Computational Biology and Chemistry 60 (2016) 9–20

Contents lists available at ScienceDirect

Computational Biology and Chemistry

journal homepage: www.elsevier.com/locate/compbiolchem

Research article

CDH1/E-cadherin and solid tumors. An updated gene-disease association analysis using bioinformatics tools

María Florencia Abascal[a,1,2], María José Besso[a,1,2], Marina Rosso[a,2], María Victoria Mencucci[a,2], Evangelina Aparicio[a,2], Gala Szapiro[a,2], Laura Inés Furlong[b], Mónica Hebe Vazquez-Levin[a,c,*]

[a] Laboratory of Cell–Cell Interaction in Cancer and Reproduction, Instituto de Biología & Medicina Experimental (IBYME), Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Fundación IBYME (FIBYME), Vuelta de Obligado 2490, Zip Code C1428ADN, Buenos Aires, Argentina
[b] Research Programme on Biomedical Informatics (GRIB) (IMIM), DCEXS, Universitat Pompeu Fabra, C/Dr Aiguader 88, Zip Code 08003, Barcelona, Spain
[c] Laboratory of Cell-Cell Interaction in Cancer and Reproduction, Instituto de Biología y Medicina Experimental (IBYME), Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Fundación IBYME (FIBYME), Vuelta de Obligado 2490, Zip Code C1428ADN, Buenos Aires, Argentina

ARTICLE INFO

Article history:
Received 24 May 2015
Received in revised form 17 October 2015
Accepted 19 October 2015
Available online 7 November 2015

Keywords:
CDH1
Epithelial Cadherin
Solid tumors
Bioinformatics
Somatic mutations

ABSTRACT

Cancer is a group of diseases that causes millions of deaths worldwide. Among cancers, Solid Tumors (ST) stand-out due to their high incidence and mortality rates. Disruption of cell–cell adhesion is highly relevant during tumor progression. Epithelial-cadherin (protein: E-cadherin, gene: *CDH1*) is a key molecule in cell–cell adhesion and an abnormal expression or/and function(s) contributes to tumor progression and is altered in ST. A systematic study was carried out to gather and summarize current knowledge on *CDH1*/E-cadherin and ST using bioinformatics resources. The DisGeNET database was exploited to survey *CDH1*-associated diseases. Reported mutations in specific ST were obtained by interrogating COSMIC and IntOGen tools. *CDH1* Single Nucleotide Polymorphisms (SNP) were retrieved from the dbSNP database.

DisGeNET analysis identified 609 genes annotated to ST, among which *CDH1* was listed. Using *CDH1* as query term, 26 disease concepts were found, 21 of which were neoplasms-related terms. Using DisGeNET ALL Databases, 172 disease concepts were identified. Of those, 80 ST disease-related terms were subjected to manual curation and 75/80 (93.75%) associations were validated. On selected ST, 489 *CDH1* somatic mutations were listed in COSMIC and IntOGen databases. Breast neoplasms had the highest *CDH1*-mutation rate. *CDH1* was positioned among the 20 genes with highest mutation frequency and was confirmed as driver gene in breast cancer. Over 14,000 SNP for *CDH1* were found in the dbSNP database.

This report used DisGeNET to gather/compile current knowledge on gene-disease association for *CDH1*/E-cadherin and ST; data curation expanded the number of terms that relate them. An updated list of *CDH1* somatic mutations was obtained with COSMIC and IntOGen databases and of SNP from dbSNP. This information can be used to further understand the role of *CDH1*/E-cadherin in health and disease.

1. Introduction

Cancer is a group of diseases characterized by an uncontrolled cell proliferation resistance to cell death, induction of angiogenesis, activation of invasion and metastasis and growth suppressor evasiveness (Negrini et al., 2010; Hanahan and Weinberg 2011). Among different cancer types, Solid Tumors (ST) stand-out due to their high incidence and mortality rates. Over 90% of ST start in the epithelium, a tissue composed of cells interconnected by intercellular junctions, among them the adherent junctions (Cooper 2000). Disruption of cell–cell adhesion is a very relevant event during tumor progression and metastasis in ST. In the transition to malignancy, down-regulation of cell–cell adhesion

* Corresponding author at: Instituto de Biología y Medicina Experimental (IBYME), Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Fundación IBYME (FIBYME), Vuelta de Obligado 2490 (C1428ADN), Buenos Aires Argentina. Fax: +5411 4786 2564.
E-mail addresses: abascalflorencia@gmail.com (M.F. Abascal), majobesso_88@hotmail.com (M.J. Besso), rossomarina84@gmail.com (M. Rosso), viki_mvm89@hotmail.com (M.V. Mencucci), aparicio.evangelina32@gmail.com (E. Aparicio), galaszapiro@gmail.com (G. Szapiro), lfurlong@imim.es (L.I. Furlong), mhvazl@gmail.com, mhvaz@levindigital.com.ar (M.H. Vazquez-Levin).
[1] These authors contributed equally to this work.
[2] Fax: +54 11 47862564.

http://dx.doi.org/10.1016/j.compbiolchem.2015.10.002
1476-9271/

molecules, cytoskeleton reorganization and signaling pathway activation that avoid adherent junction assembly accompanies an increase of proliferation and migration (Conacci-Sorrell et al., 2002; Le Bras et al., 2012; Vasioukhin 2012).

Adherent junctions participate in processes involved in keeping cellular organization. Among their functions are maintenance of cell polarity and tissue integrity, cytoskeletal dynamics and movement within epithelium proliferation, transcription, differentiation and survival (Ivanov and Naydenov, 2013; Perez-Moreno et al., 2003). They are composed of classical cadherins, a vast superfamily of membrane proteins that form mainly Ca^{2+}-dependent homophilic interactions to maintain cell–cell contact (Angst et al., 2001; Kemler 1992; Nishimura and Takeichi, 2009; Oda and Takeichi, 2011). Epithelial cadherin (E-cadherin) is the founder member of the cadherin superfamily (Takeichi 1977). It is a Ca^{2+}-dependent cell-adhesion molecule mainly expressed in epithelial cells, essential for development, cell differentiation and tissue homeostasis, as well as for maintenance of epithelial polarity and structural integrity (van Roy and Berx, 2008). E-cadherin localization is restricted to cell–cell contact sites, and part of the cell surface-located E-cadherin is subjected to endocytosis and recycling (Bryant and Stow, 2004; Mosesson et al., 2008). The human E-cadherin gene, named *CDH1,* is located on chromosome 16q22.1, spans a region of approximately 100 kb, and comprises 16 exons and 15 introns (Berx et al., 1995).

The E-cadherin mature protein is a 120 kDa glycoprotein organized in an extracellular domain (ectodomain) of five tandem cadherin motifs, a single transmembrane domain, and a highly conserved cytoplasmic domain. The E-cadherin extracellular domain mediates mainly homophilic cell–cell adhesions between adjacent cells (Nose et al., 1990; Ozawa et al., 1990). On the other hand, specific sequences of the E-cadherin intracellular domain participate in regulation of its adhesive activity (Ozawa and Kemler, 1998) and interact with several proteins, among them α-, β-, δ-, p120- and γ- (plakoglobin) catenins (official symbols CTNNA1, CTNNB1,CTNND2, CTNND1 and CTNNG) that form a complex link to the actin cytoskeleton, which regulates the strength of the cadherin-mediated cell adhesion, and are involved in signal transduction pathways (Nagafuchi et al., 1993; Hong et al., 2013).

The functional roles of E-cadherin anticipate that genetic and epigenetic alterations on the *CDH1* gene have great implications on tumor invasion and metastasis, with a loss or reduced expression of E-cadherin, resulting in a more invasive tumor (Gall and Frampton, 2013). E-cadherin has been defined as an invasion tumor suppressor since it has been frequently found down-regulated in epithelial tumors, a process that leads to cell motility and invasion. In fact, a reduced/lack of E-cadherin expression or/and loss of function contributes to cancer progression by increasing proliferation, invasion and metastasis (Berx and van Roy, 2009; Gheldof and Berx, 2013; Schneider and Kolligs, 2014; van Roy, 2014).

Disruption of E-cadherin expression and loss of its function(s) has been extensively documented in several ST. Examples are breast (Sinn et al., 2014), ovarian (Cowden Dahl et al., 2008), gastric (Schildberg et al., 2014), endometrial (Wójcik-Krowiranda et al., 2013), colorectal (Deng et al., 2014) and bladder (Bryan 2015) cancers. Several mechanisms of E-cadherin inactivation have been reported, among them are loss of heterozygosity at the 16q22.1 chromosome region (Chalmers et al., 2001), presence of inactivating mutations (Berx et al., 1998; Corso et al., 2014), CpG-island hypermethylation of the *CDH1* promoter (Caldeira et al., 2006; Gall and Frampton, 2013; Kanazawa et al., 2002), gene expression silencing by binding of specific transcription factors to sequences in the *CDH1* promoter (Zhang et al., 2014), and post-translational modifications (i.e. proteinase processing/phosphorylation/glycosylation) that negatively regulate E-cadherin functions (Rashid et al., 2001).

