E, Neutzling A, Siewert JR, Höfler H. Germline mutations of the E-cadherin(*CDH1*) and *TP53* genes, rather than of *RUNX3* and *HPP1*, contribute to genetic predisposition in German gastric cancer patients. J Med Genet. 2004;41:e89.

Knudson AG Jr. Mutation and cancer: statistical study of retinoblastoma. Proc Natl Acad Sci U S A. 1971;68:820-823.

Kodaman N, Sobota RS, Mera R, Schneider BG, Williams SM. Disrupted human-pathogen co-evolution: a model for disease. Front Genet. 2014;5:290.

Korpal M, Lee ES, Hu G, Kang Y. The miR-200 family inhibits epithelial-mesenchymal transition and cancer cell migration by direct targeting of E-cadherin transcriptional repressors *ZEB1* and *ZEB2*. J Biol Chem. 2008;283:14910-14914.

Kouvaraki MA, Shapiro SE, Perrier ND, Cote GJ, Gagel RF, Hoff AO, Sherman SI, Lee JE, Evans DB. *RET* proto-oncogene: a review and update of genotype-phenotype correlations in hereditary medullary thyroid cancer and associated endocrine tumors. Thyroid. 2005;15:531-544.

Ladeiras-Lopes R, Pereira AK, Nogueira A, Pinheiro-Torres T, Pinto I, Santos-Pereira R, Lunet N. Smoking and gastric cancer: systematic review and meta-analysis of cohort studies. Cancer Causes Control. 2008;19:689-701.

Lamouille S, Xu J, Derynck R. Molecular mechanisms of epithelial-mesenchymal transition. Nat Rev Mol Cell Biol. 2014;15:178-196.

Laurén 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;64:31-49.

Leckband DE, de Rooij J. Cadherin adhesion and mechanotransduction. Annu Rev Cell Dev Biol. 2014;30:291-315.

Lecuit T, Yap AS. E-cadherin junctions as active mechanical integrators in tissue dynamics. Nat Cell Biol. 2015;17:533-539.

Lee JY, Kong G. Roles and epigenetic regulation of epithelial-mesenchymal transition and its transcription factors in cancer initiation and progression. Cell Mol Life Sci. 2016 [Epub ahead of print].

Lee YS, Cho YS, Lee GK, Lee S, Kim YW, Jho S, Kim HM, Hong SH, Hwang JA, Kim SY, Hong D, Choi IJ, Kim BC, Kim BC, Kim CH, *et al.* Genomic profile analysis of diffuse-type gastric cancers. Genome Biol. 2014;15:R55.

Leonardi E, Martella M, Tosatto SC, Murgia A. Identification and in silico analysis of novel von Hippel-Lindau (*VHL*) gene variants from a large population. Ann Hum Genet. 2011;75:483-496.

Lewis BP, Burge CB, Bartel DP. Conserved Seed Pairing, Often Flanked by Adenosines, Indicates that Thousands of Human Genes are MicroRNA Targets. Cell. 2005;120:15-20.

Li FQ, Xu B, Wu YJ, Yang ZL, Qian JJ. Differential microRNA expression in signet-ring cell carcinoma compared with tubular adenocarcinoma of human gastric cancer. Genet Mol Res. 2015;14:739-747.

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

Li J, Woods SL, Healey S, Beesley J, Chen X, Lee JS, Sivakumaran H, Wayte N, Nones K, Waterfall JJ, Pearson J, Patch AM, Senz J, Ferreira MA, Kaurah P, *et al.* Point Mutations in Exon 1B of *APC* Reveal Gastric Adenocarcinoma and Proximal Polyposis of the Stomach as a Familial Adenomatous Polyposis Variant Am J Hum Genet. 2016;98:830-842.

Li X, Gao Y, Pan Y, Pan Y, Wang L, Xiao N, He Q, Fan Y, Wang Y. Mutation screen and RNA analysis disclose the changed splicing of the E-cadherin transcription in gastric cancer. Fam Cancer. 2013;12:547-554.

Lin Y, Wu Z, Guo W, Li J. Gene mutations in gastric cancer: a review of recent next-generation sequencing studies. Tumour Biol. 2015;36:7385-7394.

Liu J, Deng N, Xu Q, Sun L, Tu H, Wang Z, Xing C, Yuan Y. Polymorphisms of multiple genes involved in NER pathway predict prognosis of gastric cancer.

Oncotarget. 2016 [Epub ahead of print].

Liu R, Zhao X, Liu X, Chen Z, Qiu L, Geng R, Guo W, He G, Yin J, Li J, Zhu X. Influences of *ERCC1*, *ERCC2*, *XRCC1*, *GSTP1*, *GSTT1*, and *MTHFR* polymorphisms on clinical outcomes in gastric cancer patients treated with EOF chemotherapy. Tumour Biol. 2016;37:1753-1762.

Liu X, Chu KM. E-cadherin and gastric cancer: cause, consequence, and applications. Biomed Res Int. 2014;2014:637308.

Lochhead P, El-Omar EM. Gastric cancer. Br Med Bull. 2008;85:87-100.

Lordick F, Allum W, Carneiro F, Mitry E, Tabernero J, Tan P, Van Cutsem E, van de Velde C, Cervantes A. Unmet needs and challenges in gastric cancer: the way forward. Cancer Treat Rev. 2014;40:692-700.

Lordick F, Janjigian YY. Clinical impact of tumour biology in the management of gastroesophageal cancer. Nat Rev Clin Oncol. 2016;13:348-360.

Lunet N, Lacerda-Vieira A, Barros H. Fruit and vegetables consumption and gastric cancer: a systematic review and meta-analysis of cohort studies. Nutr Cancer. 2005;53:1-10.

Lynch HT, Grady W, Suriano G, Huntsman D. Gastric cancer: new genetic developments. J Surg Oncol. 2005;90:114-33.

Ohyauchi M, Imatani A, Yonechi M, Asano N, Miura A, Iijima K, Koike T, Sekine H, Ohara S, Shimosegawa T. The polymorphism interleukin 8 −251 A/T influences the susceptibility of Helicobacter pylori related gastric diseases in the Japanese population. Gut. 2005;54:330–335.

Ma G, Dai W, Sang A, Yang X, Gao C. Upregulation of microRNA-23a/b promotes tumor progression and confers poor prognosis in patients with gastric cancer. Int J Clin Exp Pathol. 2014;7:8833-8840.

Machado JC, Oliveira C, Carvalho R, Soares P, Berx G, Caldas C, Seruca R, Carneiro F, Sobrinho-Simões M. E-cadherin gene (*CDH1*) promoter methylation as the second hit in sporadic diffuse gastric carcinoma. Oncogene. 2001;20:1525-1528.

Majewski IJ, Kluijt I, Cats A, Scerri TS, de Jong D, Kluin RJ, Hansford S, Hogervorst FB, Bosma AJ, Hofland I, Winter M, Huntsman D, Jonkers J, Bahlo M, Bernards R. An α-E-catenin (*CTNNA1*) mutation in hereditary diffuse gastric cancer. J Pathol. 2013;229:621-629.

Mangold S, Norwood SJ, Yap AS, Collins BM. The juxtamembrane domain of the E-cadherin cytoplasmic tail contributes to its interaction with Myosin VI. Bioarchitecture. 2012;2:185-188.

Maxwell KN, Wubbenhorst B, D'Andrea K, Garman B, Long JM, Powers J, Rathbun K, Stopfer JE, Zhu J, Bradbury AR, Simon MS, DeMichele A, Domchek SM, Nathanson KL. Prevalence of mutations in a panel of breast cancer susceptibility genes in *BRCA1/2*-negative patients with early-onset breast cancer. Genet Med. 2015;17:630-638.

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010;20:1297-1303.

McLean MH, El-Omar EM. Genetics of gastric cancer. Nat Rev Gastroenterol Hepatol. 2014;11:664-674.

Miettinen M, Lasota J. Succinate dehydrogenase deficient gastrointestinal stromal tumors (GISTs) - a review. Int J Biochem Cell Biol. 2014;53:514-519.

Mocellin S, Verdi D, Pooley KA, Nitti D. Genetic variation and gastric cancer risk: a field synopsis and meta-analysis. Gut. 2015;64:1209-1219.

Mohamet L, Hawkins K, Ward CM. Loss of function of e-cadherin in embryonic stem cells and the relevance to models of tumorigenesis. J Oncol. 2011;2011:352616.

Molinaro V, Pensotti V, Marabelli M, Feroce I, Barile M, Pozzi S, Laghi L, Serrano D, Bernard L, Bonanni B, Ranzani GN. Complementary molecular approaches reveal heterogeneous *CDH1* germline defects in Italian patients with hereditary diffuse gastric cancer (HDGC) syndrome. Genes Chromosomes Cancer. 2014;53:432-445.

Monahan KJ, Hopkins L. Diagnosis and Management of Hereditary Gastric Cancer. Recent Results Cancer Res. 2016;205:45-60.

Moran A, O'Hara C, Khan S, Shack L, Woodward E, Maher ER, Lalloo F, Evans DG. Risk of cancer other than breast or ovarian in individuals with *BRCA1* and *BRCA2* mutations. Fam Cancer. 2012;11:235-242.

Nagar B, Overduin M, Ikura M, Rini JM. Structural basis of calcium-induced E-cadherin rigidification and dimerization. Nature. 1996;380:360-364.

Nagini S. Carcinoma of the stomach: A review of epidemiology, pathogenesis, molecular genetics and chemoprevention. World J Gastrointest Oncol. 2012;4:156-169.

Ng PC, Henikoff S. Predicting deleterious amino acid substitutions. Genome Res. 2001;11:863-874.

Ni Y, Zbuk KM, Sadler T, Patocs A, Lobo G, Edelman E, Platzer P, Orloff MS, Waite KA, Eng C. Germline mutations and variants in the succinate dehydrogenase genes in Cowden and Cowden-like syndromes. Am J Hum Genet. 2008;83:261-268.

