Transcription initiation arising from E-cadherin/CDH1 intron2: a novel protein isoform that increases gastric cancer cell invasion and angiogenesis

Hugo Pinheiro1, Joana Carvalho1,8, Patrícia Oliveira1,9, Daniel Ferreira1,9, Marta Teixeira Pinto1, Hugo Osório1, Danilo Licastro2, Renata Bordeira-Carriço1, Peter Jordan3, Dejan Lazarevic2, Remo Sanges4, Elia Stupka5, David Huntsman6, Raquel Seruca1,7 and Carla Oliveira1,7,*

1IPATIMUP-Institute of Molecular Pathology and Immunology, University of Porto, Porto 4200-465, Portugal, 2CBM S.c.r.l. Strada Statale 14 - km 163, 5 AREA Science Park, Basovizza, Trieste 34149, Italy, 3Centre of Human Genetics, National Health Institute Dr Ricardo Jorge, Lisbon, 1649-016, Portugal, 4Stazione Zoologica Anton Dohrn, Villa Comunale, Napoli 80121, Italia, 5Center for Translational Genomics and Bioinformatics, San Raffaele Scientific Institute, Via Olgettina 58, Milano 20132, Italy, 6British Columbia Cancer Agency (BCCA), Vancouver, Canada V5Z 4E6 and 7Faculty of Medicine, University of Porto, Porto 4200-465, Portugal

Received April 27, 2012; Revised and Accepted June 22, 2012

Disruption of E-cadherin (CDH1 gene) expression, subcellular localization or function arises during initiation and progression of almost 90% of all epithelial carcinomas. Nevertheless, the mechanisms through which this occurs are largely unknown. Previous studies showed the importance of CDH1 intron 2 sequences for proper gene and protein expression, supporting these as E-cadherin cis-modulators. Through RACE and RT-PCR, we searched for transcription events arising from CDH1 intron 2 and discovered several new transcripts. One, named CDH1a, with high expression in spleen and absent from normal stomach, was demonstrated to be translated into a novel isoform, differing from canonical E-cadherin in its N-terminal, as determined by mass spectrometry. Quantitative and functional assays showed that when overexpressed in an E-cadherin negative context, CDH1a replaced canonical protein interactions and functions. However, when co-expressed with canonical E-cadherin, CDH1a increased cell invasion and angiogenesis. Further, interferon-induced gene IFITM1 and IFI27 levels were increased upon CDH1a overexpression. Effects on invasion and IFITM1 and IFI27 expression were reverted upon CDH1a-specific knockdown. Importantly, CDH1a was de novo expressed in gastric cancer cell lines. This study presents a new mechanism by which E-cadherin functions are impaired by cis-regulatory mechanisms possibly with the involvement of inflammatory machinery. If confirmed in other cancer models, our data enclose potential for designing targeted therapies to rescue E-cadherin function.

INTRODUCTION

E-cadherin, a protein encoded by the CDH1 gene [ENSG00000039068] is the dominant epithelial cell-adhesion molecule and plays a crucial role in epithelial tissue polarity and structural integrity (1,2). Reduced cell–cell adhesiveness allows cancer cells to disobey their local (social) order, resulting in the destruction of histological structure; the
most prominent morphological hallmark of malignant tumors (3). The most clinically relevant point in the progression of 90% or more carcinomas is believed to be mediated by disruption of normal E-cadherin expression, subcellular localization or function (4,5). Classical gene inactivation (mutation, gene loss and promoter hypermethylation), transcriptional and post-transcriptional mechanisms (transcription repressors, RNA and protein quality control) have all been associated with E-cadherin loss and/or deregulation of its localization and function, in a wide range of epithelial tumors (6–12). Despite the strong correlation between E-cadherin loss and malignancy, the mechanism through which this occurs is not known for most sporadic and hereditary epithelial carcinomas.

Although mutations and deletions of CDH1 remain the unique germline defects described in 47% of hereditary diffuse gastric cancer (HDGC; OMIM No. 137215) (13,14), more than 70% of all HDGC families present germline mono-allelic expression imbalance at the RNA level (10), caused by so far undetected mechanisms. The later observation is consistent with the overall E-cadherin protein expression defects found in most HDGC tumors (13) and pinpoint the key role of CDH1 in this hereditary syndrome. The scenario in sporadic gastric cancers is somewhat similar, as ~90% of the cases present aberrant or absent E-cadherin protein expression and unequivocal gene inactivating mechanisms occur only in 30% of the cases (15).

While seeking for mechanisms regulating CDH1 expression during mouse development, Stemmler et al. (16,17) showed that murine CDH1 intron 2 deletions interfere with gene transcriptional activation and expression in a tissue-specific manner, strongly supporting the existence of sequences, within those regions, that may act as cis-modulators of E-cadherin expression (18). Thus, there are unaccounted mechanisms for CDH1 inactivation in sporadic and inherited tumors, potentially involving intronic-dependent regulation of this locus.