As a result of over 30 years of research since its identification, *CDH1*/E-cadherin has been the subject of numerous studies that led to a vast number of reports in scientific journals (over 21,000 publications using "E-cadherin" keyword, 11,000 publications using "E-cadherin AND cancer"; PubMed search on January 2015). This exceptional growth of information requires integrative approaches such as translational bioinformatics to transform the deluge of data into knowledge and, more importantly, to enable a deeper understanding of disease mechanisms and provide actionable information for the clinical practice (Altman, 2012; Sarkar et al., 2011). Publicly available comprehensive knowledge sources on disease genes are an important asset. The "big data" phenomenon in biomedical information is also observed in the scientific literature. Nowadays, the increasing size of literature repositories makes imperative the use of computational tools to identify relevant information and place it in the context of current biomedical knowledge. Several bioinformatics tools were developed to survey/gather information. Among them is DisGeNET, a knowledge platform on human diseases and their genes plugin for Cytoscape to query and analyze human gene-disease networks (Bauer-Mehren et al., 2010; Piñero et al., 2015). In some cases, data curation is done, and involves the identification, review and organization of the gathered information by a human expert to make it accessible to both other experts and computer systems, and it is particularly important to filter/prioritize information provided by automatic text-mining approaches (Howe et al., 2008). DisGeNET has been used for the analysis of mechanisms underlying adverse drug reactions (Bauer-Mehren et al., 2011; Grosdidier et al., 2014), the association between diabetes and Parkinson disease (Santiago and Potashkin, 2014), the prediction of disease associations for ncRNAs (Alaimo et al., 2014) and the analysis of disease-relevant nodes in metabolic pathways (Galhardo et al., 2013), among other studies.

Since tumor development has been related to the presence of gene mutations in numerous tissues, in particular in the case for the *CDH1* gene in which mutations have been reported in several publications mainly related to breast and gastric cancers (Berx et al., 1998; Corso et al., 2013, 2014; Valente et al., 2014). During the past decades, the number of reported mutations has largely increased, mainly from high-throughput approaches using next generation sequencing technologies (Pastrello et al., 2014). This information can be found in the scientific publications, and is being systematically compiled in specific databases that gather and organize the data. Among these resources are the COSMIC (Catalogue-Of-Somatic-Mutations-In-Cancer) (Forbes et al., 2010) and the IntOGen (Integrated-Onco-Genomics) (Perez-Llamas et al., 2011; Gonzalez-Perez et al., 2013) tools to search for gene mutations.

Based on the relevance of *CDH1* in human physiopathology, a systematic search was carried out to gather/summarize current knowledge on the *CDH1*/E-cadherin gene/protein and its role in human disease, in particular in cancer, using a selection of bioinformatic resources. The information contained in DisGeNET was exploited to gather diseases associated to *CDH1*, and this information was complemented with knowledge on mutations described in specific cancer samples by interrogating COSMIC and IntOGen and on SNP from dbSNP database.

2. Materials and methods

2.1. Bioinformatics tools

2.1.1. DisGeNET

Discovery platform integrating information on human diseases and their genes from expert–curated databases and the scientific literature discovered by text-mining approaches (Piñero et al.,

2015). Data is organized according to their type and level of curation: CURATED (gene-disease associations from UNIPROT and CTD human databases), PREDICTED (gene-disease associations from CTD mouse and CTD rat data, RGD and MGD), LITERATURE, GAD, LHGDN and BeFree (Bravo et al., 2014, 2015), and ALL (CURATED, PREDICTED and GAD, LHGDN and BeFree). DisGeNET classifies diseases according the MeSH hierarchy and genes according to the PANTHER protein ontology. The gene-disease associations can be ranked according to the DisGeNET score and are annotated with the DisGeNET gene-disease association type ontology. DisGeNET gene-disease score takes into account the number and type of sources (level of curation, organisms) and the number of publications supporting the association. The score ranges from 0 to 1 and it is computed according to: $S = (W_{\text{UNIPROT}} + W_{\text{WCTDhuman}}) + (W_{\text{Mouse}} + W_{\text{Rat}}) + (W_{\text{GAD}} + W_{\text{LHGDN}} + W_{\text{BeFree}})$. DisGeNET allows queries restricted to genes or diseases and identifies gene-diseases associations, type of associations and evidence that support the associations (publications). It can be accessed through the web interface (http://www.disgenet.org/web/DisGeNET/menu/dbinfo) or using a Cytoscape plugin (Bauer-Mehren et al., 2010). The current version contains 381,056 associations, between 16,666 genes and 13,172 diseases (http://www.disgenet-org/, accessed on January 2015).

2.1.2. COSMIC

Catalog of public domain data that gathers/organizes information available about somatic mutations found in various cancers, combining information manually curated from the scientific literature (PubMed) with the output derived from the "Cancer Genome Project" (Sanger Institute, United Kingdom). COSMIC searches can be performed on their online interface; further information and folders FASTA of genes can be downloaded. Genes are selected according to "Cancer Gene Census", which incorporates genes implicated in cancer with a causal relationship (Forbes et al., 2010). COSMIC provides information about mutation position and type in a gene of interest, amino acid(s) involved and (in case of substitutions), the change that occurs with the mutation. Additionally, it provides information about mutations affecting tumor-associated genes and establishes a hierarchy of the twenty most commonly mutated genes in a specific tumor. It also provides number of cases in which a mutation was reported. Moreover, it provides Mutation Impact filters derived from the FATHMM-MKL algorithm (Functional Analysis through Hidden Markov Models). The FATHMM-MKL algorithm predicts the functional, molecular and phenotypic consequences of protein missense variants using hidden Markov models. Where FATHMM-MKL scores are $\geq$0.7 the mutation is classified as 'Pathogenic', or 'Neutral' if the score is $\leq$0.5 (Shihab et al., 2013). Additionally, it provides information on Copy Number Variation (CNV). The COSMIC web interface does not include SNP.

The 2014 COSMIC release contains a major update on cancer genomes, including over a million novel mutations from ICGC sequencing projects.

2.1.3. IntOGen

Platform to search for mutations, genes and pathways involved in tumorigenesis across 4623 cancer genomes/exomes from 13 cancer sites (mainly from the International Cancer Genome Consortium, ICGC, and The Cancer Genome Atlas, TCGA). It analyzes somatic mutations in a cohort of tumors to identify cancer driver genes and pathways, and to present results of systematic analysis from most currently available large data sets of tumor somatic mutations. Analysis is based on the assumption that cancer driver genes accumulate highly functional mutation. It currently includes Oncodrive FM, a tool that detects genes significantly biased toward the accumulation of mutations with high functional impact (FM bias), and Oncodrive CLUST, which picks up genes whose mutations tend to cluster in particular regions of the protein sequence with respect to synonymous mutations (CLUST bias). Both tools detect signals of positive selection, which appear in genes whose mutations are selected during tumor development and are therefore likely drivers (Gonzalez-Perez et al., 2013).

COSMIC and IntOGen tools were browsed to survey current information on somatic mutations in *CDH1* associated to ST. The degree of similarity between the two outputs was determined by comparing the *CDH1* mutations listed by the database in each tissue analyzed.

2.1.4. dbSNP

A search was done for reported Single Nucleotide Polymorphisms (SNP) on the *CDH1* gene using the dbSNP (The Single Nucleotide Polymorphism database). The NCBI Short Genetic Variations database catalogs short variations in nucleotide sequences from a wide range of organisms. These variations include single nucleotide variations, short nucleotide insertions

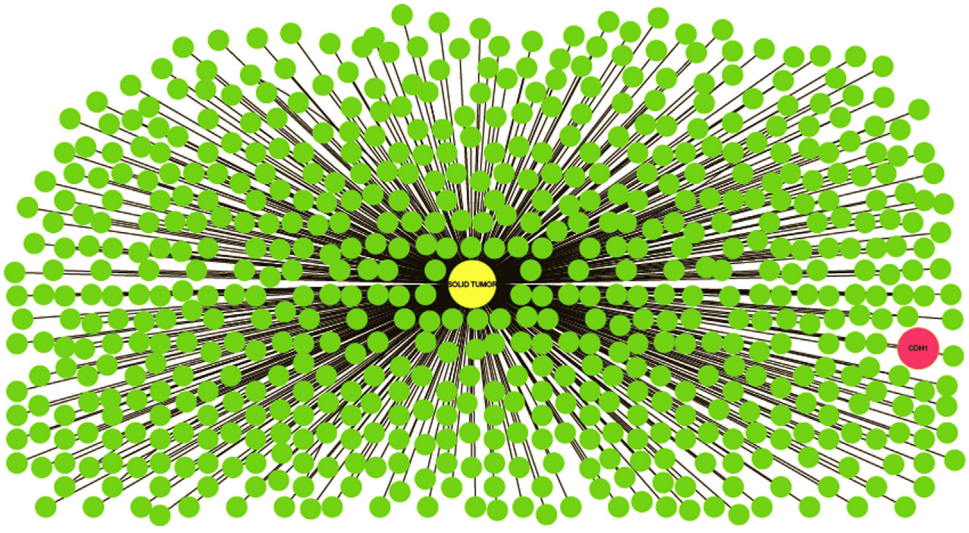

Fig. 1. Network representation of the genes annotated to the term Solid Tumor obtained with DisGeNET. A Yellow node represents the disease node Solid Tumor. Green nodes represent genes related to the disease (609 genes). In red, the *CDH1* gene is highlighted.

and deletions, short tandem repeats and microsatellites. dbSNP currently classifies nucleotide sequence variations with the following types and percentage composition of the database: (i) single nucleotide substitutions, 99.77%, (ii) small insertion/deletion polymorphisms, 0.21%, (iii) invariant regions of sequence, 0.02%, (iv) microsatellite repeats, 0.001%, (v) named variants, <0.001%, and (vi) uncharacterized heterozygous assays, <0.001%. SNPs listed are somatic and germinal (Sherry et al., 1999).