Nie X, Wang X, Lin Y, Yu Y, Liu W, Xie F, Wang X, Tan J, Huang Q. SNP rs1059234 in *CDKN1A* Gene Correlates with Prognosis in Resected Gastric Adenocarcinoma. Clin Lab. 2016;62:409-416.

Noren NK, Liu BP, Burridge K, Kreft B. p120 catenin regulates the actin cytoskeleton via Rho family GTPases. J Cell Biol. 2000;150:567-580.

Oliveira C, de Bruin J, Nabais S, Ligtenberg M, Moutinho C, Nagengast FM, Seruca R, van Krieken H, Carneiro F. Intragenic deletion of *CDH1* as the inactivating mechanism of the wild-type allele in an HDGC tumour. Oncogene. 2004;23:2236-2240.

Oliveira C, Pinheiro H, Figueiredo J, Seruca R, Carneiro F. Familial gastric cancer: genetic susceptibility, pathology, and implications for management. Lancet Oncol. 2015;16:e60-70.

Oliveira C, Senz J, Kaurah P, Pinheiro H, Sanges R, Haegert A, Corso G, Schouten J, Fitzgerald R, Vogelsang H, Keller G, Dwerryhouse S, Grimmer D, Chin SF, Yang H et al. Germline *CDH1* deletions in hereditary diffuse gastric cancer families. Hum Mol Genet. 2009;18:1545-1555. (a)

Oliveira C, Seruca R, Carneiro F. Genetics, pathology, and clinics of familial gastric cancer. Int J Surg Pathol. 2006;14:21-33.

Oliveira C, Seruca R, Hoogerbrugge N, Ligtenberg M, Carneiro F. Clinical utility gene card for: Hereditary diffuse gastric cancer (HDGC). Eur J Hum Genet. 2013;21.

Oliveira C, Sousa S, Pinheiro H, Karam R, Bordeira-Carriço R, Senz J, Kaurah P, Carvalho J, Pereira R, Gusmão L, Wen X, Cipriano MA, Yokota J, Carneiro F, Huntsman D, Seruca R. Quantification of epigenetic and genetic 2nd hits in *CDH1* during hereditary diffuse gastric cancer syndrome progression. Gastroenterology. 2009;136:2137-2148. (b)

Oliveira P, Sanges R, Huntsman D, Stupka E, Oliveira C. Characterization of the intronic portion of cadherin superfamily members, common cancer orchestrators. Eur J Hum Genet. 2012;20:878-883.

Orsulic S, Huber O, Aberle H, Arnold S, Kemler R. E-cadherin binding prevents beta-catenin nuclear localization and beta-catenin/LEF-1-mediated transactivation. J Cell Sci. 1999;112:1237-1245.

Pal S, Gupta R, Davuluri RV. Alternative transcription and alternative splicing in cancer. Pharmacol Ther. 2012;136:283-294.

Paredes J, Figueiredo J, Albergaria A, Oliveira P, Carvalho J, Ribeiro AS, Caldeira J, Costa AM, Simões-Correia J, Oliveira MJ, Pinheiro H, Pinho SS, Mateus R, Reis CA, Leite M, *et al.* Epithelial E- and P-cadherins: role and clinical significance in cancer. Biochim Biophys Acta. 2012;1826:297-311.

Parkin DM, Stjernswärd J, Muir CS. Estimates of the worldwide frequency of twelve major cancers. Bull WHO. 1984;62:163-182.

Peinado H, Olmeda D, Cano A. Snail, Zeb and bHLH factors in tumour progression: an alliance against the epithelial phenotype? Nat Rev Cancer. 2007;7:415-428.

Petersen GM. Familial Pancreatic Adenocarcinoma. Hematol Oncol Clin North Am. 2015;29:641-653.

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* downregulation and hereditary diffuse gastric cancer. Hum Mol Genet. 2010;19:943-952.

Pinheiro H, Carvalho J, Oliveira P, Ferreira D, Pinto MT, Osório H, Licastro D, Bordeira-Carriço R, Jordan P, Lazarevic D, Sanges R, Stupka E, Huntsman D, Seruca R, Oliveira C. Transcription initiation arising from E-cadherin/*CDH1* intron2: a novel protein isoform that increases gastric cancer cell invasion and angiogenesis. Hum Mol Genet. 2012;21:4253-4269.

Pinheiro H, Oliveira C, Seruca R, Carneiro F. Hereditary diffuse gastric cancer - pathophysiology and clinical management Best Pract Res Clin Gastroenterol. 2014;28:1055-1068.

Polkowski W, van Sandick JW, Offerhaus GJ, ten Kate FJ, Mulder J, Obertop H, van Lanschot JJ. Prognostic value of Laurén classification and c-erbB-2 oncogene overexpression in adenocarcinoma of the esophagus and gastroesophageal junction. Ann Surg Oncol. 1999;6:290-297.

Qian X, Karpova T, Sheppard AM, McNally J, Lowy DR. E-cadherin-mediated adhesion inhibits ligand-dependent activation of diverse receptor tyrosine kinases EMBO J. 2004;23:1739-1748.

Rački N, Morisset D, Gutierrez-Aguirre I, Ravnikar M. One-step RT-droplet digital PCR: a breakthrough in the quantification of waterborne RNA viruses. Anal Bioanal Chem. 2014;406:661-667.

Ren JS, Kamangar F, Forman D, Islami F. Pickled food and risk of gastric cancer--a systematic review and meta-analysis of English and Chinese literature. Cancer Epidemiol Biomarkers Prev. 2012;21:905-915.

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

Richards FM, McKee SA, Rajpar MH, Cole TR, Evans DG, Jankowski JA, McKeown C, Sanders DS, Maher ER. Germline E-cadherin gene (*CDH1*) mutations predispose to familial gastric cancer and colorectal cancer. Hum Mol Genet. 1999;8:607-610.

Riquelme I, Tapia O, Leal P, Sandoval A, Varga MG, Letelier P, Buchegger K, Bizama C, Espinoza JA, Peek RM, Araya JC, Roa JC. miR-101-2, miR-125b-2 and miR-451a act as potential tumor suppressors in gastric cancer through regulation of the PI3K/AKT/mTOR pathway. Cell Oncol (Dordr). 2016;39:23-33.

Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. Integrative Genomics Viewer. Nature Biotechnol. 2011;29:24–26.

Rohmer V, Vidal-Trecan G, Bourdelot A, Niccoli P, Murat A, Wemeau JL, Borson-Chazot F, Schvartz C, Tabarin A, Chabre O, Chabrier G, Caron P, Rodien P, Schlumberger M, Baudin E; Groupe Français des Tumeurs Endocrines. Prognostic factors of disease-free survival after thyroidectomy in 170 young patients with a *RET* germline mutation: a multicenter study of the Groupe Francais d'Etude des Tumeurs Endocrines. J Clin Endocrinol Metab. 2011;96:E509-518.

Rosivatz E, Becker I, Specht K, Fricke E, Luber B, Busch R, Höfler H, Becker KF. Differential expression of the epithelial-mesenchymal transition regulators snail, SIP1, and twist in gastric cancer. Am J Pathol. 2002;161:1881-1891.

Salahshor S, Haixin L, Huo H, Kristensen VN, Loman N, Sjöberg-Margolin S, Borg A, Børresen-Dale AL, Vorechovsky I, Lindblom A. Low frequency of E-cadherin alterations in familial breast cancer. Breast Cancer Res. 2001;3:199-207.

Scaini MC, Minervini G, Elefanti L, Ghiorzo P, Pastorino L, Tognazzo S, Agata S, Quaggio M, Zullato D, Bianchi-Scarrà G, Montagna M, D'Andrea E, Menin C, Tosatto SC. *CDKN2A* unclassified variants in familial malignant melanoma: combining functional and computational approaches for their assessment. Hum Mutat.

2014;35:828-840.

Schmidt TP, Perna AM, Fugmann T, Böhm M, Jan Hiss, Haller S, Götz C, Tegtmeyer N, Hoy B, Rau TT, Neri D, Backert S, Schneider G, Wessler S. Identification of E-cadherin signature motifs functioning as cleavage sites for Helicobacter pylori HtrA. Sci Rep. 2016;6:23264.

Schmittgen TD, Livak KJ. Analyzing real-time PCR data by the comparative C(T) method. Nat Protoc. 2008;3:1101-1108.

Schneider MR, Kolligs FT. E-cadherin's role in development, tissue homeostasis and disease: Insights from mouse models: Tissue-specific inactivation of the adhesion protein E-cadherin in mice reveals its functions in health and disease. Bioessays. 2015;37:294-304.

Schrader KA, Masciari S, Boyd N, Salamanca C, Senz J, Saunders DN, Yorida E, Maines-Bandiera S, Kaurah P, Tung N, Robson ME, Ryan PD, Olopade OI, Domchek SM, Ford J, *et al.* Germline mutations in *CDH1* are infrequent in women with early-onset or familial lobular breast cancers. J Med Genet. 2011;48:64-68.

Schrader KA, Masciari S, Boyd N, Wiyrick S, Kaurah P, Senz J, Burke W, Lynch HT, Garber JE, Huntsman DG. Hereditary diffuse gastric cancer: association with lobular breast cancer. Fam Cancer. 2008;7:73-82.

Schuetz JM, Leach S, Kaurah P, Jeyes J, Butterfield Y, Huntsman D, Brooks-Wilson AR. Catenin family genes are not commonly mutated in hereditary diffuse gastric cancer. Cancer Epidemiol Biomarkers Prev. 2012;21:2272-2274.

Shah MA, Salo-Mullen E, Stadler Z, Ruggeri JM, Mirander M, Pristyazhnyuk Y, Zhang L. *De novo CDH1* mutation in a family presenting with early-onset diffuse gastric cancer. Clin Genet. 2012;82:283-287.

Shapiro L, Weis WI. Structure and biochemistry of cadherins and catenins. Cold Spring Harb Perspect Biol. 2009;1:a003053.