The specificity and complexity of gene expression patterns in cells and tissues is achieved not only by increases and decreases in expression levels of cell-specific genes, but also through alternative splicing, alternative transcription initiation, alternative polyadenylation and RNA editing (19). Recent uses of RNA sequencing (RNA-seq) have been pivotal to unveil transcript quantification, assess alternative splicing and detect novel gene structures. This overwhelming and increasing diversity on mRNA population occurs because 95–100% of all human pre-mRNAs, which contain sequences corresponding to more than one exon, are processed to yield multiple mRNAs (20,21). Several of these new transcripts are known to regulate the canonical RNA form by many different mechanisms (22,23). As an example, a recent paper suggests that an alternative PTK6 transcript is able to negatively regulate growth and modulate PTK6 activity, protein–protein associations and/or subcellular localization (24). To date, similar findings have not been reported for CDH1, as no transcripts alternative to the canonical are described, besides those resulting from exon skipping events in cancerous samples. Herein, we describe transcripts arising from within CDH1 intron 2, and address their putative role as modulators of E-cadherin expression and function in cancer cells.

RESULTS

In the present work, we intended to understand whether novel long transcripts arise from within CDH1 intron 2 to characterize their expression pattern and to ascribe their putative role as modulators of E-cadherin expression, localization and function.

CDH1 locus gives rise to several transcripts in addition to the canonic

Our working hypothesis was that novel exons overlapping with annotated expressed sequence tags (ESTs) within intron 2 could be initiating exons of new CDH1 transcripts. To address this, we have analyzed CDH1 expression upstream and downstream of intron 2 in several normal human tissues by real-time PCR and using TaqMan assays covering exon 1–2 and exon 6–7 borders. Interestingly, the expression ratio (exons 6–7)/(exons 1–2) was higher than 1 for all normal tissues analyzed suggesting increased transcription levels downstream of CDH1 intron 2 (Fig. 1). This increased transcription was the highest for peripheral blood lymphocytes (PBLs) followed by spleen, where exon 6–7 probe expression was 2-fold higher when compared with exon 1–2 levels.

The information from Public databases (UCSC Genome Browser http://genome.ucsc.edu/ and Ensembl http://www.ensembl.com) was used to retrieve the available information on the CDH1 locus. Four transcripts are currently annotated at Ensembl database for CDH1: the canonic transcript and other three transcripts either including one additional canonical exon or excluding one or three canonical exons. Importantly for the present work, none of the annotated transcripts encompass intron 2 sequences. Several overlapping ESTs are annotated for CDH1 intron 2, potentially indicating that transcription from this locus could be more complex than previously anticipated (Fig. 2). As spleen was one of the tissues with higher (exon 6–7)/(exon 1–2) expression ratio, we used commercial spleen RNA to test the existence of CDH1 transcripts encompassing any of the selected ESTs, and their splicing with further downstream CDH1 exons. Among ESTs tested, we identified two new exons, encoded by intron 2 sequences, each splicing with exon 3 at its canonical splice site (Fig. 2). These two novel transcripts were called CDH1a and CDH1b (Fig. 2). Moreover, when amplifying CDH1a, we systematically verified the appearance of a higher molecular weight band that, upon sequencing, revealed to be another transcript, named CDH1j. A similar scenario was verified for CDH1b, where another splice site was identified 10 bp upstream of the one initially recognized. This variant was coined CDH1b-10 (Fig. 2). All four new transcripts were found to splice with the canonical exon 3 splice site and to include all downstream CDH1 canonical exons (Fig. 2), as determined by primer-walking PCR (data not shown). Additionally, only CDH1a and CDH1b-10 were found to have in-frame initiation codons (AUG). CDH1a presents a competent Kozak sequence upstream of the AUG (data not shown) in contrast to the one from CDH1b-10. We have also assessed the polyadenylation status of these transcripts and inquired about the DNA strand from which they were derived. CDH1a and CDH1j were found to be polyadenylated
In order to determine the transcription start site (TSS) of novel CDH1 transcripts, we used spleen and stomach 5’-RACE-ready cDNA libraries (built from polyadenylated RNA). These tissues were chosen for this purpose because spleen, as a non-epithelial tissue, is described to lack canonical E-cadherin expression and presented a high (exon 6–7)/(exon 1–2) CDH1 expression ratio, indicating the presence of additional transcripts. Differently, in stomach, an epithelial tissue, E-cadherin canonical expression is well recognized and was shown to have a low (exon 6–7)/(exon 1–2) expression ratio (Fig. 1). Using this strategy, we observed that, in spleen, CDH1a and CDH1j represented more than half of the total polyadenylated RNA molecules amplified. Additionally, we observed the existence of two additional transcripts, one starting in intron 1 (CDH1Seq1) and other in exon 3 (CDH1Seq5, which probably corresponds to the Ensembl annotated transcript, ENST00000422392—Ensembl release 61—February 2011). As expected, the canonical CDH1 transcript was not detected in spleen (Fig. 2). For stomach, the canonical transcript was the main RNA form found (45%), although transcripts starting in intron 1 (CDH1Seq1) and exon 3 (CDH1Seq11-13), similar to those identified in spleen, were also identified in stomach, the remaining CDH1a, CDH1j, CDH1b and CDH1b-10 were not found (Fig. 2 and Supplementary Material, Fig. S1).