The complete contents of dbSNP are available to the public at http://www.ncbi.nlm.nih.gov/SNP

2.2. Expert curation

A manual curation of *CDH1*/E-cadherin related diseases identified from DisGeNET All Databases sources was performed by reviewing publications supporting each *CDH1*-disease association. Gene-disease associations were "confirmed" when a relevant association between *CDH1* and the disease had been identified (Mutation/Genetic Variation/Post-translational modification/Biomarker). Associations were "inconclusive" when the relevance of *CDH1* wasn't concluded. Finally, associations were "unconfirmed" when no association between *CDH1* and the disease under study was identified. Some terms were related to more than one publication; in these cases, all the publications were reviewed, but the term was "validated" when the gene-disease association was confirmed in at least one publication.

3. Results

3.1. Identification of genes associated to ST

A first analysis was conducted to gather genes annotated to the term "Solid Tumors (ST)" from the DisGeNET database, resulting in 609 genes, among which *CDH1* was identified. These results are shown as a network of genes associated to the term "ST" (Fig. 1, *CDH1* is highlighted). *CDH1* was classified as a Biomarker gene associated to ST (*Source BeFree*, PMID: 9581841). The supporting evidence for this association referred to a publication describing a key role of *CDH1* loss or reduced expression in proliferation, invasion and metastasis of breast cancer. *CDH1* expression, genetic mutation and promoter methylation was analyzed in 10 established breast cancer cell lines (Hiraguri et al., 1998).

The 609 genes associated to ST were classified according to the PANTHER Protein Classification (data not shown). *CDH1*/E-cadherin was classified among protein classes "Cell adhesion molecule" and "Cell junction protein", together with other ST associated proteins (Table 1A). Regarding the protein pathways involved, *CDH1*/E-cadherin was associated to apoptosis, cell communication, extracellular matrix organization and immune system (Table 1B).

3.2. Identification of CDH1-associated diseases using DisGeNET

To further explore the specific diseases associated to *CDH1*, DisGeNET was interrogated using *CDH1* as query term. First, the search was restricted to expert curated databases (CURATED), which resulted in 26 disease concepts associated to *CDH1* (grey central node, Fig. 2). Five of them were classified as non-neoplastic diseases (orange nodes Fig. 2) and 21 as neoplasms-related terms. Among these 21 terms, 4 terms were related to tumor properties (green nodes, Fig. 2), and the remaining 17 terms did correspond to specific ST (pink nodes, Fig. 2). The associations found between *CDH1* and the 26 nodes were from the classes Biomarker, Genetic Variation and Therapeutic (DisGeNET gene-disease association type ontology; Fig. 3 and Supplementary Table 1).

Next, *CDH1*-associated diseases were explored in DisGeNET ALL. This source includes data automatically extracted from

Table 1
CDH1 and Panther Classification.

A. Protein class

Panther protein class	Gene
Cell adhesion molecule (30)	***CDH1***-*CD9-CD47-CD53-CD55-CD63-CD82- CEACAM3-CEACAM5-CEACAM6-CEACAM7-CLEC4D-CXADR-ICAM1-GPNMB-ICAM2-KITLG-MAGEA3-MAGEA6-MAGED4-MAGED4B-NEDD9-NRP2-PCDH10-POSTN-PSG2-ROBO1-SDC1-SPP1-VWF*
Cell Junction Protein (4)	***CDH1***-*CLDN4-GJA1-TMF1*

Left column: Panther protein class terms for Solid Tumor-related genes are shown (only classes in which *CDH1* encoding E-cadherin are listed). In parenthesis, the number of genes listed in the right column. *Right column*: Solid Tumor-related genes (Hugo Gene Nomenclature Committee). The gene ***CDH1*** is highlight in bold and it is classified as "cell adhesion molecule" and "cell junction protein".

B. Protein classification

Pathway	Gene
Apoptosis (17)	*APC-BAK1-BAX-BCL2-BCL2L11-CASP8-**CDH1**-CLSPN-DAPK1-DYNLL1-FAS-TNF-TNFSF10-TNFRSF10B TP53-VIM-RIPK1*
Cell communication (8)	***CDH1***-*CD47-CLDN4-KRT5-PIK3CA-PIK3CB-PTK2B-PTPN11*
Extracellular matrix organization (27)	*ADAM10-ADAMTS1-ADAMTS8-BSG-**CDH1**-CD44-CD47-COL10A1-COL18A1-COL4A2-CTSB-DDR1-FGF2-ICAM1-ICAM2-KDR-LOX-MMP7-MMP9-MMP13-PECAM1-PLG- SDC1-SERPINE1- SPP1-TGFB2-THBS1-VWF*
Immune system (112)	*ABI1-ABL1-ADAR-ATF1-BCL2-BCL10-CALM1-CALM2-CALM3-CASP1-CASP8-**CDH1**-CD200-CD274-CD4-CD40-CD40LG-CD55-CD74-CD80-CD8A-CSF2-CD44-CREB1-CTLA4-CTSB-CXADR-DCTN3-DEFA4-DRB4-DYNC1H1-DYNLL1-EGF-EGFR-EGR1-ERBB2-ERBB4-FGF2-FGFR3-FGFR4-FLNB-FOXO1-FOXO3-FZR1-HGF-HLA-HLA-A-HLA-E-HLA-G-HRAS-ICAM1-ICAM2-IFNA1-IFNA13-IFNB1-IFNG-IL1B-IL1R1-IL1RN-IL18-IL2-IL2RG-IL7-IRF8-IRF9-IRS1-ITK-IL1A-JAK2-KIF2C-KIF22-KIF4A-KIR3DL1- KIT-KITLG-KLRC4-KLRK1-KRAS-LGMN-MASP2-MBL2-MDM2-MEFV-MRC1-NFKB1-PDCD1-PDGFRA-PDGFRB-PDPK1-PIK3CA-PIK3CB-PIK3CD-PML-IL6-PRKAR1A- PTEN-PTK2B-PTPN11-PTPRC-RAB7A-RAF1-RIPK1-STAT3-SOCS1-SOCS3SOS1-STIM1-TUBB3-UBE2C-TRAT1-TXK-VHL-ZBTB16*

Left column: Panther pathway terms for Solid Tumor-related genes are shown (only pathways in which *CDH1* encoding E-cadherin are listed). In parenthesis, the number of genes listed in the right column. Right column: Solid Tumor-related genes (Hugo Gene Nomenclature Committee). The gene ***CDH1*** is highlight in bold and it is associated to "Apoptosis", "Cell comunication", "Extracellular matrix organization" and "Immune system". Source: DisGeNET.

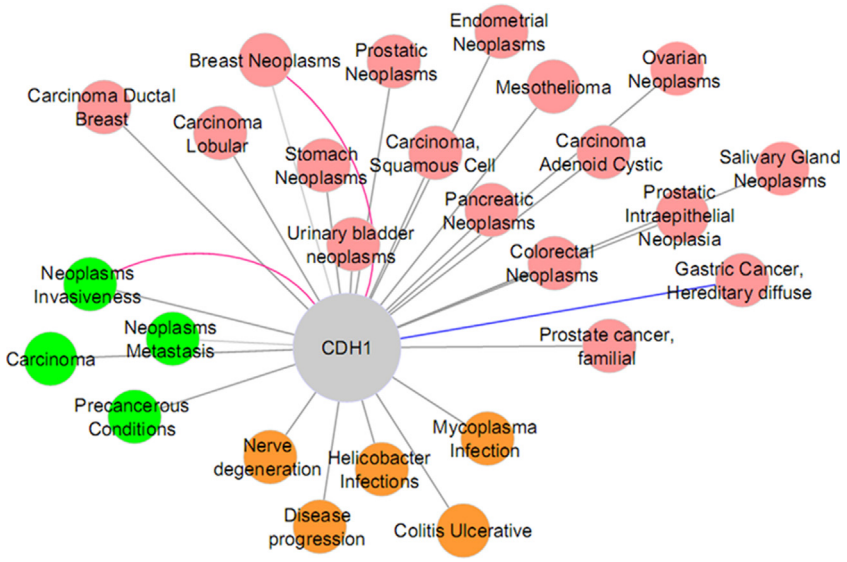

Fig. 2. Network representation of the Gene-Disease associations for *CDH1* obtained with DisGeNET curated sources. A grey node represents the restriction term *CDH1*. Pink nodes represent different types of solid tumors (17/26). Green nodes correspond to tumor properties related terms associated with *CDH1* (4/26) and orange nodes to other non-neoplastic diseases (5/26). The edges between nodes represent the association types: blue: genetic variation; pink: therapeutic; grey: biomarker. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

current literature by text-mining approaches in addition to the information from expert-curated databases. *CDH1* was found associated to 172 disease concepts: 132 neoplasms-related (green, pink, purple and light-blue nodes, Fig. 4) and 40 to non-neoplastic diseases (yellow nodes, Fig. 4). From the 132 neoplasms-related concepts, 24 were tumor properties/neoplasm-related concepts (green nodes), 97 were ST (pink), 8 were neoplastic processes (purple nodes) and 3 were non-solid tumors neoplasms (light-blue nodes) (Fig. 5 and Supplementary Table 2). The *CDH1* gene-disease association types were Biomarker, Altered Expression, Genetic Variation, Postranslational Modification and Therapeutic.