Shi Y, Tan YJ, Zeng DZ, Qian F, Yu PW. miR-203 suppression in gastric carcinoma promotes Slug-mediated cancer metastasis. Tumour Biol. 2015 [Epub ahead of print].

Shrestha S, Hsu SD, Huang WY, Huang HY, Chen W, Weng SL, Huang HD. A systematic review of microRNA expression profiling studies in human gastric cancer. Cancer Med. 2014;3:878-888.

Sijmons RH, Greenblatt MS, Genuardi M. Gene variants of unknown clinical significance in Lynch syndrome. An introduction for clinicians. Fam Cancer.

2013;12:181-187.

Song P, Wu L, Guan W. Dietary Nitrates, Nitrites, and Nitrosamines Intake and the Risk of Gastric Cancer: A Meta-Analysis. Nutrients. 2015;7:9872-9895.

Song S, Ajani JA. The role of microRNAs in cancers of the upper gastrointestinal tract. Nat Rev Gastroenterol Hepatol. 2013;10:109-118.

Sparmann A, van Lohuizen M. Polycomb silencers control cell fate, development and cancer. Nat Rev Cancer. 2006;6:846-856.

Stănculescu D, Mărgăritescu C, Stepan A, Mitruţ AO. E-cadherin in gastric carcinomas related to histological prognostic parameters. Rom J Morphol Embryol. 2011;52:1107-1112.

Stemmler MP, Hecht A, Kemler R. E-cadherin intron 2 contains *cis*-regulatory elements essential for gene expression. Development. 2005;132:965-976.

Stemmler MP, Hecht A, Kinzel B, Kemler R. Analysis of regulatory elements of E-cadherin with reporter gene constructs in transgenic mouse embryos. Dev Dyn. 2003;227:238-245.

Suárez-Arriaga MC, Ribas-Aparicio RM, Ruiz-Tachiquín ME. MicroRNAs in hereditary diffuse gastric cancer. Biomed Rep. 2016;5:151-154.

Sugimoto S, Yamada H, Takahashi M, Morohoshi Y, Yamaguchi N, Tsunoda Y, Hayashi H, Sugimura H, Komatsu H. Early-onset diffuse gastric cancer associated with a *de novo* large genomic deletion of *CDH1* gene. Gastric Cancer. 2014;17:745-749.

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, *et al.* Characterization of a recurrent germ line mutation of the E-cadherin gene: implications for genetic testing and clinical management. Clin Cancer Res. 2005;11:5401-5409.

Takahashi K, Yan IK, Kim C, Kim J, Patel T. Analysis of extracellular RNA by digital PCR. Front Oncol. 2014;4:129.

Tan P, Yeoh KG. Genetics and molecular pathogenesis of gastric adenocarcinoma. Gastroenterology. 2015;149:1153-1162.e3.

Tan RY, Ngeow J. Hereditary diffuse gastric cancer: What the clinician should know. World J Gastrointest Oncol. 2015;7:153-160.

Tapia O, Riquelme I, Leal P, Sandoval A, Aedo S, Weber H, Letelier P, Bellolio E, Villaseca M, Garcia P, Roa JC. The PI3K/AKT/mTOR pathway is activated in gastric cancer with potential prognostic and predictive significance. Virchows Arch. 2014;465:25-33.

Thoreson MA, Anastasiadis PZ, Daniel JM, Ireton RC, Wheelock MJ, Johnson KR, Hummingbird DK, Reynolds AB. Selective uncoupling of p120(ctn) from E-cadherin disrupts strong adhesion. J Cell Biol. 2000;148:189-202.

Tsao H, Chin L, Garraway LA, Fisher DE. Melanoma: from mutations to medicine. Genes Dev. 2012;26:1131-1155.

Ueda T, Volinia S, Okumura H, Shimizu M, Taccioli C, Rossi S, Alder H, Liu CG, Oue N, Yasui W, Yoshida K, Sasaki H, Nomura S, Seto Y, Kaminishi M, *et al.* Relation between microRNA expression and progression and prognosis of gastric cancer: a microRNA expression analysis. Lancet Oncol. 2010;11:136-146.

Van Cutsem E, Sagaert X, Topal B, Haustermans K, Prenen H. Gastric cancer. Lancet. 2016 [Epub ahead of print].

Van den Bossche J, Malissen B, Mantovani A, De Baetselier P, Van Ginderachter JA. Regulation and function of the E-cadherin/catenin complex in cells of the monocyte-macrophage lineage and DCs. Blood. 2012;119:1623-1633.

van der Post RS, Vogelaar IP, Carneiro F, Guilford P, Huntsman D, Hoogerbrugge N, Caldas C, Schreiber KE, Hardwick RH, Ausems MG, Bardram L, Benusiglio PR, Bisseling TM, Blair V, Bleiker E, *et al.* Hereditary diffuse gastric cancer: updated clinical guidelines with an emphasis on germline *CDH1* mutation carriers. J Med Genet. 2015;52:361-374.

van Roy F, Berx G. The cell-cell adhesion molecule E-cadherin. Cell Mol Life Sci. 2008;65:3756-88.

van Roy F. Beyond E-cadherin: roles of other cadherin superfamily members in cancer. Nat Rev Cancer. 2014;14:121-134.

Wadhwa R, Song S, Lee JS, Yao Y, Wei Q, Ajani JA. Gastric cancer-molecular and clinical dimensions. Nat Rev Clin Oncol. 2013;10:643-55.

Wang HJ, Ruan HJ, He XJ, Ma YY, Jiang XT, Xia YJ, Ye ZY, Tao HQ. MicroRNA-101 is down-regulated in gastric cancer and involved in cell migration and invasion. Eur J Cancer. 2010;46:2295-2303.

Wang K, Kan J, Yuen ST, Shi ST, Chu KM, Law S, Chan TL, Kan Z, Chan AS, Tsui

WY, Lee SP, Ho SL, Chan AK, Cheng GH, Roberts PC, *et al.* Exome sequencing identifies frequent mutation of *ARID1A* in molecular subtypes of gastric cancer. Nat Genet. 2011;43:1219-1223.

Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 2010;38:e164.

Wang K, Yuen ST, Xu J, Lee SP, Yan HH, Shi ST, Siu HC, Deng S, Chu KM, Law S, Chan KH, Chan AS, Tsui WY, Ho SL, Chan AK, *et al.* Whole-genome sequencing and comprehensive molecular profiling identify new driver mutations in gastric cancer. Nat Genet. 2014;46:573-582.

Wang W, Chen S, Brune KA, Hruban RH, Parmigiani G, Klein AP. PancPRO: risk assessment for individuals with a family history of pancreatic cancer. J Clin Oncol. 2007;25:1417-1422.

Wang X. Improving microRNA target prediction by modeling with unambiguously identified microRNA-target pairs from CLIP-ligation studies. Bioinformatics. 2016;32:1316-1322.

Wang Z, Liu C. MiR-153 regulates metastases of gastric cancer through Snail. Tumour Biol. 2015 [Epub ahead of print].

Watabe-Uchida M, Uchida N, Imamura Y, Nagafuchi A, Fujimoto K, Uemura T, Vermeulen S, van Roy F, Adamson ED, Takeichi M. alpha-Catenin-vinculin interaction functions to organize the apical junctional complex in epithelial cells. J Cell Biol. 1998;142:847-857.

Wirtz D, Konstantopoulos K, Searson PC. The physics of cancer: the role of physical interactions and mechanical forces in metastasis. Nat Rev Cancer. 2011;11:512-522.

Wong N, Wang X. miRDB: an online resource for microRNA target prediction and functional annotations. Nucleic Acids Res. 2015;43:D146-152.

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

Wu B, Lei D, Wang L, Yang X, Jia S, Yang Z, Shan C, Yang X, Zhang C, Lu B. MiRNA-101 inhibits oral squamous-cell carcinoma growth and metastasis by targeting zinc finger E-box binding homeobox 1. Am J Cancer Res. 2016;6:1396-1407.

Xiao K, Oas RG, Chiasson CM, Kowalczyk AP. Role of p120-catenin in cadherin trafficking. Biochim Biophys Acta. 2007;1773:8-16.

Xu J, Ma J, Zong HT, Wang SY, Zhou JW. Pharmacogenetic role of *XRCC1* polymorphisms on the clinical outcome of gastric cancer patients with platinum-based chemotherapy: a systematic review and meta-analysis. Genet Mol Res. 2014;13:1438-1446.

Yaghoobi M, Bijarchi R, Narod SA. Family history and the risk of gastric cancer. Br J Cancer. 2010;102:237-242.

Yakirevich E, Resnick MB. Pathology of gastric cancer and its precursor lesions. Gastroenterol Clin North Am. 2013;42:261-84.

Yanaka Y, Muramatsu T, Uetake H, Kozaki K, Inazawa J. miR-544a induces epithelial-mesenchymal transition through the activation of WNT signaling pathway in gastric cancer. Carcinogenesis. 2015;36:1363-1371.

Yang M, Kim HS, Cho MY. Different methylation profiles between intestinal and diffuse sporadic gastric carcinogenesis. Clin Res Hepatol Gastroenterol. 2014;38:613-620.

Ying J, Xu Q, Liu B, Zhang G, Chen L, Pan H. The expression of the PI3K/AKT/mTOR pathway in gastric cancer and its role in gastric cancer prognosis. Onco Targets Ther. 2015;8:2427-2433.

Yonemura S, Wada Y, Watanabe T, Nagafuchi A, Shibata M. alpha-Catenin as a tension transducer that induces adherens junction development. Nat Cell Biol. 2010;12:533-542.

Yoon KA, Ku JL, Yang HK, Kim WH, Park SY, Park JG. Germline mutations of E-cadherin gene in Korean familial gastric cancer patients. J Hum Genet. 1999;44:177-180.

Yuasa Y. Control of gut differentiation and intestinal-type gastric carcinogenesis. Nat Rev Cancer. 2003;3:592-600.