For CDH1a, two 5’-sequences upstream of the same in-frame AUG were recurrently identified (CDH1Seq2 and CDH1Seq3 in Fig. 2) and therefore assumed as two different TSSs. The presence of these was further confirmed when its sequence was compared with CAGE data extracted from the Encode Project database at UCSC genome browser (http://genome.ucsc.edu/). For CDH1j transcript, the right location of potential TSSs was harder to infer, due to high variation in the length of sequences obtained by 5’-RACE. This transcript is most probably a non-coding RNA, due to the lack of identifiable AUG with a long open-reading frame and Kozak sequence. CDH1b and CDH1b-10 transcripts, shown earlier to be non-polyadenylated, were not found in this 5’-RACE experiment (Fig. 2).

Figure 1. The CDH1 qRT–PCR expression ratio using two probes targeting different gene regions, exon 6/7 and exon 1/2 boundaries. Normal tissues were tested and all presented higher expression levels with the downstream probe indicating the existence of transcripts encompassing exons 6-7 but excluding exons 1–2.

### The sequence of novel CDH1 exons overlaps or is in close vicinity of genomic regulatory features and conserved non-coding sequences

Approximately 15% of all CDH1 transcripts in spleen and 60% in stomach start within or upstream of CDH1 intron 1. These transcripts, together with CDH1Seq10 from stomach (Fig. 2), overlap a CpG island that is well known to modulate the expression of the canonical transcript as first described by Berx et al. (2). Thus, similarly to the canonical transcript, the expression of these novel transcripts may also be influenced by methylation at this CpG island. Additionally, a First Exon (EF) element is predicted to overlap the initial exon of the canonical transcript (exclusive of stomach), and the novel exon encoded by a portion of intron 1 belonging to a transcript that is present in both tissues studied (Fig. 2). The presence of this element and its proximity to the latter novel exon encoded by intron 1 (CDH1Seq1) indicates that this exon is likely the first exon of a novel CDH1 transcript.

Several DNAse I hypersensitive sites, which are commonly associated with new areas of transcription and/or gene regulation, were annotated in the vicinity of CDH1a and CDH1b sequences (25) (Fig. 2). Moreover, an AluSc repeat was found to overlap the sequence of the longer form of CDH1a that encodes the CDH1j transcript, likely indicating an Alu-mediated exonization event (26) (Fig. 2).

At the beginning of the canonical exon 3, a predicted CCCTC-binding factor (CTCF) binding site was also found, what may be important to explain the percentage of RNA transcripts starting immediately downstream of this site in stomach (~40%) and in spleen (~15%) (Fig. 2).

Figure 2 was constructed based on several tracks of UCSC, the Ensembl Genome Browser and our own data and depicts in detail the data described above.

We also investigated the extent of mammalian sequence conservation for the CDH1a and CDH1j transcripts identified. We verified the nucleotide level GERP score (27) for each transcript, focusing only on the novel portion (i.e. the part of the transcript that had not been previously reported). As shown in Supplementary Material, File S1, CDH1a transcripts appear to be evolutionarily constrained, with mammalian sequence conservation across most of their sequence in line with that of UTR portions of genes (median GERP score 1.72, max GERP score = 3.67 for both the longer CDH1a_70 transcript and shorter CDH1a_34 transcript). CDH1j, on the other hand, shows a very different conservation pattern. The median GERP score indicates that there is no long-range evolutionary constraint across the transcript, but the sliding window analysis indicates clearly that the final portion of the CDH1j exon is highly conserved (nt 320–352), which has a median GERP score of 2.3. Interestingly, the highly conserved portion also corresponds to the DNAase I hypersensitivity site.
CDH1a expression is tissue specific

We characterized CDH1a in detail as, unlike the other CDH1 novel transcripts, it has cumulatively a polyadenylated open-reading frame encoding a novel 5'-exon transcribed from within intron 2, encompasses two possible TSSs, an adequate Kozak consensus sequence upstream of the AUG, and splices with the canonical exon 3, sharing all downstream exons with the CDH1 canonical transcript. This RNA transcript encloses most of the necessary features to be translated into a protein isoform and is also particularly interesting because it is the most abundant CDH1 transcript in spleen (~50%) and is completely absent in stomach, according to the 5'-RACE results.