Since some associations were only retrieved by automatic text-mining approaches, a manual curation was performed to assess their relevance. Of the 172 terms found in all DisGeNET databases, 26 had already been curated by the DisGeNET team (Fig. 4). The focus was made on 80 disease terms related to ST not reported as associated to *CDH1* by expert-curated databases in DisGeNET (Supplementary Table 2). A total of 286 publications were reviewed, since several associations were supported by more than one publication. As result of the analysis, the association between *CDH1* and the disease was validated in 199/286 (70%) publications.

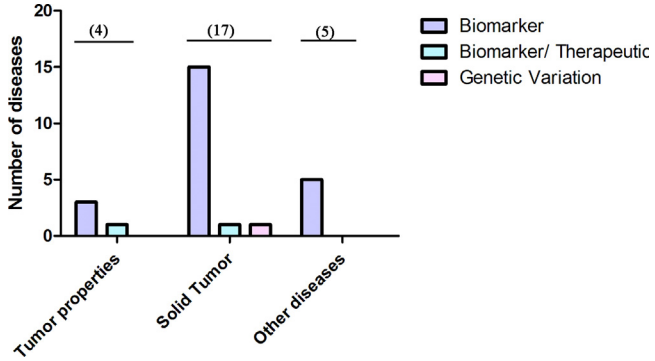

Fig. 3. *CDH1*-related diseases and Association Types resulting from DisGeNET curated databases. A graphical representation; detailed information is presented in Supplementary Table 1.

In 67/286 (23%) publications, the association was "inconclusive" and in the remaining 20/286 (7%) it was "unconfirmed" (Fig. 6, upper panel). Finally, 75/80 (93.75%) associations between *CDH1* and ST concepts were validated (Fig. 6, lower panel). This curation effort added new disease terms related to ST to those already available in DisGeNET CURATED. These terms correspond to bladder, breast, colorectal, endometrial, liver, ovarian, pancreatic, prostatic, stomach and gastrointestinal neoplasms. In addition, it validated new *CDH1*-related tumor tissues, among them bile ducts, brain, oesophagus, gallbladder, kidney, head and neck tissues, larynges, lung, meninges, mouth, muscles, parathyroid and thyroid, pituitary gland, rectum, skin, tonsil, thymus.

3.3. CDH1 *somatic mutations and SNP in ST*

To evaluate current knowledge on somatic mutations in *CDH1* in ST, the COSMIC and IntOGen bioinformatics tools were surveyed. The evaluation was done based on the 20 disease terms obtained from the output of DisGeNET (only curated databases) between *CDH1* and ST. Selected terms were: breast, endometrial, ovarian, pancreatic, prostate, stomach and urinary bladder neoplasms, as well as lobular and ductal breast carcinoma and squamous cell carcinoma. Since "colorectal carcinoma", was not listed in the COSMIC tool as such, the terms "colon" and "rectum" carcinoma found were grouped together and included in the analysis. In some cases, no results or an equivalent term corresponding to the DisGeNET output were found in the database cancer browser term.

As a result of this survey, 395 and 94 somatic mutations in *CDH1* were listed for selected ST in COSMIC and IntOGen databases, respectively. A further analysis of the COSMIC output revealed breast neoplasms as the ST with highest rate of *CDH1* mutations compared to the total mutations found (207/395). In addition, the IntOGen output confirmed that *CDH1* is considered a driver gene in breast cancer (p value $= 7.76 \times 10^{-13}$). Colorectal and stomach neoplasms also had a high rate of mutations in the *CDH1* gene (77/395, 19.5%, 59/395, 14.9%, respectively). Neoplasms of the endometrium, bladder, ovary, prostate and squamous cell carcinomas had the lowest reported *CDH1* somatic mutations (Table 2).

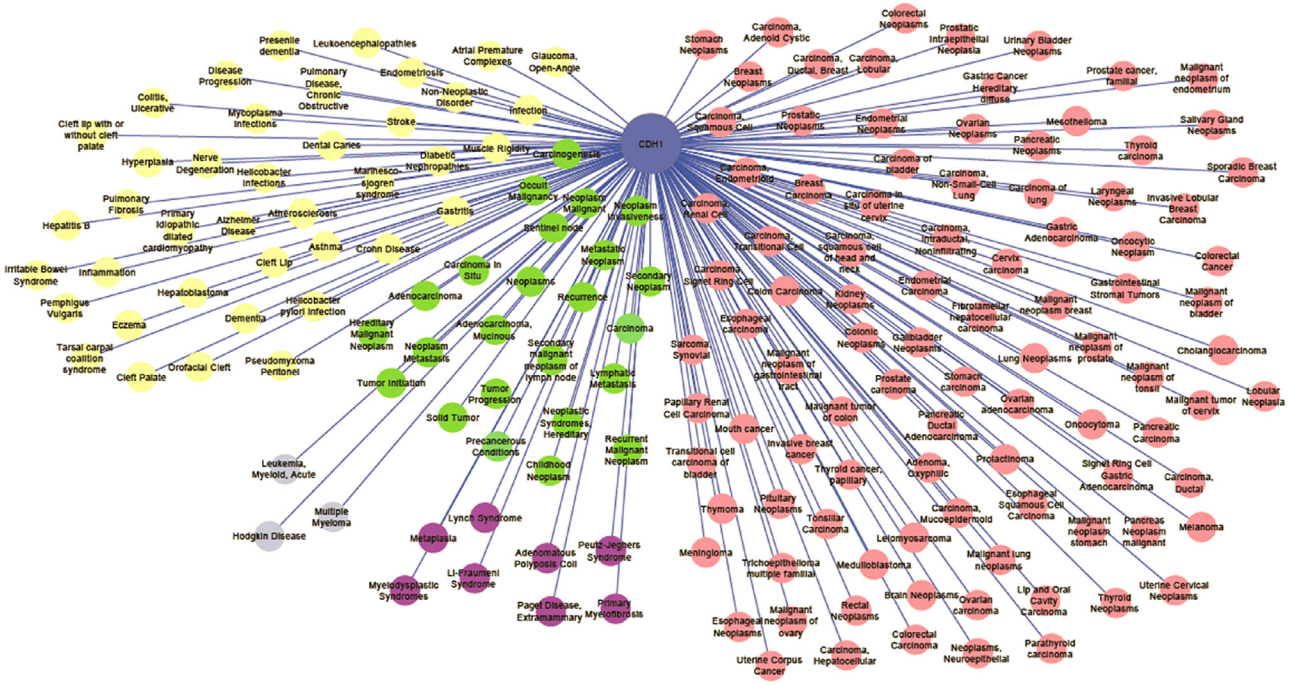

Fig. 4. Network representation of the Gene-Disease associations for *CDH1* obtained with DisGeNET ALL sources. From the 172 terms, 132 related to neoplasms (green, pink, purple and light-blue nodes) and 40 terms related to non-neoplasic diseases (yellow nodes). From the 132 neoplasms-related terms, 24 were tumor properties/neoplasm-related terms (green nodes), 8 were neoplastic processes (purple nodes) and 3 were non-solid tumors neoplasms (light-blue nodes) and 97 were ST (pink and yellow nodes). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Substitution-missense mutations (136/354, 38.4%) were predominant, followed by deletions to shift in the reading frame (67/354, 18.9%), unknown mutations (43/354, 12.1%) and insertions with shift in the reading frame (41/354, 11.6%) (Fig. 7).

In some tumor types, *CDH1* was positioned among the 20 genes with the highest mutation frequency. This was the case of breast neoplasms (3rd position out of 20). *CDH1* was ranked as second mutation carrier gene in lobular breast carcinomas (2nd of 20) and eleventh in ductal breast carcinoma. It was also found in the "top 20" in stomach (2nd of 20) and colorectal neoplasms (colon: 17th of 20, rectum: 19th of 20).

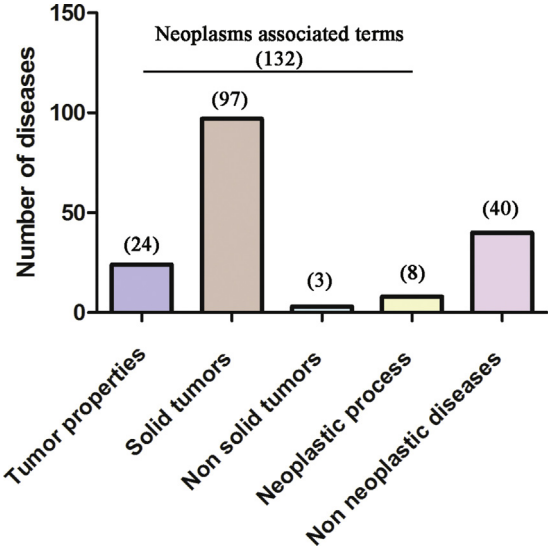

Fig. 5. *CDH1* related diseases and Association Types resulting from All DisGeNET databases. A graphical representation; detailed information is presented in Supplementary Table 2.