Zang ZJ, Cutcutache I, Poon SL, Zhang SL, McPherson JR, Tao J, Rajasegaran V, Heng HL, Deng N, Gan A, Lim KH, Ong CK, Huang D, Chin SY, Tan IB, *et al.* Exome sequencing of gastric adenocarcinoma identifies recurrent somatic mutations in cell adhesion and chromatin remodeling genes. Nat Genet. 2012;44:570-574.

Zhao X, He L, Li T, Lu Y, Miao Y, Liang S, Guo H, Bai M, Xie H, Luo G, Zhou L, Shen G, Guo C, Bai F, Sun S, *et al.* SRF expedites metastasis and modulates the epithelial to mesenchymal transition by regulating miR-199a-5p expression in human

gastric cancer. Cell Death Differ. 2014;21:1900-1913.

Zhou RM, Niu CX, Wang N, Liu L, Huang X, Chen ZF, Huo XR, Hao YL, Li Y. *XPG* Gene Polymorphisms and the Risk of Gastric Cardia Adenocarcinoma. Genet Test Mol Biomarkers. 2016;20:432-437.

Zhou X, Xia Y, Li L, Zhang G. MiR-101 inhibits cell growth and tumorigenesis of Helicobacter pylori related gastric cancer by repression of *SOCS2*. Cancer Biol Ther. 2015;16:160-169.

Additional contributions

During my PhD I took part in a molecular screening of a cohort of Italian patients with multiple adenomas and familial adenomatous polyposis (FAP). Germline mutations in the *APC*, *MUTYH*, *POLE* and *POLD1* disease-genes were analyzed and genotype-phenotype correlations were investigated. The paper resulting from this work, entitled "Colorectal Adenomatous Polyposis: Heterogeneity of Susceptibility Gene Mutations and Phenotypes in a Cohort of Italian Patients", is here included.

In addition, I was involved in the characterization of a FAP family that has remained genetically unexplained until now. The disease proved to be associated with a novel deletion of the promoter 1B of the *APC* gene. Fine-mapping the breakpoints allowed us to design "diagnostic" primers and to extend the gentic test to the proband's family members. Interestingly, by analyzing different branches of the proband's family tree we could demonstrate that this germline mutation shows a founder effect in Italy. The manuscript for this work, entitled "A novel *APC* promoter-1B deletion shows a founder effect in Italian patients with familial adenomatous polyposis", is currently in preparation.

List of original manuscripts

Marabelli M, Molinaro V, Abou Khouzam R, Berrino E, Panero M, Balsamo A, Venesio T, Ranzani GN. Colorectal Adenomatous Polyposis: Heterogeneity of Susceptibility Gene Mutations and Phenotypes in a Cohort of Italian Patients. Genet. Test Mol Biomarkers. 2016 [Epub ahead of print]. DOI: 10.1089/gtmb.2016.0198.

Pdf of the proof has been included

Abou Khouzam R, Molinari C, Salvi S, Marabelli M, Molinaro V, Orioli D, Saragoni L, Morgagni P, Calistri D, Ranzani GN. Digital PCR identifies changes in *CDH1* (E-cadherin) transcription pattern in intestinal-type gastric cancer. Oncotarget. *Accepted*

Marabelli M*, Gismondi V*, Ricci MT, Vetro A, Abou Khouzam R, Rea V, Zuffardi O, Varesco L, Ranzani GN. A novel *APC* promoter-1B deletion shows a founder effect in Italian patients with familial adenomatous polyposis. *In preparation*
*Equal contributors

GENETIC TESTING AND MOLECULAR BIOMARKERS
Volume 00, Number 00, 2016

Pp. 1–9
DOI: 10.1089/gtmb.2016.0198

ORIGINAL ARTICLE

Colorectal Adenomatous Polyposis: Heterogeneity of Susceptibility Gene Mutations and Phenotypes in a Cohort of Italian Patients

Monica Marabelli,[1] Valeria Molinaro,[1] Raefa Abou Khouzam,[1] Enrico Berrino,[2] Mara Panero,[2] Antonella Balsamo,[2] Tiziana Venesio,[2] and Guglielmina Nadia Ranzani[1]

Aims: Colorectal adenomatous polyposis entailing cancer predisposition is caused by constitutional mutations in different genes. *APC* is associated with the familial adenomatous polyposis (FAP/AFAP) and *MUTYH* with the *MUTYH*-associated polyposis (MAP), while *POLE* and *POLD1* mutations cause the polymerase proofreading-associated polyposis (PPAP). *Methods:* We screened for mutations in patients with multiple adenomas/FAP: 121 patients were analyzed for *APC* and *MUTYH* mutations, and 36 patients were also evaluated for *POLE* and *POLD1* gene mutations. *Results:* We found 20 FAP/AFAP, 15 MAP, and no PPAP subjects: pathogenic mutations proved to be heterogeneous, and included 5 *APC* and 1 *MUTYH* novel mutations. The mutation detection rate was significantly different between patients with 5–100 polyps and those with >100 polyps ($p = 8.154 \times 10^{-7}$), with *APC* mutations being associated with an aggressive phenotype ($p = 1.279 \times 10^{-9}$). Mean age at diagnosis was lower in FAP/AFAP compared to MAP ($p = 3.055 \times 10^{-4}$). Mutation-negative probands showed a mean age at diagnosis that was significantly higher than FAP/AFAP ($p = 3.46986 \times 10^{-7}$) and included 45.3% of patients with <30 polyps and 70.9% of patients with no family history. *Conclusions:* This study enlarges the *APC* and *MUTYH* mutational spectra, and also evaluated variants of uncertain significance, including the MUTYH p.Gln338His mutation. Moreover this study underscores the phenotypic heterogeneity and genotype–phenotype correlations in a cohort of Italian patients.

Keywords: FAP-MAP, genetic predisposition to colorectal polyposis, genotype-phenotype in colorectal polyposis

Introduction

MENDELIAN SYNDROMES EXPLAIN about 5% of the total burden of colorectal cancer (CRC), with syndromes characterized by multiple polyps representing about 1% of cases. The great majority of patients with multiple polyps are ascribable to familial adenomatous polyposis, a dominantly inherited disease that, both in its classical (familial adenomatous polyposis [FAP]) and attenuated (AFAP) forms, is associated with constitutional inactivation of *APC* tumor suppressor gene (Half *et al.*, 2009). Less common is *MUTYH*-associated polyposis (MAP), a recessive disease caused by the impairment of *MUTYH* DNA base-excision repair gene, which is estimated to only account for about 0.7% of all CRCs (Cleary *et al.*, 2009; Lubbe *et al.*, 2009; Venesio *et al.*, 2012). Other polyposis types, dominantly inherited, are much rarer, including juvenile polyposis (JPS), linked to *SMAD4* and *BMPR1A* genes; Peutz–Jeghers syndrome, linked to *STK11*; and Cowden disease, linked to *PTEN* (Tomlinson, 2015). While a clear genetic etiology is still to be defined for serrated polyposis (SPS) and hyperplastic polyposis (HPPS), complementary molecular methods, including high-throughput sequencing, have recently associated the *GREM1* gene with the dominantly inherited mixed polyposis (HMPS) (Jaeger *et al.*, 2012; Rohlin *et al.*, 2016). Mutations in other genes have been identified by whole-exome and whole-genome sequencing in a subset of patients; new genes include *NTHL1* base-excision repair gene, which causes a recessive form of predisposition to multiple adenomas (NAP from *NTHL1*-associated polyposis) (Weren *et al.*, 2015), and *POLE* and *POLD1* encoding for polymerases ε and δ, respectively. Constitutional mutations within the exonuclease domain of these proteins have been found in patients with a dominantly inherited CRC predisposition, now called polymerase proofreading-associated polyposis (PPAP) (Palles *et al.*, 2013).

The above syndromes are characterized by different polyp distribution and histology, clinical presentation, and risk of

[1]Department of Biology and Biotechnology, University of Pavia, Pavia, Italy.
[2]Candiolo Cancer Institute, FPO-IRCCS, Torino, Italy.

malignancy; accordingly, the phenotype can direct the search for the underlying genetic defect. However, there is great phenotypic variability within each genetically defined condition, including number and location of polyps with/without CRC, polyp size and appearance, cancer age of onset, and extracolonic manifestations (Syngal *et al.*, 2015). Moreover, the issue can be further complicated by the phenotypic overlapping among the different syndromes, including polyp histology itself (Lucci-Cordisco *et al.*, 2013).

In this complex scenario, differences in mutation patterns, genetic background, and environmental factors might exist among different populations, making it effectual to investigate genotype–phenotype correlations in patients from different regions. On this basis, we examined a cohort of Italian patients with multiple adenomas/FAP to evaluate the contribution of *APC*, *MUTYH*, and *POLE/POLD1* genes to CRC burden, explore the possible role of Variants of Uncertain Significance (VUSs), and compare clinical features in mutation carriers and noncarriers.

Materials and Methods

Patients and genetic tests

A total of 300 Italian unrelated patients, diagnosed with multiple adenomas/FAP, with/without CRC, were enrolled for genetic testing at the Candiolo Cancer Institute-FPO-IRCCS (Torino, Italy), through a 15-year counseling service (1997–2012). Each patient provided a written informed consent and analyses were carried out following protocols approved by the Institutional Ethics Committee.

Since complete clinical data, reliable information on family history retrieved from medical records, and comprehensive molecular analysis were available for 121 probands, we only considered these patients. Patients, with/without family history of polyposis/CRC, presented the symptoms of either classical or attenuated polyposis, with age at diagnosis ranging from 14 to 78 years. Based on endoscopic examination, the number of polyps proved to be highly variable, ranging from 5 to >1000. All polyps were histologically classified as adenomatous, serrated, or hyperplastic.

Probands' genomic DNA was extracted from blood samples by standard procedures. Patients were analyzed for either *APC* or *MUTYH*, according to the pattern of disease inheritance (dominant vs. recessive) and the clinical phenotype (very severe vs. severe/attenuated). Most patients tested negative for the first gene were analyzed for the second one.