Using PCR followed by Q-SnapShot (10), we addressed CDH1a pattern of expression in several normal tissues. We have confirmed high CDH1a expression for spleen and its absence in stomach (Fig. 3A), contrasting with the scenario obtained for CDH1 canonical transcript (Fig. 2). Although highest CDH1a expression levels have been found for spleen, other normal tissues displayed variable expression of the transcript. Interestingly, the expression of CDH1a in PBLs was not comparable to that observed when measuring the (exon 6–7)/(exon 1–2) expression ratio depicted in Figure 1 which indicates that this is likely due to the expression of other transcripts starting downstream of exon 2 in PBLs.

Given that CDH1a mRNA levels were high for spleen, no canonical transcript was present and no other open-reading frames were identified by 5'-RACE, we reasoned that if a protein isoform was being produced from this transcript, it would be caught by immunohistochemistry using an antibody against any portion of the canonical protein, encoded by exons...
downstream of exon 3. For that, we stained both normal human spleen and stomach sections with an E-cadherin anti-body that recognizes the protein (E-cadherin and putative CDH1a isoform) cytoplasmic domain. In spleen, we observed a very specific and localized staining pattern (Fig. 3B), in contrast to the typical staining of the canonical E-cadherin at the basolateral surface of adjacent cells in gastric glands (Fig. 3C). Although fairly similar to small blood vessels in aspect, the stained structures observed in spleen showed no co-localization with CD34, a marker of endothelial cells (Supplementary Material, Fig. S1). Together, these observations support that CDH1a is translated and encodes a novel E-cadherin isoform in spleen.

Exogenous overexpression of CDH1a generates a protein that is less efficiently processed to a mature form than canonical E-cadherin

In order to assess CDH1a protein isoform translation and potential processing, we cloned the full transcript sequence from the AUG to the stop codon in a Lentiviral-derived expression vector (pLenti, Invitrogen). Empty vector (Mock) and the vector carrying canonical CDH1 cDNA (E-cadherin) were also generated as negative and positive controls, respectively. The Chinese Hamster Ovary (CHO) cell line was chosen to be transduced due to its complete lack of E-cadherin expression. The predicted amino acidic differences between E-cadherin and CDH1a sequences are depicted in Figure 4A. The canonical sequence MetGPWSRSLSA...RHLERGRVLGR transcribed from the exon 1 and 2 sequences is replaced by MetKLKLSRKQIQGHDKAASVSL, the amino acids encoded by the new exon 1a. Through western blot and using an antibody capable of detecting both the canonical and also CDH1a isoforms (Fig. 4A), we were able to prove the effectiveness of transduction and to confirm that, upon lentiviral promoter influence, CDH1a is efficiently translated into a protein of approximated size to that of the canonical E-cadherin (Fig. 4B). As expected, the empty vector (Mock) cells did not induce E-cadherin expression.

Since the immature form of E-cadherin is predictably processed to its mature form by cleavage of the 154-amino acid-residue precursor sequence, we analyzed whether CDH1a isoform would be cleaved in a similar way. Using a drug that blocks protein transfer to the Golgi complex and its processing (brefeldin A) (28), we verified for both (CDH1a isoform and E-cadherin), bands corresponding to the mature proteins displayed comparable molecular weight, whereas the immature proteins did not. This result confirms that the CDH1a pre-protein is smaller than the canonical E-cadherin and further that the two proteins are cleaved, if not at the same site, in very close proximity (Fig. 4C).

To confirm the relative amount of mature and immature proteins between cells expressing the canonical or the CDH1a protein isoform, we proceeded to protein sequencing by mass spectrometry using bands extracted from the gel (Fig. 4E). Protein lysates were immunoprecipitated using an anti-E-cadherin antibody and the resulting eluate was separated by 1D SDS–PAGE from both CHO E-cadherin and CHO CDH1a expressing cell lines. For each cell line, two bands with different molecular weights were observed, possibly related to E-cadherin immature (higher molecular weight) and mature (lower molecular weight) forms as seen in Figure 4C. For CHO E-cadherin, both bands were identified by MALDI-TOF/TOF mass spectrometry as E-cadherin with a CI of 100% (Supplementary Material, Table S2). Thirty peptides were found to be associated with E-cadherin (UniProt accession ID P12830) from CHO E-cadherin, both in higher and lower molecular weight mass.