For some histological subtypes, *CDH1* mutations varied according to the organ in which the tumor originated. In squamous cell carcinomas, *CDH1* mutations were found in lung cancer, upper aerodigestive tract and urinary tract, but not in cervix, endometrium, stomach, intestine, esophagus, among others. Table 3 displays the search results on identified *CDH1* mutations in selected ST, total number of mutations found, mutation type, citation (PMID) and COSMIC link to find the full output of somatic mutations are indicated. While most of the mutations in *CDH1* were specific for each type of tumor, some mutations were shared by two or more ST in 5% of the cases. The mutation impact is included (Table 4). In addition to these results, data on CNV was retrieved from the COSMIC database and listed in Table 5.

To further explore changes in the *CDH1* sequence, a survey was conducted using the dbSNP database to obtain the number of reported SNPs for this gene. As a result of this evaluation, a total of 14,566 SNP were found. Using clinical significance filters, 222 SNPs were classified as benign (21 results), likely benign (17 results), likely pathogenic (4 results), pathogenic (20 results), uncertain significance (124 results), untested (9 results) and other (27 results) (data not shown). Less than 10% were assigned as of germinal origin.

4. Discussion

The growth of biomedical information in the last decades has been a positive force toward the development/implementation of integrated flexible text-mining systems. Computational methods for mining of biomedical literature are becoming indispensable to manage this flux of information. Bioinformatics tools have proven useful to browse in an organized and systematic fashion, available information in several biomedical databases to better understand biological processes (Harmston et al., 2010; Rebholz-Schuhmann

et al., 2012). In addition, they have shown to increase manual literature searches for associations between genes/proteins to study human diseases (Goh and Choi, 2012; Kann, 2010). Zhu et al. (2013) reported a substantial growth in the number of publications obtained from PubMed using "text-mining" as the query word in the title or abstract since 2000.

Among human pathologies, cancer is known to cause millions of deaths worldwide (Ferlay et al., 2015). Of all cancers, ST are the subject of studies aimed at improving diagnosis/treatment (Ascierto et al., 2014; Neal and Sledge, 2014). Several genes/proteins and mechanisms involved in this pathology have been identified, however it is still a disease far from being understood and effectively treated.

Researchers have already taken advantage of the text-mining technologies to discover novel knowledge to support biomedical research in cancer (Ahmed et al., 2011). Beck et al. (2014) reported advantages and limitations of applying bioinformatics tools to study epithelial-to-mesenchymal transition in lung cancer, and Banwait and Bastola (2015) discussed the role of bioinformatics in studying miRNAs in the context of human cancer.

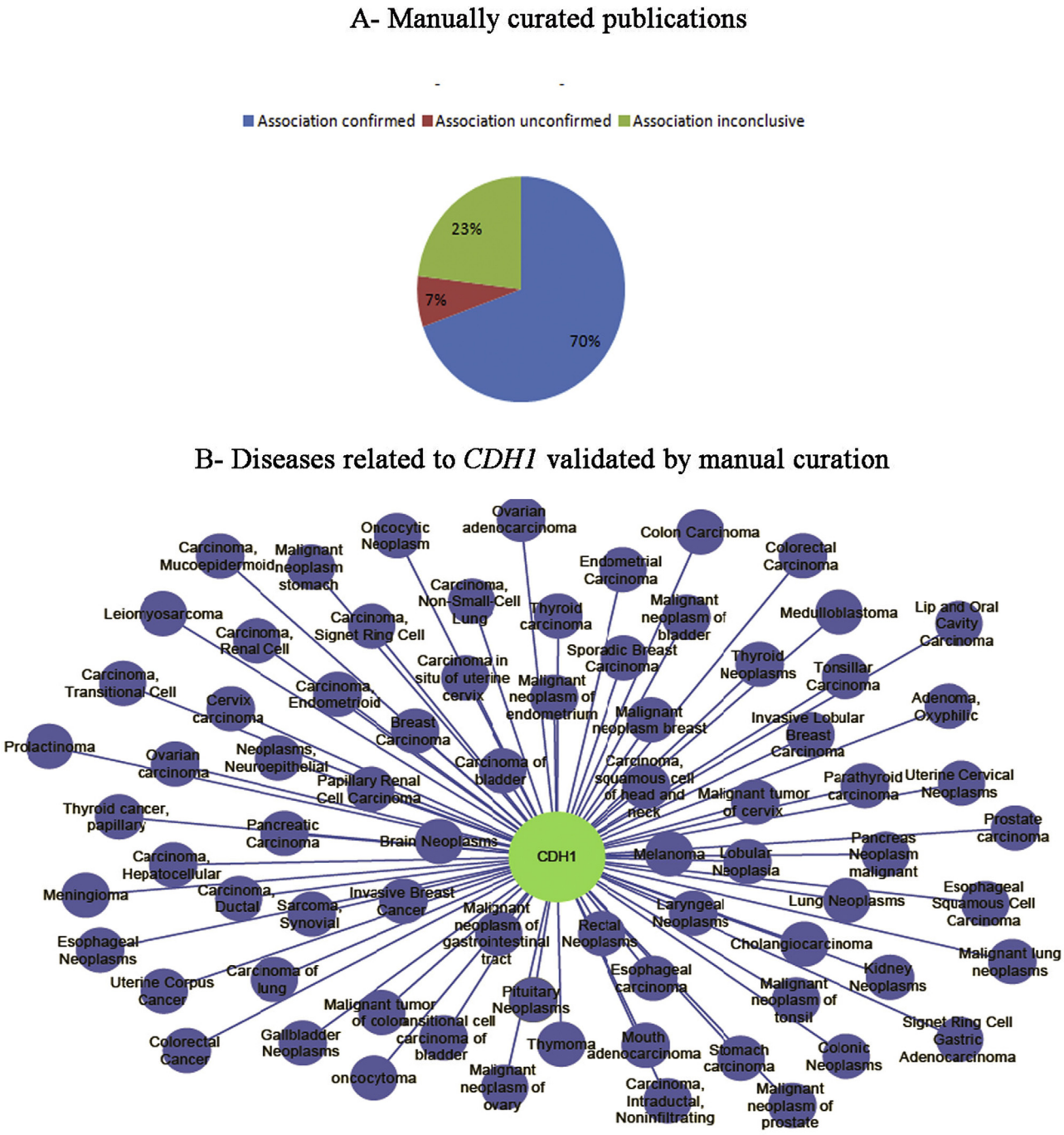

Fig. 6. Curation of disease terms associated to *CDH1* from DisGeNET. (A) Number of publications manually reviewed. From 286 publications that support the association between *CDH1* and Solid Tumors terms, 199/286 (70%) publications were confirmed, 67/286 (23%) publicationswere inconclusive, and the remaining 20 publications did not support the association between *CDH1* and Solid Tumors. (B) Seventy five diseases validated by manual curation. The 5 terms not validated as related to *CDH1* were Fibrolamellar hepatocellular carcinoma, gastric adenocarcinoma, gastrointestinal stromal tumors, pancreatic ductal adenocarcinoma and trichoepithelioma multiple familial.

Table 2
COSMIC and IntOGen output for *CDH1* somatic mutations.

Solid tumor	Number of *CDH1* somatic mutations (COSMIC/IntOgen/Both databases)	Percentage of *CDH1* mutations	Mutation frequency (Mutated samples/Tested samples)
Bladder neoplasms	12/7/3	3.0/7.5/13.0	COSMIC: 9/171 IntOgen: 6/98
Breast Neoplasms	207/72/10	52.4/76.6/43.5	
Ductal breast carcinoma	8/0/0	2.0/0/0	COSMIC: 215/1835
Lobular breast carcinoma	75/0/0	19.0/0/0	IntOgen: 75/1148
Colorectal neoplasms	77/1/1	19.5/1.1/4.4	COSMIC: Colon:64/443 Rectum: 14/154 IntOgen: Colon: 1/229
Endometrial neoplams	20/6/2	5.1/6.4/8.7	COSMIC: 20/343 IntOgen: 7/230
Ovarian neoplasms	4/1/1	1.0/1.1/4.4	COSMIC: 4/673 IntOgen: 1/316
Pancreatic neoplasms	3/1/1	0.8/1.1/4.4	COSMIC: 1/709 IntOgen: 1/214
Prostatic neoplasms	4/0/0	1.0/0/0	COSMIC: 3/493 IntOgen: 0
Squamous cell carcinoma	9/4/4	2.3/4.3/17.4	COSMIC: Lung: 3/246 Upper aerodigestive tract: 2/169 Urinary Tract: 1/2 IntOgen: Oropharynx: 9/375 Lung and bronchus: 7/665
Stomach neoplasms	59/2/1	14.9/2.1/4.4	COSMIC: 72/509 IntOgen: 2/22
Total number of mutations	395/94/23	100	–

Specifically regarding the tumor suppressor gene *CDH1*/E-cadherin, a large contribution of experimental data and scientific reports has greatly helped cancer prevention, diagnosis and treatment. However, data is overwhelming and increased worldwide in the last decade. Bioinformatics tools have recently highlighted E-cadherin as an important molecule in breast (Shargh et al., 2014) and gastric (Liu and Chu, 2014) cancer.