Thirty-six selected patients were also investigated by direct sequencing for mutations in the proofreading domains of Polε and Polδ polymerases. These subjects, with either dominant (12 cases) or absent family history of polyposis, showed 15–100 polyps (with CRC in 15 cases) and were negative for both *APC* and *MUTYH* genetic tests.

Molecular analysis

APC analysis was performed by Protein Truncation Test (PTT) of exon 15 and by direct sequencing of exons 1–14 (Venesio *et al.*, 2003a, 2003b). *MUTYH* gene analysis was performed by direct sequencing of exons 1–16 (Venesio *et al.*, 2004).

As for *POLE* (NM_006231.2) and *POLD1* (NM_002691.3), we amplified and sequenced the full regions encoding for the 3′-5′ exonuclease domains of Polε and Polδ polymerases (*POLE*: codons 268–471; *POLD1*: codons 304–533). Polymerase chain reactions were carried out with annealing temperatures and primers reported in Table 1.

In silico *analysis of VUSs*

Pathogenicity of missense VUSs identified in *APC* and *MUTYH* was predicted by using eight bioinformatics tools: SIFT (http://sift.jcvi.org/); Provean (http://provean.jcvi.org/seq_submit.php); PolyPhen-2 (http://genetics.bwh.harvard.edu/pph2/); MUpro (http://mupro.proteomics.ics.uci.edu/); PhD-SNP (http://snps.biofold.org/phd-snp/phd-snp.html); Mutation Taster (www.mutationtaster.org/); SNAP2 (www.rostlab.org/services/snap/); and HANSA (www.cdfd.org.in/HANSA/).

Statistical analysis

Chi-squared test with Yates' continuity correction was used to assess differences in proportions. Shapiro–Wilk test (<30 cases) and Kolmogorov–Smirnov test (>30 cases) were used to check normality distribution of data, while Levene's

TABLE 1. POLYMERASE CHAIN REACTION ANALYSIS OF *POLE* AND *POLD1* GENES: SEQUENCE OF PRIMERS AND LENGTH OF AMPLICONS

Gene	*Exons*	*Primer (5′→3′)*	*T annealing (°C)*	*Amplicon length (bp)*
POLE	Exons 9–10	F: TGGTTAGTCTTAGGGTCCTTCT R: CCTGAGAACAAGCTCATGGA	61	828
POLE	Exons 11–12	F: GCTGCTGCTTCTGAACTTTG R: CATACTCTTGGGTGACCTGAAA	61	650
POLE	Exons 13–14	F: TGGGTGCCTGTTAGGAACTT R: AGAATTCTACTGCCGCCTCT	64	784
POLD1	Exons 7–8	F: AACAGGGCATCCGTGTGGCA R: CCTCTGTGCTCACTCCATCT	61	593
POLD1	Exons 9–10	F: AGATGGAGTGAGCACAGAGG R: CACAGAGCAAGATGCACCAA	64	838
POLD1	Exons 11–12	F: GTGTCCCCAAATCTCTTCCT R: GTCAGGTAAGGTCACAGGTC	61	786
POLD1	Exon 13	F: TATGACCTTGTCTCTGCTTTCC R: TTGGGAGTGGGGAGAAAAG	61	560

test was used to assess homogeneity of variances between groups. Student's *t*-test was used to compare the age at diagnosis (normally distributed) between FAP/AFAP and MAP patients. Kruskal–Wallis analysis, followed by Dunn's *post hoc* test with Bonferroni correction, was performed to compare the age at diagnosis in mutation-negative patients (not normally distributed) with that in FAP/AFAP and MAP patients. In each comparison, a $p < 0.05$ was considered statistically significant.

Results

Pathogenic mutations

The present analysis is focused on 121 patients with multiple adenomas/FAP. In total, we could genetically explain 35 cases out of 121 (28.9%). We found pathogenic mutations of *APC* in 20 patients (16.5%) and biallelic pathogenic mutations of *MUTYH* in 15 cases (12.4%). *APC* and *MUTYH* mutations and patients' phenotypes are reported in Tables 2 and 3, respectively.

APC mutations were nine nonsense, eight frameshift, and one small deletion, all predicted to result in a truncated protein (Table 2). To the best of our knowledge, five of these truncating mutations are novel (in bold in Table 2). The p.Arg499Ter and p.Ser1068GlyfsTer57 *APC* lesions, which are recurrent in polyposis patients (http://chromium.lovd.nl/LOVD2/colon_cancer/home.php?select_db=APC), were both found twice in this cohort.

Among *MUTYH* mutations, there were four nonsense, one frameshift, five missense, one splice site, and one small in-frame deletion (Table 3). To the best of our knowledge, the p.Arg109Gly amino acid substitution is novel (in bold in Table 3). This mutation was present in a p.Arg109Gly/p.Tyr104Ter compound heterozygous subject; by testing her healthy parents, we could demonstrate that the two mutations were inherited in trans. Due to family size, we could not verify the cosegregation of the novel mutation with polyposis/CRC. However, we considered p.Arg109Gly as pathogenic according to the following evidence: it is not reported in the 1000 Genomes Project (www.1000genomes.org/home) and in the ExAC variation databases (http://exac.broadinstitute.org), it is predicted to be damaging by 8/8 bioinformatic tools (Table 4), and the substitution of the highly conserved arginine residue at position 109 by tryptophan has already been found in polyposis patients and classified as pathogenic (http://chromium.lovd.nl/LOVD2/colon_cancer/home.php?select_db=MUTYH).

The p.Tyr179Cys and p.Gly396Asp amino acid substitutions (formerly: p.Tyr165Cys and p.Gly382Asp) are by far the most common mutations in polyposis patients and, in our survey, they were present in nine and five subjects, respectively.

None of the 36 patients we selected for *POLE/POLD1* analysis (see Materials and Methods section) proved to carry pathogenic mutations or VUSs in these genes.

Missense VUSs

Besides the 35 probands genetically explained by clear-cut pathogenic mutations, we identified the following: 2 patients with two different *APC* missense VUSs (p.Ser1631Cys and p.Ala1670Val), 2 carriers of both a single *MUTYH* pathogenic mutation and a single VUS (p.Ile223Val/p.Gln338His and p.Gln338Ter/p.Gly457Trp), and 3 heterozygous carriers of the MUTYH p.Val22Met common variant. The MUTYH p.Gln338His polymorphic variant was detected in 21 patients, 4 of whom were homozygous.

To predict the functional effect of the above VUSs, we used eight bioinformatic tools that, however, produced inconclusive results (Table 4).

Genotype–phenotype correlations and family history

By considering the 88 patients with the attenuated phenotype (5–100 polyps) and the 33 patients with a more aggressive disease (>100 polyps), mutation detection rates were 15.9% and 63.6% in first and second group, respectively ($p = 8.154 \times 10^{-7}$). Within the first group, we identified 3 carriers of *APC* mutations (3.4%) and 11 carriers of *MUTYH* biallelic mutations (12.5%). Within the second group, 4 patients were genetically explained by *MUTYH* (12.1%) and 17 by *APC* (51.5%), showing that *APC* impairment is more likely associated with an aggressive phenotype (3/88 vs. 17/33: $p = 1.279 \times 10^{-9}$) (Fig. 1).

Compared with *MUTYH*, *APC* mutations were associated not only with a higher number of polyps but also with an earlier age of disease onset. Indeed, FAP/AFAP patients showed a mean age at diagnosis of 33.4 ± 9.4 years (range: 14–51), while MAP patients showed a mean age of 47.1 ± 10.8 years (range: 31–62) ($p = 3.055 \times 10^{-4}$); in particular, patients with disease diagnosis <30 years were only ascribable to FAP/AFAP, while disease diagnosis at >55 years was limited to MAP (Fig. 2).

By considering family history (48 cases) and number of polyps, mutation detection rates were 32.1% among patients with an attenuated phenotype (28 cases) and 70% among those with a more aggressive disease (20 cases) ($p = 0.02171$) (Fig. 3).

We found six carriers of *MUTYH* biallelic mutations and six carriers of *APC* lesions among subjects without family history of polyposis/CRC. For the first gene, small family sizes limited our possibility to recognize an MAP-recessive pattern of inheritance, while the occurrence of a *de novo* mutation is the most likely explanation for *APC* mutation carriers, indicating a *de novo* mutation frequency of 30% (6/20). Two MAP patients (n. 25 and 26 in Table 3) showed a dominant family history of the disease: the father of the first patient had been diagnosed with polyposis and CRC at 53 years, while the mother of the second one developed a CRC at 44 years of age.

In our survey, 46.7% of MAP patients developed different extracolonic manifestations, including malignancies (Table 3). On the other hand, 60% of FAP/AFAP subjects displayed extracolonic manifestations belonging to the FAP/AFAP phenotypic spectrum, including gastric polyps, CHRPE, and desmoids, as well as different types of neoplasia (Table 2).

The 86 patients who tested mutation negative showed a mean age of 52.2 ± 14.1 years (range: 15–78) at diagnosis, which is higher than those of FAP/AFAP and MAP (Fig. 4), and a phenotype similar to MAP: 29% of cases developed extracolonic manifestations and 86% were characterized by an attenuated phenotype (5–100 polyps). In particular, the group of nonmutated cases included 9.3% of patients with <10 polyps and 36% with 10–30 polyps. For the great majority of patients (70.9%), there was no evidence of family history.