CDH1a bands also presented 19 peptides associated with E-cadherin (UniProt accession ID P12830) that were
Figure 4. CDH1a is translated, in vitro, into an E-cadherin isoform that interacts with β- and p120 catenins at the membrane of CHO (E-cadherin-negative) cells and elicits cell aggregation and invasion suppression. (A) Schematic representation of the canonic protein and from the putative CDH1a isoform. In grey and blue areas are the specific regions from both proteins, respectively. The red and the white areas correspond to the common part that is predicted to be cleaved during protein processing that occurs at the site marked by a grey line. The black area corresponds to the expected mature proteins and the asterisk marks the epitope recognized by the antibody used. (B) CDH1a mature protein has a similar size to the canonic E-cadherin. (C) CDH1a immature form is, as expected, smaller than the canonic protein as exons 1 and 2 are replaced by exon 1a. (D) CDH1a is able to bind to β-catenin and p120ctn, components of the adhesion complex as happens with the canonic E-cadherin protein. (E) Coomassie blue-stained gel showing E-cadherin and CDH1a processed and unprocessed bands. (F and G) CDH1a isoform recruits and co-localizes with β-catenin and p120ctn at the cell membrane as occurs for E-cadherin. No staining is observed for mock-transfected cells. Scale bar: 15 μm. (H). CDH1a isoform is able to elicit cell–cell aggregation as for E-cadherin, in contrast to the empty vector (mock) transduced cells, Scale bar: 200 μm. (I) CDH1a confers statistically significant strong invasion suppression capacity as the canonic protein when compared with mock cells (P = 0.026 and 0.019, respectively).
common to the canonical E-cadherin. In addition, an extra peak was identified with a mass of 1722.85 Da that was absent for CHO E-cadherin (Supplementary Material, Fig. S2). To verify that this novel peak was associated with CDH1a N-terminus, we downloaded the Swiss-Prot/UniProt protein sequence database and manually inserted the full CDH1a sequence. A Mascot Peptide Mass FingerPrint (PMF) analysis successfully associated this 1722.85 Da peak as well as identified a smaller one with 953.52 Da with CDH1a sequence, corresponding, respectively, to the N-terminal peptides AAASLVNFEDC+TGR 16-31 (C+, carbamidomethylation of cysteine) and KQIQHGDK 8-15 (Supplementary Material, Table S2). To further validate this result, the 1722.85 Da peak, which presented the best signal-to-noise ratio, was subjected to MS/MS peptide sequencing followed by a PMF + MS/MS combined analysis. A Mascot ion score of 99.9% was obtained for this peptide (Supplementary Material, Table S2). Peptide sequencing was not possible for the smaller peak (953.52 Da); nevertheless, its presence is consistent in the two bands analyzed and specific from CDH1a expressing cells (two independent replicas).

It is also worth noting that, for the lower molecular weight E-cadherin band, presumably indicating the mature form (Fig. 4E), the combination of the first 3 N-terminal peptides (amino acids 55–74), observed after tryptic digestion, are decreased 8.8-fold in E-cadherin and the first four N-terminal peptides (amino acids 8–42) decrease 3.5-fold in CDH1a when compared with all the other peptides (Supplementary Material, Table S2). This indicates that CDH1a processing is 2.5 (8.8/3.5) times less efficient than E-cadherin processing. In the immature E-cadherin and CDH1a higher molecular weight bands, the intensity of the peptide peaks associated with the protein’s N-terminal does not significantly change, when compared with the peptide peaks associated with other regions of E-cadherin, as expected when protein processing is efficient.

The results obtained for both E-cadherin isoforms confirm their identity and reveal that CDH1a is less efficiently processed to a mature form than canonical E-cadherin, but are not sufficient to determine the potential sites of cleavage for each isoform. They show, nevertheless, that CDH1 and CDH1a encode two distinct proteins with different N-terminal regions that, independently of the processing, exist in the cell with potentially different functions.

**CDH1a isoform mimics canonical E-cadherin localization, adhesion complex interactions, cell-adhesion and invasion suppression properties when overexpressed in canonical E-cadherin-negative cells**

To assess CDH1a function, we conducted immunoprecipitation in protein extracts from cells expressing CDH1a (and the canonical form, for positive control) using the same anti-E-cadherin antibody, followed by western blot for proteins that classically interact with E-cadherin to form the adhesion complex (β-catenin and p120(ctn)). Our results demonstrate the effective binding of both proteins to these catenins (Fig. 4D) and are further supported by immunofluorescence results showing that CDH1a is able to induce β-catenin (Fig. 4E) and p120-catenin (Fig. 4F) recruitment to the cell membrane. This argues towards the putative functionality of the CDH1a protein in establishing an effective adhesion complex. Moreover, we observed that CDH1a, similarly to E-cadherin, was able to confer aggregation capacity to the otherwise non-aggregating CHO cells (Fig. 4G) and to reduce the intrinsic invasion levels of parental CHO cells (Fig. 4H). Overall, these results show that CDH1a isoform, when expressed alone, mimics the canonical E-cadherin function.