In the current study, the information offered by some of the *state-of-the-art* resources on human diseases and cancer (namely DisGeNET, COSMIC and IntOGen, dbSNP) were exploited to assess current knowledge on the relationship between *CDH1* and ST. Using DisGeNET, *CDH1*/E-cadherin and other genes associated to ST were identified; in addition, information about the relationship between *CDH1*/E-cadherin and ST was obtained, identifying a list of terms of disease-related properties as well as of specific diseases, and determining the association type between *CDH1* and each specific term.

Since a large amount of extracted data is predominantly from non-curated databases, a manual curation was done to assess the relevance of the information automatically extracted by text-mining. This evaluation resulted in a high percentage of validated associations, expanding the current *CDH1*-related disease terms in DisGeNET-curated databases. Moreover, these results represent an independent evaluation on the text-mining approach used to extract gene-disease associations by DisGeNET, at least for this specific use.

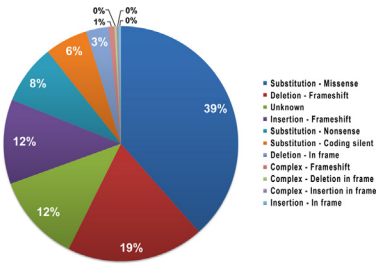

Fig. 7. *CDH1* somatic mutations in Solid Tumors. The percentage assigned to each mutation type was calculated considering as 100% the total *CDH1* mutations in Solid Tumors from the COSMIC output. Since some mutations are common in some Solid Tumors, the amount of total mutations is less than the number of mutations considered as 100% in Fig.

In addition to the studies summarized above, the result of a survey of annotated somatic mutations and SNP (germline, somatic, unknown) was included in this report. The advance of NGS technology has rapidly changed the approach to understanding cancer genomics (Forbes et al., 2010). These data contributes to our understanding of cancer causation and development, providing the foundation for prevention and treatment (Pleasance et al., 2010). In the present study, COSMIC and IntOGen databases were utilized to survey somatic *CDH1* mutations in several cancer types and over 350 mutations were identified in a group of ST. CNV data retrieved from COSMIC also revealed either gain or loss of *CDH1* copy number. Moreover, dbSNP listed over 14,000 SNP on *CDH1*.

The present report is the first showing the use of DisGeNET as a valuable tool to survey the gene-disease association between *CDH1*/E-cadherin and ST. Its use helped to find, extract, compile and mine the information available from the literature and categorize these data for the gene encoding the cell–cell adhesion protein. Data curation done expanded the number of terms that relate the disease-association between *CDH1* and ST. The survey of COSMIC and IntOGen databases rendered an updated list of somatic mutations reported for *CDH1* in several ST. Moreover, the survey done on dbSNP revealed a large amount of SNP reported for *CDH1*, some of which would have clinical impact. In any case, since it is well accepted that complex diseases such as cancer are multigenic, further investigations on *CDH1*-related genes participating in the progression of ST is warranted. DisGeNET has proven to be a useful bridge between basic research data and computational tasks, and this information can be used subsequently as a

Table 3
CDH1 somatic mutations. COSMIC output.

Solid tumor (number of mutations in *CDH1*)	Mutation type	PMIDs	Link out to COSMIC
Breast neoplasms (192)	Substitution—missense	22495314, 11857408, 20668451,	http://cancer.sanger.ac.uk/cosmic/gene/analysis?ln=CDH1&ss=all&sh=all&sn=breast&hn=all
	Deletion—frameshift	21512767	
	Deletion—In frame	7961105	
	Susbtitution—coding silent	8557030	
	Susbstitution—nonsense	8934538	
		10094558	
		10584868	
	Complex—insertion in frame	11291078	
	Complex—deletion in frame	12800196	
	Complex—frameshift	19191266	
	Unknown	23575477	
		11322170	
		15696125	
		19593635	
		9581841	
		11196175	
		22722201	
Colorectal neoplasms (81)	Substitution—nonsense	12096341	http://cancer.sanger.ac.uk/cosmic/gene/analysis?ln=CDH1&ss=colon&sh=all&sn=large_intestine&hn=all
	Substitution—missense	22810696	
	Substitution—nonsense	22895193	http://cancer.sanger.ac.uk/cosmic/gene/analysis?ln=CDH1&ss=rectum&sh=all&sn=large_intestine&hn=all
	Deletion—frameshift	23856246	
	Substitution—coding silent	22810696	
	Unknown		
Lobular breast carcinoma (75)	Substitution—missense	7961105	http://cancer.sanger.ac.uk/cosmic/gene/analysis?ln=CDH1&ss=all&sh=lobular_carcinoma&sn=breast&hn=carcinoma
	Deletion—frameshift	8557030	
	Deletion—in frame	8934538	
	Susbtitution—coding silent	10094558	
	Susbstitution—nonsense	10584868	
		11196175	
		11291078	
	Unknown	11857408	
		12800196	
		19191266	
		23575477	
Stomach neoplasms (59)	Substitution—missense	8033105	http://cancer.sanger.ac.uk/cosmic/gene/analysis?ln=CDH1&ss=all&sh=all&sn=stomach&hn=all
	Deletion—frameshift	8127895	
	Deletion—in frame	8797891	
	Susbtitution—coding silent	9045944	
	Susbstitution—nonsense	10094558	
		10211998	
		11058874	
	Unknown	11313896	
		11598162	
		16610016	
		17376510	
		22037554	
		22152101	
		23196062	
		23341533	
		22290393	
Endometrial neoplasms (20)		8075649	

Table 3 (*Continued*)

Solid tumor (number of mutations in *CDH1*)	Mutation type	PMIDs	Link out to COSMIC
	Substitution—missense		http://cancer.sanger.ac.uk/cosmic/gene/samples?src=gene&coords=AA%3AAA&end=883&ln=CDH1&ss=NS&sn=endometrium&id=924&seqlen=883&start=1
	Substitution—coding-silent		
	Deletion—frameshift		
	Unknown		
Urinary bladder neoplasms (10)	Substitution—missense	10891567	http://cancer.sanger.ac.uk/cosmic/gene/analysis?ln=CDH1&ss=bladder&sh=all&sn=urinary_tract&hn=all
	Substitution—coding silent	10891567	
	Unknown	23887298	
Breast ductal carcinoma (8)	Deletion—frameshift	11857408	http://cancer.sanger.ac.uk/cosmic/gene/analysis?ln=CDH1&ss=all&sh=ductal_carcinoma&sn=breast&hn=carcinoma
	Complex—frameshift	11196175	
	Substitution—missense	11322170	
	Deletion—in frame	11857408	
	Unknown		
Squamous cell carcinoma (6)	Substitution—missense	21798893	http://cancer.sanger.ac.uk/cosmic/gene/analysis?ln=CDH1&ss=all&sh=squamous_cell_carcinoma&sn=upper_aerodigestive_tract&hn=all
	Unknown	10891567	http://cancer.sanger.ac.uk/cosmic/gene/analysis?ln=CDH1&ss=all&sh=squamous_cell_carcinoma&sn=urinary_tract&hn=all
			http://cancer.sanger.ac.uk/cosmic/gene/analysis?ln=CDH1&ss=all&sh=squamous_cell_carcinoma&sn=lung&hn=all
Prostate neoplasms (3)	Substitution—missense	22610119	http://cancer.sanger.ac.uk/cosmic/gene/analysis?ln=CDH1&ss=&sh=&sn=prostate&hn=all
		22722839	
Ovarian neoplasms (4)	Substitution—missense	21720365	http://cancer.sanger.ac.uk/cosmic/gene/analysis?ln=CDH1&ss=all&sh=all&sn=ovary&hn=all
	Unknown	8075649	http://cancer.sanger.ac.uk/cosmic/browse/tissue#sn=ovary&ss=all&hn=all&sh=neoplasm&in=t&src=tissue
		23791828	
Pancreatic neoplasms (1)	Substitution—missense	No reported	http://cancer.sanger.ac.uk/cosmic/gene/samples?src=gene&coords=AA%3AAA&end=883&ln=CDH1&sn=pancreas&id=924&seqlen=883&start=1

The total number of *CDH1* mutations found in COSMIC is detailed. Left column: Number of mutations (parenthesis) in breast neoplasms, colorectal neoplasms, breast lobular carcinoma, stomach neoplasms, endometrial neoplasms, urinary bladder neoplasms, breast ductal carcinoma, squamous cell carcinoma, prostate neoplasms, ovarian neoplasm. Middle columns mutation type, and citation (PMID). Right column: COSMIC link to find the full output of somatic mutations in each case.
Note: PMIDs do not correspond to the mutations listed in the "Mutation type" column.

Table 4
Shared mutations found in two or more solid tumors.