Table 2. Familial Adenomatous Polyposis/AFAP Patients: *APC* Pathogenic Mutations and Clinical Data

ID	APC *pathogenic mutation*	*Exon*	*Mutation type*	*Sex*	*N. polyps*	*Age at diagnosis*	*CRC at diagnosis*	*Extracolonic manifestations*	*Family history*
1	c.637C>T; p.Arg213Ter	Exon 5	Nonsense	M	60–100	22	No[a]	No	AD
2	c.745A>T; p.Lys249Ter	Exon 7	Nonsense	M	>100	51	No	No	No
3	c.847C>T; p.Arg283Ter	Exon 8	Nonsense	F	>100	28	No[a]	Benign thyroid nodules; gastric polyps	AD
4	**c.1219delC; p.Leu407PhefsTer47**	**Exon 9**	**Frameshift**	F	~40	29	No[a]	Ovarian cancer	AD
5	c.1495C>T; p.Arg499Ter	Exon 11	Nonsense	F	>100	26	No[a]	Endometrial polyps; carcinoma of the Ampulla of Vater; thyroid cancer	AD
6	c.1495C>T; p.Arg499Ter	Exon 11	Nonsense	F	>1000	27	Yes	No	AD
7	**c.1628delT; p.Ile544LeußTer5**	**Exon 13**	**Frameshift**	F	>100	42	Yes	No	No
8	c.1748_1756delCAACCCTCA; p.Ser583Ter	Exon 14	Small deletion[b]	F	~40	14	No[a]	Gastric polyps; CHRPE; odontoma	AD
9	c.2468C>A; p.Ser823Ter	Exon 15	Nonsense	F	300	34	No[a]	No	AD
10	c.2701C>T; p.Gln901Ter	Exon 15	Nonsense	M	>100	40	No	CHRPE; sebaceous cysts; gastric polyps; desmoids	AD
11	c.2771-2772insT; p.Arg924SerfsTer16	Exon 15	Frameshift	M	>100	33	Yes	Gastric polyps	No
12	c.2805C>G; p.Tyr935Ter	Exon 15	Nonsense	M	>100	23	No	No	AD
13	c.3180_3184delAAAAC; p.Gln1062Ter	Exon 15	Frameshift	F	>100	36	No[a]	No	No
14	c.3202_3205delTCAA; p.Ser1068GlyfsTer57	Exon 15	Frameshift	M	>100	42	Yes	Gastric polyps; duodenal polyps; CHRPE	AD
15	c.3202_3205delTCAA; p.Ser1068GlyfsTer57	Exon 15	Frameshift	F	>100	45	Yes	Vaginal cancer	AD
16	**c.3331G>T; p.Glu1111Ter**	**Exon 15**	**Nonsense**	M	>100	23	Yes	Gastric polyps	No
17	c.3595_3596delAA; p.Lys1199GlufsTer8	Exon 15	Frameshift	F	>100	32	No[a]	Multiple osteomas; gastric polyps; desmoids	AD
18	**c.3676_3677delAA; p.Lys1226GlufsTer13**	**Exon 15**	**Frameshift**	F	>100	40	No	Gastric polyps	AD
19	c.3769G>T; p.Glu1257Ter	Exon 15	Nonsense	F	>100	36	No	Gastric polyps	AD
20	**c.3805_3806delAT; p.Ile1269MetfsTer6**[c]	**Exon 15**	**Frameshift**	M	>100	44	Yes	No	No

Mutation nomenclature refers to NM_000038.5. Novel mutations are depicted in bold.
[a]These patients underwent prophylactic colectomy.
[b]TCA ACC CTG AAA: the deleted nucleotides are underlined; TAA is a stop codon.
[c]This mutation is novel at the DNA level, but results in the same truncated protein already reported in other patients (c.3807_3808delAT; p.Ile1269MetfsTer6).
AD, autosomal dominant; CHRPE, congenital hypertrophy of the retinal pigment epithelium; CRC, colorectal cancer; N. polyps, cumulative number of polyps.

Table 3. *MUTYH*-Associated Polyposis Patients: *MUTYH* Pathogenic Mutations and Clinical Data

ID	MUTYH *pathogenic mutations*	*Exon*	*Mutation type*	*Sex*	*N. polyps*	*Age at diagnosis*	*CRC at diagnosis*	*Extracolonic manifestations*	*Family history*
21	c.312C>A; p.Tyr104Ter	Exon 3	Nonsense	F	10	49	Yes	No	AR
	c.325C>G; p.Arg109Gly	**Exon 3**	**Missense**						
22	c.536A>G; p.Tyr179Cys	Exon 7	Missense	M	>100	41	Yes (at 58 years)	Duodenal polyps	No
	c.536A>G; p.Tyr179Cys	Exon 7	Missense						
23	c.536A>G; p.Tyr179Cys	Exon 7	Missense	F	50–100	31	No	No	No
	c.536A>G; p.Tyr179Cys	Exon 7	Missense						
24	c.536A>G; p.Tyr179Cys	Exon 7	Missense	M	>100	47	Yes	Chronic gastritis	AR
	c.536A>G; p.Tyr179Cys	Exon 7	Missense						
25	c.536A>G; p.Tyr179Cys	Exon 7	Missense	M	>100	60	Yes	No	AD
	c.667A>G; p.Ile223Val	Exon 8	Missense						
26	c.536A>G; p.Tyr179Cys	Exon 7	Missense	M	15–30	41	No	No	AD
	c.1147delC; p.Ala385ProfsTer23	Exon 12	Frameshift						
27	c.536A>G; p.Tyr179Cys	Exon 7	Missense	F	~30	45	Yes	Breast cancer	No
	c.1171 C>T; p.Gln391Ter	Exon 12	Nonsense						
28	c.536A>G; p.Tyr179Cys	Exon 7	Missense	M	30–100	52	No	No	AR
	c.1187G>A; p.Gly396Asp	Exon 13	Missense						
29	c.536A>G; p.Tyr179Cys	Exon 7	Missense	M	30–100	61	No	No	No
	c.1187G>A; p.Gly396Asp	Exon 13	Missense						
30	c.536A>G; p.Tyr179Cys	Exon 7	Missense	M	~25	55	Yes	Chronic myeloid leukemia	AR
	c.1187G>A; p.Gly396Asp	Exon 13	Missense						
31	c.545G>A; p.Arg182His	Exon 7	Missense	M	12	35	No	Chronic gastritis	AR
	c.1437_1439delGGA; p.Glu480del	Exon 14	Small in-frame deletion						
32	c.998-1G>T; p.Ala333_Ser395del	Intron 11	Splice site mutation	F	30–100	32	Yes	Duodenal polyps	No
	c.1437_1439delGGA; p.Glu480del	Exon 14	Small in-frame deletion						
33	c.1171 C>T; p.Gln391Ter	Exon 12	Nonsense	F	~25	59	No	Reactive lymphoid hyperplasia	No
	c.1187G>A; p.Gly396Asp	Exon 13	Missense						
34	c.1187G>A; p.Gly396Asp	Exon 13	Missense	M	>100	37	No	No	AR
	c.1187G>A; p.Gly396Asp	Exon 13	Missense						
35	c.1240C>T; p.Gln414Ter	Exon 13	Nonsense	F	15–30	62	Yes	No	AR
	c.1438G>T; p.Glu480Ter	Exon 14	Nonsense						

Mutation nomenclature refers to NM_001128425.1. The novel mutation is depicted in bold.
AR, autosomal recessive.

TABLE 4. *IN SILICO* ANALYSIS OF *MUTYH* AND *APC* MISSENSE MUTATIONS WITH UNKNOWN PATHOGENIC SIGNIFICANCE

Variant	*SIFT*	*Provean*	*PolyPhen-2*	*MUpro*	*PhD-SNP*	*Mutation taster*	*SNAP2*	*HANSA*
MUTYH c.325C>G; p.Arg109Gly[a]	Damaging	Damaging	Damaging	Damaging	Damaging	Damaging	Damaging	Damaging
MUTYH c.64G>A; p.Val22Met[b]	Damaging	Neutral	Neutral	Neutral	Neutral	Neutral	Damaging	Damaging
MUTYH c.1014G>C; p.Gln338His[b]	Neutral	Neutral	Neutral	Damaging	Neutral	Neutral	Neutral	Damaging
MUTYH c.1369G>T; p.Gly457Trp[a]	Neutral	Neutral	Neutral	Damaging	Neutral	Neutral	Neutral	Damaging
APC c.4891A>T; p.Ser1631Cys[a]	Damaging	Neutral	Damaging	Damaging	Neutral	Damaging	Neutral	Damaging
APC c.5009C>T; p.Ala1670Val[c]	Neutral	Neutral	Neutral	Damaging	Neutral	Damaging	Neutral	Damaging

[a]These variants are not listed in the 1000 Genomes Project (www.1000genomes.org/home) nor in the ExAC population databases (http://exac.broadinstitute.org).
[b]Polymorphic variants.
[c]Rare variant.

Discussion

In our cohort, constitutional genetic lesions proved to be heterogeneous. We identified five *APC* and one *MUTYH* novel disease-causing mutations, thus enlarging the clinical spectrum of gene alterations. We repeatedly found mutations that are known to be recurrent in the disease; these also include MUTYH p.Glu480del (formerly: 1395del GGA; c.1395_1397del, p.Glu466del), which has been reported in a number of Italian polyposis patients, possibly reflecting a higher frequency of the mutation in Southern Europe (Gismondi *et al.*, 2004; Aceto *et al.*, 2005; de Leon *et al.*, 2013). On the other hand, we did not detect c.933+3A>C (IVS10+3A>C) transversion, which has been reported to account for nearly 1/5 of mutations in MAP patients from north-eastern Italy (Pin *et al.*, 2013).

We found two MAP patients (n. 25 and n. 26) with a dominant family history of polyposis/CRC. Due to the relatively high frequency of some *MUTYH* mutations in the general population, a fraction of MAP patients is expected to have affected children with partners who carry a single mutation, thus mimicking a dominant inheritance of the disease. This should be the case of patients n. 25 and n. 26, both of whom were carriers of the p.Tyr179Cys recurrent mutation. Indeed, p.Tyr179Cys, together with p.Gly396Asp, is predominant, reaching subpolymorphic frequencies in Southern Europe (Cattaneo *et al.*, 2007).

Two patients proved to carry a single *MUTYH* pathogenic mutation. Monoallelic mutation carriers have been reported to have a small increased risk of CRC (Theodoratou *et al.*, 2010) and in monoallelic-mutated patients with a first-degree relative affected by CRC, such a risk is estimated to be sufficiently high to warrant more intensive surveillance than is offered to the general population (Win *et al.*, 2014). In our survey, the two patients with a monoallelic pathogenic mutation also carried a second variant (p.Gly457Trp and p.Gln338His, respectively) that might further increase the disease risk.