**CDH1a promotes cell invasion and angiogenesis when co-expressed with the canonical E-cadherin**

We next addressed the CDH1a-induced expression effect in cells endogenously producing functional E-cadherin. For that, we chose the gastric cancer-derived cell line MKN28. Due to suboptimal proportion of MKN28 CDH1a expressing cells, using the pLenti vector, we subcloned CDH1a and the canonical CDH1 into a pRES2-EGFP vector that expresses our protein of interest and GFP in the same cell. This approach allowed sorting GFP-positive cells and monitoring transfection efficiency which reached ~90% (Fig. 5A and B). Overexpression from either the canonical E-cadherin or CDH1a has not produced visible alterations in the overall E-cadherin expression levels or cellular localization, as ascertained by immunocytochemistry (Fig. 5C).

We confirmed mRNA overexpression of the canonical CDH1 or CDH1a using the previously described strategy, where exon 1–2 probe expression measures specifically canonical transcript expression and exon 6–7 expression measures overexpression of both transcripts. As expected, an increase in canonical CDH1 mRNA expression was detected by both probes, for cells transfected with the canonical CDH1 expression vector in comparison to Mock cells; whereas an increase in CDH1a mRNA overexpression was detected only by the exon 6–7 probe (Fig. 5D).

Surprisingly, the results obtained by western blot did not mimic the previous RNA expression data, since similar levels of E-cadherin protein were detected for the three cell lines (Fig. 5E). To unveil whether the excess of translated protein, due to the forced production of high levels of mRNA from both isoforms, was being degraded, we treated the three cell lines with MG132, a proteasome function inhibitor. We confirmed that this was the case, as significantly increased protein levels were detected both in E-cadherin and CDH1a overexpressing cells (P = 0.001 and P = 0.020, respectively) in contrast to Mock cells, after the treatment with the drug (Fig. 5F).

Assuming that the proteasome blindly degrades both isoforms in CDH1a overexpressing cells, a proportion of the detected protein by western blot and immunohistochemistry is expected to be the CDH1a isoform. Moreover, in CDH1a overexpressing MKN28 cells, high amounts of unprocessed CDH1a seem to occur, as shown by the abnormal superior thickening of the western blot signal (Fig. 5F). In contrast to canonical E-cadherin overexpressing MKN28 cells, separation between the upper and the main lower band (presumably the mature processed form) is not clear, due to the smaller difference between the unprocessed and processed proteins when compared with the canonical form.

To prove this, we have repeated mass spectrometry analysis and showed that the CDH1a-specific peaks (953.52 Da and
1722.85 Da) are found in MKN28 CDH1a overexpressing cells (Supplementary Material, Tables S3 and S4).

Next, we carried out aggregation, motility and invasion experiments to access the effect of concomitant expression of the two isoforms. We verified that CDH1a overexpression increased cell aggregation capacity in comparison with Mock cells; nevertheless, this augmented aggregation ability was lower than that induced by the canonic isoform (Fig. 6A). No differences were found in motility levels for the three cell lines (data not shown), which contrasted with results obtained for the invasion assay. CDH1a overexpression in MKN28 cells led to a significant increase in the number of invasive cells when compared with cells overexpressing E-cadherin or the empty vector (Fig. 6B). Taken together, these data indicate that CDH1a overexpression-induced less efficient cell aggregation did not affect cell motility but conferred an invasive phenotype to otherwise poorly invasive cells. No differences were observed in the three cell lines regarding cell proliferation or apoptosis (data not shown).

We further tested the effect of CDH1a overexpression in MKN28 cells regarding tumor-induced angiogenesis, using the classical chick embryo chorioallantoic membrane (CAM) in vivo model (29–31). The angiogenic response levels were quantified by counting the number of novel radial blood vessels formed in tumors induced by every cell line, and verified that tumors formed by CDH1a overexpressing cells were