Mutation	Mutation type	AA Change	Solid tumors that share the mutation	Mutation impact (COSMIC)
c.2537G > C	Substitution—missense	p.S846T	Bladder neoplasms, squamous cell carcinoma (urinary tract)	*Pathogenic*
c.67C > T	Substitution—nonsense	p.Q23*	Colorectal neoplasms, lobular breast carcinoma	*Neutral*
c.372delC	Deletion—frameshift	p.P126fs*89	Colorectal neoplasms, ductal breast carcinoma	*Neutral*
c.1711G > A	Substitution—missense	p.G571S	Colorectal neoplasms, lobular breast carcinoma	*Pathogenic*
c.1901C > T	Substitution—missense	p.A634V	Colorectal neoplasms, stomach neoplasms	*Pathogenic*
c.2195G > A	Substitution—missense	p.R732Q	Colorectal neoplasms, breast neoplasms, endometrial neoplasms	*Pathogenic*
c.2253C > T	Substitution—coding silent	p.N751N	Colorectal neoplasms, stomach neoplasms	*Pathogenic*
c.2512A > G	Substitution—missense	p.S838G	Colorectal neoplasms, ovarian neoplasms	*Pathogenic*
c.1105A > G	Substitution—missense	p.N369D	Colorectal neoplasms, stomach neoplasms	*Pathogenic*
c.1320 + 1G > C	Unknown	Unknown	Stomach neoplasms, breast neoplasms	*Pathogenic*
c.1008G > A	Substitution—coding silent	p.E336E	Stomach neoplasms, breast neoplasms	*Pathogenic*
c.1138_1320del183	Deletion—in frame	p.Y380K440del61,	Stomach neoplasms, ductal breast carcinoma	*Neutral*
c.1223C > T	Substitution—missense	p.A408V	Stomach neoplasms, lobular breast carcinoma	*Pathogenic*
c.1774G > A	Substitution—missense	p.A592T	Stomach neoplasms, breast neoplasms	*Pathogenic*
c.532-1G > T	Unknown	Unknown	Breast neoplasms, lobular breast carcinoma	*Pathogenic*
c.687 + 1_687 + 2delgt	Unknown	Unknown	Breast neoplasms, ductal breast carcinoma	*Neutral*
c.1009-2_1009-1delag	Unknown	Unknown	Breast neoplasms, lobular breast carcinoma	*Neutral*
c.1199A > G	Substitution—missense	p.D400G	Pancreatic neoplasms, colorectal neoplasms	*Pathogenic*
c.1849G > A	Substitution—missense	p.A617T	Endometrial neoplasms, colorectal neoplasms	*Neutral*

Table 5
CDH1—gene copy number variation in ST.

NEOPLASM	Tissue	Sub Tissue	Histology	Sub Histology	Copy Number Variation Variant%
Bladder neoplasms	Urinary tract	Bladder	Carcinoma	NS	Total Samples tested: 214 Sample(s) with CNV gain: 2 Total% of samples with gain: 0.93
Breast neoplasms	Breast	NS	Carcinoma	Ductal carcinoma	Total Samples tested: 28 Samples with CNV loss: 1 Total% of samples with loss: 3.57
Colorectal neoplasms	Large intestine	Rectum	Carcinoma	Adenocarcinoma	Total Samples tested: 153 Sample(s) with CNV gain: 1 Total% of samples with gain:0.65
Endometrial neoplasms	Endometrium	NS	Carcinoma	Endometrioid carcinoma	Total Samples tested: 355 Samples with CNV loss: 1 Total% of samples with loss: 0.28.
Ovarian neoplasms	Ovary	NS	Carcinoma	Serous carcinoma	Total Samples tested: 568 Samples with CNV loss: 9 Total% of samples with loss: 1.58
Prostatic neoplasms	Prostate	NS	Carcinoma	Adenocarcinoma	Total Samples tested: 288 Samples with CNV loss: 2 Total% of samples with loss: 0.69
Squamous cell carcinoma	Upper aerodigestive tract	Head neck	Carcinoma		Total samples tested: 432 Sample(s) with CNV gain: 3 Total% of samples with gain: 0.69
Stomach neoplasms	Stomach	NS	Carcinoma	Adenocarcinoma	Total samples tested: 336 Sample(s) with CNV gain: 1 Total% of samples with gain: 0.3

"starting point" for future bioinformatics analysis such as network analysis to identify gene-disease modules and novel biomarkers, pathways and therapeutic targets valuable for the development of cancer diagnosis, prognosis and treatment strategies. Such approaches will offer to the researcher complementary strategies to understand current and new findings about *CDH1*/E-cadherin in health and disease.

Acknowledgements

Studies related to the preparation of this manuscript were supported by grants from the National Agency to Promote Science and Technology (Agencia Nacional de Promoción de Ciencia y Tecnología, ANPCyT, PICT-SU-2012#1072), National Institute of Cancer (Instituto Nacional del Cáncer, INC, 2014-2015) and Roemmers Foundation (Fundación Roemmers, 2011-2013) to M.H.V.L.

L.I.F received support from Instituto de Salud Carlos III, Fondo Europeo de Desarrollo Regional (PI13/00082). The Research Programme on Biomedical Informatics (GRIB) is a node of the Spanish National Institute of Bioinformatics (INB).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.compbiolchem.2015.10.002.

References

Ahmed, J., Meinel, T., Dunkel, M., et al., 2011. Cancer resource: a comprehensive database of cancer-relevant proteins and compound interactions supported by experimental knowledge. Nucleic Acids Res. J. 39, D960–D967 (Database issue).
Alaimo, S., Giugno, R., Pulvirenti, A., 2014. ncRNA-Disease association prediction through tripartite network based inference. Front. Bioeng. Biotechnol. 2, 71.
Altman, R.B., 2012. Introduction to translational bioinformatics collection. PLoS Comput. Biol. 8, e1002796.
Angst, B.D., Marcozzi, C., Magee, A.I., 2001. The cadherin superfamily: diversity in form and function. J. Cell Sci. 114 (Pt 4), 629–641.
Ascierto, P.A., Addeo, R., Cartenì, G., et al., 2014. The role of immunotherapy in solid tumors: report from the Campania Society of Oncology Immunotherapy (SCITO) meeting, Naples 2014. J. Trans. Med. 12, 291.
Banwait, J.K., Bastola, D.R., 2015. Contribution of bioinformatics prediction in microRNA-based cancer therapeutics. Adv. Drug Deliv. Rev. 81C, 94–103.
Bauer-Mehren, A., Rautschka, M., Sanz, F., et al., 2010. DisGeNET: a Cytoscape plugin to visualize, integrate, search and analyze gene-disease networks. Bioinformatics 26, 2924–2926.
Bauer-Mehren, A., Bundschus, M., Rautschka, M., et al., 2011. Gene-disease network analysis reveals functional modules in mendelian, complex and environmental diseases. PLoS One 6, e20284.
Beck, T.N., Chikwem, A.J., Solanki, N.R., et al., 2014. Bioinformatic approaches to augment study of epithelial-to-mesenchymal transition in lung cancer. Physiol. Genomics 46, 699–724.
Berx, G., Staes, K., van Hengel, J., et al., 1995. Cloning and characterization of the human invasion suppressor gene E-cadherin (CDH1). Genomics 26, 281–289.
Berx, G., van Roy, F., 2009. Involvement of members of the cadherin superfamily in cancer. Cold Spring Harb. Perspect. Biol. 1, a003129.
Berx, G., Becker, K.F., Höfler, H., et al., 1998. Mutations of the human E-cadherin (CDH1) gene. Hum. Mutat. 12, 226–237.
Bravo, À., Cases, M., Queralt-Rosinach, N., et al., 2014. 2014. A knowledge-driven approach to extract disease-related biomarkers from the literature. Biomed. Res. Int. 253128.
Bravo, À., Piñero, J., Queralt, N., et al., 2015. Extraction of relations between genes and diseases from text and large-scale data analysis: implications for translational research. BMC Bioinf. 16, 55.
Bryan, R.T., 2015. Cell adhesion and urothelial bladder cancer: the role of cadherin switching and related phenomena. Philos. Trans. R. Soc. Lond. B. Biol. Sci 370, 1661.
Bryant, D.M., Stow, J.L., 2004. The ins and outs of E-cadherin trafficking. Trends Cell Biol. 14, 427–434.
Caldeira, J.R., Prando, E.C., Quevedo, F.C., et al., 2006. *CDH1* promoter hypermethylation and E-cadherin protein expression in infiltrating breast cancer. BMC Cancer 6, 48.
Chalmers, I.J., Aubele, M., Hartmann, E., et al., 2001. Mapping the chromosome 16 cadherin gene cluster to a minimal deleted region in ductal breast cancer. Cancer Genet. Cytogenet. 126, 39–44.
Conacci-Sorrell, M., Zhurinsky, J., Ben-Ze'ev, A., 2002. The cadherin–catenin adhesion system in signaling and cancer. J. Clin. Invest. 109, 987–991.
Cooper, G. 2000. The Cell 2nd Edition. A Molecular Approach. *Sunderland (MA)-Sinauer Associates*. Chapter 12.
Corso, G., Carvalho, J., Marrelli, D., et al., 2013. Somatic mutations and deletions of the E-cadherin gene predict poor survival of patients with gastric cancer. J. Clin. Oncol. 31, 868–875.
Corso, G., Figueiredo, J., Biffi, R., et al., 2014. E-cadherin germline mutation carriers: clinical management and genetic implications. Cancer Metastasis Rev. 33, 1081–1094.
Cowden Dahl, K.D., Symowicz, J., Ning, Y., et al., 2008. Matrix metalloproteinase 9 is a mediator of epidermal growth factor-dependent E-cadherin loss in ovarian carcinoma cells. Cancer Res. 68, 4606–4613.