Bioinformatics produced inconclusive results for p.Gln338His polymorphic variant. However, clinical studies suggested an

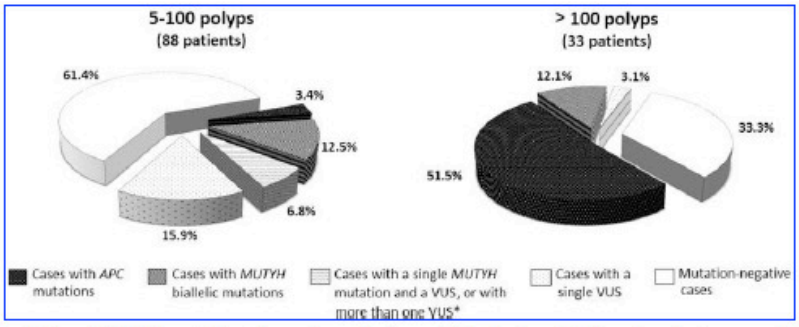

FIG. 1. *APC* and *MUTYH* mutations by number of polyps. *This fraction includes six patients: one p.Ile223Val/p.Gln338His, one p.Gln338Ter/p.Gly457Trp, and four p.Gln338His homozygous carriers.

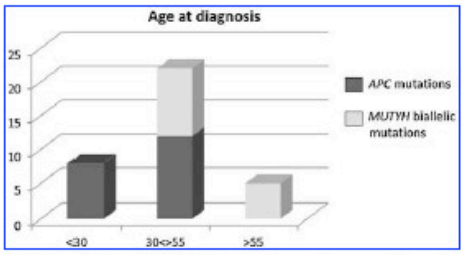

FIG. 2. *APC*- and *MUTYH*-mutated cases by age at diagnosis.

increased risk to CRC for p.Gln338His carriers (Picelli *et al.*, 2010), and a positive statistical association under a recessive model has recently been reported between this variant and CRC risk (Kasahara *et al.*, 2008; Tao *et al.*, 2008; Picelli *et al.*, 2013). Moreover, the mutated protein leads to increased sensitivity to oxidant treatment and defective removal of DNA damage when expressed in Mutyh null mouse embryonic fibroblasts (Raetz *et al.*, 2012; Turco *et al.*, 2013). The *in vivo* repair-defective phenotype has been explained by the impairment of p.Gln338His-mutated protein in its capability of interacting with RAD9:RAD1:HUS1, a heterotrimeric complex acting as a replication checkpoint sensor in response to DNA damage (Turco *et al.*, 2013; Brinkmeyer and David, 2015). Indeed, the 338 residue, where glutamine is replaced by histidine, is far from the catalytic site of MUTYH, but is located in its interconnecting domain that interacts with the HUS1 component of the heterotrimeric complex. On the basis of all the above considerations, p.Gln338His should be classified as a medium-penetrance variant. Accordingly, the p.Gln338His mutation might explain the clinical phenotype in the four homozygous carriers. The presence of a single pathogenic mutation with a VUS might also explain the two patients with p.Ile223Val/p.Gln338His and p.Gln338Ter/p.Gly457Trp genotypes (Table 5).

Constitutional mutations of *POLE* and *POLD1* have been associated with PPAP (Palles *et al.*, 2013). We analyzed for the first time a sample of Italian patients: none of the 36 subjects proved to carry *POLE/POLD1* mutations or VUSs in the exonuclease proofreading domain of DNA Polε and Polδ. This finding can be explained by the rarity of PPAP. Otherwise, it can be attributed to the limited size of our sample or to the patient selection process itself.

In our sample, clinical phenotype was quite variable among mutated and nonmutated cases, including number of polyps, age at diagnosis, CRC presence, and extracolonic manifestations. We found a significant correlation between the presence of *APC* mutations and a more aggressive phenotype, that is, the presence of at least 100 polyps. Interestingly, in a very large cohort of patients of various ancestry, including Western/Northern European and Central/East European, Grover *et al.* (2012) reported that the *APC* mutation rate was progressively increasing with increasing polyp burden, while *MUTYH* mutation rates were relatively constant across patients belonging to different polyp-count categories. Compared to *MUTYH*, we also found a significant association between *APC* genetic lesions and early disease onset; in particular, carriers of pathogenic mutations with a disease onset before 30 years of age were all carriers of *APC* mutations.

In our series, we could not identify a disease-causing mutation in 86 cases. In some of these cases, the clinical phenotype could be ascribable to constitutional mutations only detectable by specific molecular approaches. Previous messenger RNA (mRNA) analysis on mutation-negative patients revealed the existence of *APC/MUTYH* constitutional expression defects, including reduced/absent allelic expression (often due to unknown molecular mechanisms), and presence of aberrant transcripts (Renkonen *et al.*, 2005; Venesio *et al.*, 2007; Aceto *et al.*, 2015). In particular, up to 8% of negative cases have been reported to show aberrant transcription products associated with deep intronic *APC* variants (Spier *et al.*, 2012). A number of mutation-negative cases can also be explained by *APC* allelic deletions. In the largest report of genetic testing, including *APC* deletion analysis by multiplex ligation-dependent probe amplification, 6% of the mutated cases proved to be ascribable to gross *APC* rearrangements (Kerr *et al.*, 2013). It is also possible that some negative polyposis cases result from different combinations of low- or medium-penetrance variants along

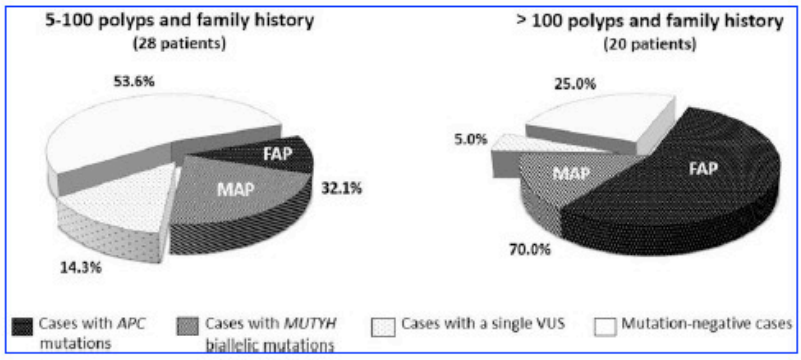

FIG. 3. *APC*- and *MUTYH*-mutated cases by number of polyps and family history of polyposis/CRC. CRC, colorectal cancer.

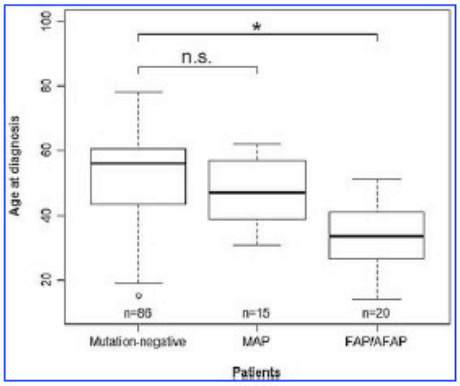

FIG. 4. Age at diagnosis per patients' group. In mutation-negative patients, the age of onset was significantly higher compared to FAP/AFAP patients, but not significantly different from MAP (Dunn's test *p*-values: $p = 3.46986 \times 10^{-7}$ and 0.24, respectively). *Black lines*: median values. *Whiskers*: minimum and maximum values that lie within 1.5× interquartile range. *White dot*: outlier. *Asterisk*: significant. FAP, familial adenomatous polyposis; MAP, *MUTYH*-associated polyposis; *n*, number of individuals within each category; n.s., not significant.

APC/MUTYH loci or in different susceptibility genes (Carvajal-Carmona *et al.*, 2013; Hes *et al.*, 2014; Cheng *et al.*, 2015). This is a more likely explanation for patients with a very mild phenotype and no evidence of family history. Of relevance, our 86 mutation-negative patients showed a mean age at diagnosis higher than those of FAP/AFAP and MAP and comprised 36% of subjects with 10–30 polyps and 9.3% of subjects with <10 polyps; moreover, in 61/86 cases (70.9%), there was no evidence of family history. On the whole, it has to be said that physicians recurrently suggest genetic testing to exclude hereditary polyposis, even though *a priori* probability of finding constitutional mutations appears extremely low.

TABLE 5. POLYPOSIS PATIENTS LIKELY DUE TO THE *MUTYH* GENOTYPE

ID	MUTYH *genotype*	*Sex*	*N. polyps*	*Age at diagnosis*	*CRC at diagnosis*
36	**p.Gln338Ter/** p.Gly457Trp	F	50–100	37	Yes
37	**p.Ile223Val/** p.Gln338His	M	60–70	48	Yes
38	p.Gln338His/ p.Gln338His	F	~30	59	No
39	p.Gln338His/ p.Gln338His	M	~40	66	Yes
40	p.Gln338His/ p.Gln338His	F	~15	55	No
41	p.Gln338His/ p.Gln338His	F	~50	23	No

Recognized pathogenic mutations are depicted in bold. None of these patients showed family history.

Finally, as recently emerged from next-generation sequencing (NGS) screening on other cohorts, some mutation-negative cases might be explained by rare high-penetrant mutations in new genes or by mutations in cancer predisposition genes that are unexpected based on patients' clinical features (Jaeger *et al.*, 2012; Weren *et al.*, 2015; Yurgelun *et al.*, 2015). At any rate, although NGS is rapidly increasing our knowledge on the molecular basis of inherited polyposis, patients characterized by a very mild phenotype and no or "weak" family history still represent a "diagnostic nightmare" (Lynch and Smyrk, 1998), and their genetic test remains a challenging problem in terms of cost/benefit (Syngal *et al.*, 2015).

Acknowledgments

This work was supported by the Cancer Prevention Screening Program of Regione Piemonte and by the Fondo Ricerca Docente of the University of Pavia. M.M. was recipient of a Giovanni Magni postdoctoral fellowship from the Adriano Buzzati-Traverso Foundation and R.A.K. was recipient of a PhD fellowship from Dunia Beam-Erasmus Mundus Action 2.