Figure 5. CDH1a overexpression in E-cadherin expressing MKN28 cells does not interfere with canonic protein localization. (A) Flow cytometry was performed to monitor MKN28 transfection efficiencies of pIREs_E-cadherin, Mock and CDH1a constructs. (B) GFP fluorescence levels of MKN28 E-cadherin, Mock and CDH1a cells. Scale bar: 15 μm. (C) E-cadherin immunofluorescence of MKN28 E-cadherin, Mock and CDH1a cells. Overexpression of the transcripts has not influenced normal membrane protein localization. Scale bar: 15 μm. (D) QRT-PCR levels with two assays targeting different regions, 1–2 and 6–7 CDH1 exon boundaries. As expected, CDH1 overexpression results in higher levels of expression for both probes while for CDH1a only the second assay detects increased transcription levels. (E) E-cadherin western blot reveals an equivalent level of protein expression in each cell line. (F) E-cadherin western blot of MG132 treated cells shows that transfection results in an effective increase in protein translation that is masked by proteasome degradation.
the ones eliciting significantly more blood vessels when compared with cells transfected with the empty vector (Fig. 6C and D). By pan-cytokeratin immunoperoxidase, we confirmed that tumors growing in the CAM were exclusively formed by the human MKN28 cells (Fig. 6E) and demonstrated by PCR that these maintain CDH1 and CDH1a overexpression patterns (Supplementary Material, Fig. S3).

**CDH1a overexpression in MKN28 cells increases IFITM1 and IFI27 mRNA levels**

In order to determine the effect of CDH1a overexpression on the overall mRNA expression pattern of pLenti-transduced MKN28 cells, we performed a genome-wide mRNA expression array analysis. This experiment identified 50 genes (76 probes) whose expression was upregulated and 33 genes (42 probes) whose expression was downregulated (Supplementary Material, Fig. S4 and File S3). In Supplementary Material, Table S5, we have presented the 24 most altered (positively or negatively) genes, from which we selected preferentially upregulated genes for qRT–PCR validation. The eight genes with top fold-change and lower P-values were tested as well as two of the top downregulated genes. We were able to confirm IFN-induced transmembrane protein 1 (IFITM1) and IFI27 overexpression by qRT–PCR, specifically in cells overexpressing CDH1a (Fig. 7A and B) while for the other genes no consistent results were found (Supplementary Material, Fig. S5). IFITM1 levels were also upregulated upon CDH1a overexpression in HEK cells (data not shown).

We performed Ingenuity analysis on the list of significant differentially expressed genes to select the functional classes (defined as network) overrepresented. The highest Ingenuity scoring networks contain the most statistically robust candidates for hypothesis building. The Inflammatory response network was the most significant (Supplementary Material, Fig. S6, see also Supplementary Material, File S3). Interestingly, 4 out of 15 genes of the network known to be involved in gastric cancer are present in the list of differentially expressed. The other significant networks overrepresented in CDH1a responsive genes are cellular growth and proliferation, cellular movement and cancer and cell death (Supplementary Material, File S2). Through a Parametric Gene Set Enrichment Analysis (PGSEA), the transforming growth factor-β (TGF-β) pathway was predicted to be upregulated in MKN28 CDH1a
cells when compared with controls (Supplementary Material, Fig. S7).

**Invasion levels are restored and IFITM1 and IFI27 are lowered upon siRNA-mediated CDH1a downregulation**

In order to definitively demonstrate that the observed increases in invasion, and IFITM1 and IFI27 expression levels are driven by CDH1a overexpression, we designed a siRNA to target specifically this transcript. A 35% reduction in CDH1a levels (Fig. 8A) was sufficient to lower significantly invasion levels by 73% (Fig. 8B) and IFITM1 and IFI27 by 31 and 45%, respectively. Canonical CDH1a expression displayed no alteration (Fig. 8C).

**CDH1a transcript is overexpressed in gastric cancer cell lines**

Having observed the CDH1a potentially deleterious effects over canonical E-cadherin function, and given that CDH1a mRNA was not expressed in normal stomach, we tested its expression in gastric cancer-derived cell lines along with canonical E-cadherin expression. Interestingly, we observed an inverse correlation between CDH1a and canonical CDH1a mRNA expression in normal stomach and gastric cancer cell lines (Fig. 9A and B). While the canonical form was highly expressed in normal stomach and presented an overall down-regulation in gastric cancer cell lines, CDH1a mRNA was absent in normal stomach, as previously observed, and over-expressed in most gastric cancer cell lines (Fig. 9A and B). Moreover, most gastric cancer cell lines were found to co-express the canonical CDH1 and CDH1a mRNAs.

**DISCUSSION**

Loss of the epithelial adhesion molecule E-cadherin is thought to be the earliest and one of the most important steps in metastatic dissemination of epithelial cancer (7,32). Although impairment of E-cadherin gene and protein expression has been the subject of many studies, the causes for this disruption are, in many cases, unknown and the mechanisms so far described insufficient. It is possible that driver events for this impairment could be embedded in the genomic structure of the CDH1 gene itself. The pioneer studies by Stemmler et al. (16,17) showed that regulation at the CDH1 locus can be driven by powerful yet unidentified regulatory sequences with the 65 kb intron 2 of this gene. In the present study, we studied novel and conserved coding intron 2 sequences, and described the tissue specificity and potential biological function of one of them.