Deng, Q.W., He, B.S., Pan, Y.Q., et al., 2014. Roles of E-cadherin (CDH1) genetic variations in cancer risk: a meta-analysis. Asian Pac. J. Cancer Prev. 15, 3705–3713.
Ferlay, J., Soerjomataram, I., Dikshit, R., et al., 2015. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. Int. J. Cancer 136, E359–86.
Forbes, S.A., Tang, G., Bindal, N., et al., 2010. COSMIC (the catalogue of somatic mutations in cancer): a resource to investigate acquired mutations in human cancer. Nucleic Acids Res. 38, D652–D657 (Database issue).
Galhardo, M., Sinkkonen, L., Berninger, P., et al., 2013. Integrated analysis of transcript-level regulation of metabolism reveals disease-relevant nodes of the human metabolic network. Nucleic Acids Res. 42 (3), 1474–1496.
Gall, T.M., Frampton, A.E., 2013. Gene of the month E-cadherin (CDH1). J. Clin. Pathol. 66, 928–932.
Gheldof, A., Berx, G., 2013. Cadherins and epithelial-to-mesenchymal transition. Prog. Mol. Biol. Transl. Sci. 116, 317–336.
Goh, K.I., Choi, I.G., 2012. Exploring the human diseasome: the human disease network. Brief Funct. Genomics 11, 533–542.
Gonzalez-Perez, A., Perez-Llamas, C., Deu-Pons, J., et al., 2013. IntOGen-mutations identifies cancer drivers across tumor types. Nat. Methods 10, 1081–1082.
Grosdidier, S., Ferrer, A., Faner, R., et al., 2014. Network medicine analysis of COPD multimorbidities. Respir. Res. 15 (1), 111.
Hanahan, D., Weinberg, R.A., 2011. Hallmarks of cancer: the next generation. Cell 144, 646–674.
Harmston, N., Filsell, W., Stumpf, M.P., 2010. What the papers say: text mining for genomics and systems biology. Hum. Genomics 5, 17–29.
Hiraguri, S., Godfrey, T., Nakamura, H., et al., 1998. Mechanisms of inactivation of E-cadherin in breast cancer cell lines. Cancer Res. 58, 1972–1977.
Hong, S., Troyanovsky, R.B., Troyanovsky, S.M., 2013. Binding to F-actin guides cadherin cluster assembly, stability, and movement. J. Cell Biol. 201, 131–143.
Howe, D., Costanzo, M., Fey, P., et al., 2008. Big data: the future of biocuration. Nature 455, 47–50.
Ivanov, A.I., Naydenov, N.G., 2013. Dynamics and regulation of epithelial adherens junctions: recent discoveries and controversies. Int. Rev. Cell. Mol. Biol. 303, 27–99.
Kanazawa, T., Watanabe, T., Kazama, S., et al., 2002. Poorly differentiated adenocarcinoma and mucinous carcinoma of the colon and rectum show higher rates of loss of heterozygosity and loss of E-cadherin expression due to methylation of promoter region. Int. J. Cancer 102, 225–2259.
Kann, M.G., 2010. Advances in translational bioinformatics: computational approaches for the hunting of disease genes. Brief Bioinform. 11, 96–110.
Kemler, R., 1992. Classical cadherins. Semin. Cell Biol. 3, 149–155.
Le Bras, G.F., Taubenslag, K.J., Andl, C.D., 2012. The regulation of cell–cell adhesion during epithelial-mesenchymal transition, motility and tumor progression. Cell Adh. Migr. 6, 365–373.
Liu, X., Chu, K.M., 2014. E-cadherin and gastric cancer: cause, consequence, and applications. Biomed. Res. Int. 637308.
Mosesson, Y., Mills, G.B., Yarden, Y., 2008. Derailed endocytosis: an emerging feature of cancer. Nat. Rev. Cancer 8, 835–850.
Nagafuchi, A., Tsukita, S., Takeichi, M., 1993. Transmembrane control of cadherin-mediated cell–cell adhesion. Semin. Cell Biol. 4, 175–181.
Neal, J.W., Sledge, G.W., 2014. Decade in review-targeted therapy: successes, toxicities and challenges in solid tumours. Nat. Rev. Clin. Oncol. 11, 627–628.
Negrini, S., Gorgoulis, V.G., Halazonetis, T.D., 2010. Genomic instability—an evolving hallmark of cancer. Nat. Rev. Mol. Cell. Biol. 11, 220–228.
Nishimura, T., Takeichi, M., 2009. Remodeling of the adherens junctions during morphogenesis. Curr. Top Dev. Biol. 89, 33–54.
Nose, A., Tsuji, K., Takeichi, M., 1990. Localization of specificity determining sites in cadherin cell adhesion molecules. Cell 61, 147–155.
Oda, H., Takeichi, M., 2011. Evolution: structural and functional diversity of cadherin at the adherens junction. J. Cell Biol. 193, 1137–1146.
Ozawa, M., Kemler, R., 1998. The membrane-proximal region of the E-cadherin cytoplasmic domain prevents dimerization and negatively regulates adhesion activity. J. Cell Biol. 142, 1605–1613.
Ozawa, M., Engel, J., Kemler, R., 1990. Single amino acid substitutions in one Ca^{2+} binding site of uvomorulin abolish the adhesive function. Cell 63, 1033–1038.
Pastrello, C., Pasini, E., Kotlyar, M., et al., 2014. Integration, visualization and analysis of human interactome. Biochem. Biophys. Res. Commun. 445 757–573.
Perez-Llamas, C., Gundem, G., Lopez-Bigas, N., 2011. Integrative cancer genomics (IntOGen) in Biomart. Database (Oxford) bar039.
Perez-Moreno, M., Jamora, C., Fuchs, E., 2003. Sticky business: orchestrating cellular signals at adherens junctions. Cell 112, 535–548.
Piñero, J., Queralt-Rosinach, N., Bravo, À., et al., 2015. DisGeNET: a discovery platform for the dynamical exploration of human diseases and their genes. Database (Oxford) bav028.
Pleasance, E.D., Cheetham, R.K., Stephens, P.J., et al., 2010. A comprehensive catalogue of somatic mutations from a human cancer genome. Nature 463, 191–196.
Rashid, M.G., Sanda, M.G., Vallorosi, C.J., et al., 2001. Posttranslational truncation and inactivation of human E-cadherin distinguishes prostate cancer from matched normal prostate. Cancer Res. 61, 489–492.
Rebholz-Schuhmann, D., Oellrich, A., Hoehndorf, R., 2012. Text-mining solutions for biomedical research: enabling integrative biology. Nat. Rev. Genet. 13, 829–839.
Santiago, J.A., Potashkin, J.A., 2014. System-based approaches to decode the molecular links in Parkinson's disease and diabetes. Neurobiol. Dis. doi:http://dx.doi.org/10.1016/j.nbd.2014.03.019.
Sarkar, I.N., Butte, A.J., Lussier, Y.A., et al., 2011. Translational bioinformatics: linking knowledge across biological and clinical realms. J. Am. Med. Inform. Assoc. 18, 354–357.
Schildberg, C.W., Abba, M., Merkel, S., et al., 2014. Gastric cancer patients less than 50 years of age exhibit significant downregulation of E-cadherin and CDX2 compared to older reference populations. Adv. Med. Sci. 59, 142–146.
Schneider, M.R., Kolligs, F.T., 2014. 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 doi:http://dx.doi.org/10.1002/bies.201400141.
Shargh, S.A., Sakizli, M., Khalaj, V., et al., 2014. Downregulation of E-cadherin expression in breast cancer by promoter hypermethylation and its relation with progression and prognosis of tumor. Med. Oncol. 31, 250.
Sherry, S.T., Ward, M., Sirotkin, K., 1999. db SNP-database for single nucleotide polymorphisms and other classes of minor genetic variation. Genome Res. 9, 677–679.
Shihab, H.A., Gough, J., Cooper, D.N., et al., 2013. Predicting the functional, molecular and phenotypic consequences of amino acid substitutions using hidden Markov models. Hum. Mutat. 34, 57–65.
Sinn, H.P., Helmchen, B., Heil, J., et al., 2014. Lobular neoplasms and invasive lobular breast cancer. Pathologe 35, 45–53.
Takeichi, M., 1977. Functional correlation between cell adhesive properties and some cell surface proteins. J. Cell Biol. 75, 464–474.
Valente, A.L., Rummel, S., Shriver, C.D., et al., 2014. Sequence-based detection of mutations in cadherin 1 to determine the prevalence of germline mutations in patients with invasive lobular carcinoma of the breast. Hered. Cancer Clin. Pract. 12, 17.
van Roy, F., Berx, G., 2008. The cell–cell adhesion molecule E-cadherin. Cell Mol. Life Sci. 65, 3756–3788.
van Roy, F., 2014. Beyond E-cadherin: roles of other cadherin superfamily members in cancer. Nat. Rev. Cancer 14, 121–134.
Vasioukhin, V., 2012. Adherens junctions and cancer. Subcell. Biochem. 60, 379–414.
Wójcik-Krowiranda, K., Forma, E., Zaczek, A., et al., 2013. Expression of E-cadherin and beta1-integrin mRNA in endometrial cancer. Ginekol. Pol. 84, 910–914.
Zhang, P., Hu, P., Shen, H., et al., 2014. Prognostic role of twist or snail in various carcinomas: a systematic review and meta-analysis. Eur. J. Clin. Invest. 44, 1072–1094.
Zhu, F., Patumcharoenpol, P., Zhang, C., et al., 2013. Biomedical text mining and its applications in cancer research. J. Biomed. Inform. 46, 200–211.