Author Disclosure Statement

No competing financial interests exist.

References

Aceto G, Curia MC, Veschi S, *et al.* (2005) Mutations of APC and MYH in unrelated Italian patients with adenomatous polyposis coli. Hum Mutat 26:394.

Aceto GM, Fantini F, De Iure S, *et al.* (2015) Correlation between mutations and mRNA expression of APC and MUTYH genes: new insight into hereditary colorectal polyposis predisposition. J Exp Clin Cancer Res 34:131–140.

Brinkmeyer MK, David SS (2015) Distinct functional consequences of MUTYH variants associated with colorectal cancer: damaged DNA affinity, glycosylase activity and interaction with PCNA and Hus1. DNA Repair (Amst) 34:39–51.

Carvajal-Carmona LG, Zauber AG, Jones AM, *et al.* (2013) Much of the genetic risk of colorectal cancer is likely to be mediated through susceptibility to adenomas. Gastroenterology 144:53–55.

Cattaneo F, Molatore S, Mihalatos M, *et al.* (2007) Heterogeneous molecular mechanisms underlie attenuated familial adenomatous polyposis. Genet Med 9:836–841.

Cheng TH, Gorman M, Martin L, *et al.* (2015) Common colorectal cancer risk alleles contribute to the multiple colorectal adenoma phenotype, but do not influence colonic polyposis in FAP. Eur J Hum Genet 23:260–263.

Cleary SP, Cotterchio M, Jenkins MA, *et al.* (2009) Germline MutY human homologue mutations and colorectal cancer: a multisite case-control study. Gastroenterology 136:1251–1260.

de Leon MP, Urso ED, Pucciarelli S, *et al.* (2013) Clinical and molecular features of attenuated adenomatous polyposis in northern Italy. Tech Coloproctol 17:79–87.

Gismondi V, Meta M, Bonelli L, *et al.* (2004) Prevalence of the Y165C, G382D and 1395delGGA germline mutations of the MYH gene in Italian patients with adenomatous polyposis coli and colorectal adenomas. Int J Cancer 109:680–684.

Grover S, Kastrinos F, Steyerberg EW, *et al.* (2012) Prevalence and phenotypes of APC and MUTYH mutations in patients with multiple colorectal adenomas. JAMA 308:485–492.
Half E, Bercovich D, Rozen P (2009) Familial adenomatous polyposis. Orphanet J Rare Dis 4:22.
Hes FJ, Ruano D, Nieuwenhuis M, *et al.* (2014) Colorectal cancer risk variants on 11q23 and 15q13 are associated with unexplained adenomatous polyposis. J Med Genet 51:55–60.
Jaeger E, Leedham S, Lewis A, *et al.* (2012) Hereditary mixed polyposis syndrome is caused by a 40-kb upstream duplication that leads to increased and ectopic expression of the BMP antagonist GREM1. Nat Genet 44:699–703.
Kasahara M, Osawa K, Yoshida K, *et al.* (2008) Association of MUTYH Gln324His and APEX1 Asp148Glu with colorectal cancer and smoking in a Japanese population. J Exp Clin Cancer Res 27:49.
Kerr SE, Thomas CB, Thibodeau SN, *et al.* (2013) APC germline mutations in individuals being evaluated for familial adenomatous polyposis: a review of the Mayo Clinic experience with 1591 consecutive tests. J Mol Diagn 15:31–43.
Lubbe SJ, Di Bernardo MC, Chandler IP, *et al.* (2009) Clinical implications of the colorectal cancer risk associated with MUTYH mutation. J Clin Oncol 27:3975–3980.
Lucci-Cordisco E, Risio M, Venesio T, *et al.* (2013) The growing complexity of the intestinal polyposis syndromes. Am J Med Genet A 161A:2777–2787.
Lynch HT, Smyrk TC (1998) Classification of familial adenomatous polyposis: a diagnostic nightmare. Am J Hum Genet 62:1288–1289.
Palles C, Cazier JB, Howarth KM, *et al.* (2013) Germline mutations affecting the proofreading domains of POLE and POLD1 predispose to colorectal adenomas and carcinomas. Nat Genet 45:136–144.
Picelli S, Lorenzo Bermejo J, Chang-Claude J, *et al.* (2013) Meta-analysis of mismatch repair polymorphisms within the cogent consortium for colorectal cancer susceptibility. PLoS One 8:e72091.
Picelli S, Zajac P, Zhou XL, *et al.* (2010) Common variants in human CRC genes as low-risk alleles. Eur J Cancer 46:1041–1048.
Pin E, Pastrello C, Tricarico R, *et al.* (2013) MUTYH c.933+3A>C, associated with a severely impaired gene expression, is the first Italian founder mutation in MUTYH-associated polyposis. Int J Cancer 132:1060–1069.
Raetz AG, Xie Y, Kundu S, *et al.* (2012) Cancer-associated variants and a common polymorphism of MUTYH exhibit reduced repair of oxidative DNA damage using a GFP-based assay in mammalian cells. Carcinogenesis 33:2301–2309.
Renkonen ET, Nieminen P, Abdel-Rahman WM, *et al.* (2005) Adenomatous polyposis families that screen APC mutation-negative by conventional methods are genetically heterogeneous. J Clin Oncol 23:5651–5659.
Rohlin A, Eiengård F, Lundstam U, *et al.* (2016) GREM1 and POLE variants in hereditary colorectal cancer syndromes. Genes Chromosomes Cancer 55:95–106.
Spier I, Horpaopan S, Vogt S, *et al.* (2012) Deep intronic APC mutations explain a substantial proportion of patients with familial or early-onset adenomatous polyposis. Hum Mutat 33:1045–1050.
Syngal S, Brand RE, Church JM, *et al.* (2015) ACG clinical guideline: genetic testing and management of hereditary gastrointestinal cancer syndromes. Am J Gastroenterol 110:223–262.
Tao H, Shinmura K, Suzuki M, *et al.* (2008) Association between genetic polymorphisms of the base excision repair gene MUTYH and increased colorectal cancer risk in a Japanese population. Cancer Sci 99:355–360.
Theodoratou E, Campbell H, Tenesa A, *et al.* (2010) A large-scale meta-analysis to refine colorectal cancer risk estimates associated with MUTYH variants. Br J Cancer 103:1875–1884.
Tomlinson I (2015) The Mendelian colorectal cancer syndromes. Ann Clin Biochem 52:690–692.
Turco E, Ventura I, Minoprio A, *et al.* (2013) Understanding the role of the Q338H MUTYH variant in oxidative damage repair. Nucleic Acids Res 41:4093–4103.
Venesio T, Balsamo A, D'Agostino VG, *et al.* (2012) MUTYH-associated polyposis (MAP), the syndrome implicating base excision repair in inherited predisposition to colorectal tumors. Front Oncol 2:83.
Venesio T, Balsamo A, Rondo-Spaudo M, *et al.* (2003a) APC haploinsufficiency, but not CTNNB1 or CDH1 gene mutations, accounts for a fraction of familial adenomatous polyposis patients without APC truncating mutations. Lab Invest 83:1859–1866.
Venesio T, Balsamo A, Scordamaglia A, *et al.* (2003b) Germline APC mutation on the beta-catenin binding site is associated with a decreased apoptotic level in colorectal adenomas. Mod Pathol 16:57–65.
Venesio T, Balsamo A, Sfiligoi C, *et al.* (2007) Constitutional high expression of an APC mRNA isoform in a subset of attenuated familial adenomatous polyposis patients. J Mol Med 85:305–312.
Venesio T, Molatore S, Cattaneo F, *et al.* (2004) High frequency of MYH gene mutations in a subset of patients with familial adenomatous polyposis. Gastroenterology 126:1681–1685.
Weren RD, Ligtenberg MJ, Kets CM, *et al.* (2015) A germline homozygous mutation in the base-excision repair gene NTHL1 causes adenomatous polyposis and colorectal cancer. Nat Genet 47:668–671.
Win AK, Dowty JG, Cleary SP, *et al.* (2014) Risk of colorectal cancer for carriers of mutations in MUTYH, with and without a family history of cancer. Gastroenterology 146:1208–1211.
Yurgelun MB, Allen B, Kaldate RR, *et al.* (2015) Identification of a variety of mutations in cancer predisposition genes in patients with suspected lynch syndrome. Gastroenterology 149:604–613.

Address correspondence to:
Guglielmina Nadia Ranzani, PhD
Department of Biology and Biotechnology
University of Pavia
Via Ferrata 9
27100 Pavia
Italy

E-mail: guglielmina.ranzani@unipv.it

Acknowledgements

First and foremost I want to thank the Dunia Beam-Erasmus Mundus Action 2, for providing me with the scholarship necessary to pursue my PhD here at the University of Pavia. An equal weight of gratitude goes to Prof G. Nadia Ranzani for taking a chance on me as a PhD candidate at her laboratory of Human Genetics-Cancer Genetics and for the guidance and care she has shown me over the past three years.

A heartfelt thank you is in place to the members of the "Biosciences Laboratory" of the "Istituto Scientifico Romagnolo per lo Studio e la Cura dei Tumori (IRST)-IRCCS" of Meldola, not only for giving me access to freely plan and perform my experiments at their facility, but also for providing the necessary samples and support to see the project through. This is especially true for Dr. Daniele Calistri and Dr. Chiara Molinari.

Thanks from my part is also due to our collaborators at the "European Institute of Oncology-IEO" of Milano, for their input regarding the selection of the HDGC cases.

A grand thank you is additionally in place to Dr Monica Marabelli (Laboratory of Human Genetics-Cancer Genetics), Alessandro Raveane (Laboratory of Human Genetics of Population) and Simone Savino (Laboratory of Biocrystallography), for their quintessential help and support throughout my PhD program.

Last but not least, I would like to take a moment to acknowledge the love and patience of my family and especially that of my fiancé, Fouad Salha; without them I could not imagine being where I am today.