This is, to the best of our knowledge, the first report specifically addressing CDH1 alternative transcription, besides the alternative splicing variants, involving canonical exons, deposited in public databases. After proving that the mRNA sequences downstream of intron 2 were more represented in several normal tissues than those upstream of this intron, we tested whether annotated ESTs within CDH1 intron 2 could be transcribed into novel CDH1 exons. Two of these ESTs were shown to encompass novel exons able to splice with exon 3, sharing the remaining exons and termination with the canonical sequence. Both exons are targets of alternative splicing and each new exon generates at least one in-frame transcript and one long-noncoding RNA. This finding reveals the existence of novel E-cadherin transcripts in normal contexts. Somewhat surprisingly, spleen, which was shown not to express canonical transcript, expresses high levels of CDH1a, a novel protein encoding transcript that localizes at the cell membrane of structures from the splenic red medulla.

We were further able to map important genomic elements overlapping or near the novel transcription units herein described, such as a CpG island, Dnase I hypersensitive sites and a CTCF-binding site, that re-enforce their significance and potential function (Fig. 2). CpG islands are regions commonly found near TSSs and frequently associated with promoter regions. This fact constitutes evidence towards the possible regulation of novel transcripts through promoter methylation, as happens for the canonical CDH1 transcript. Dnase I hypersensitive sites, uncovered by ChIP-Seq experiments (UCSC Genome Browser tracks), tend to be near active genes, which are regularly transcribed (25). At the beginning of CDH1 exon 3, a predicted CCCTC-binding factor (CTCF) is found (UCSC Genome Browser tracks). Its characteristic insulator function (33,34) may justify the low percentage of RNAs starting after this region in stomach. Here, as an epithelial tissue, E-cadherin is known to exert well-established functions and must be, therefore, tightly regulated. This would allow preventing putative deleterious effects from the
concomitant expression of other transcripts from the CDH1 locus. In contrast, for spleen, as non-epithelial tissue, the CDH1 locus may somehow be more loosely controlled originating a broader variety of coding and non-coding transcripts of thus far uncertain functional relevance. Interestingly, one of the non-coding transcripts (CDH1j, depicted as Seq4 in Fig. 2B) was found to encompass an AluSc repeat. The Alu family is composed of over one million repetitive elements of about 300 bp long that are interspersed throughout the human genome and embrace more than 10% of it (9,35). Alu repeats have been implicated in the etiology of rearrangement-based deleterious mutations reported of CDH1 and other genes (9). In addition, such transposable elements have been shown to cause alternative splicing by providing the 5′ or 3′ splicing sites in so-called exonization events (36). Other regulatory mechanisms also involving Alu elements have been recently reviewed (37).

Overall evidence indicated that while CDH1a is a bona fide novel protein-coding transcript, while the other transcripts identified are more likely to play a role at the non-coding level. The conservation analysis underlined this difference, since CDH1a showed constant overall conservation, while CDH1j showed a conservation pattern which is more typical of enhancer like elements, with short blocks of high conservation. Transcribed enhancers have been shown to exist extensively, especially across highly conserved elements (38); thus, it is possible that CDH1j is a transcribed enhancer.

We further characterized one of the novel transcripts, CDH1a, selected because it possesses a long ORF, an adequate Kozak sequence, is polyadenylated and is therefore

---

**Figure 8.** Invasion, IFITM1 and IFI27 levels are reduced upon CDH1a-specific downregulation. (A) MKN28 CDH1a overexpressing cells were treated with a specific siRNA leading to 35% of knockdown when compared with the non-silencing siRNA. (B) Invasion levels were reduced in 73% after siRNA treatment. (C) IFITM1 and IFI27 levels were also reduced by 31 and 45%, respectively.

**Figure 9.** CDH1a transcript is overexpressed in gastric cancer cell lines. (A). Stomach qRT–PCR expression of CDH1 and CDH1a transcripts. Cell lines present high expression heterogeneity when compared with stomach values, although the majority express lower CDH1 levels than the normal tissue. (B). Q-SNapShot CDH1a expression levels show that this transcript is completely absent in normal stomach but de novo in most of the gastric cell lines tested.
predicted to encode a novel E-cadherin protein isoform. We then determined that CDH1a, an isoform that differs from the canonical one only at the most N-terminal residues, which were predicted to be cleaved from the mature form of the protein, could work as a negative modulator of the canonical E-cadherin function. This kind of effect has been previously described (23).

We were able to show that CDH1a isoform, driven by a pLenti vector, gave rise to a mature protein similar in size to the canonical one only at the most N-terminal residues, which were predicted to be cleaved from the mature form of the protein. We then determined that CDH1a, an isoform that differs from the canonical one only at the most N-terminal residues, was able to replace the canonical protein function, in that its protein location at the membrane, partner interactions co-localizing with β-catenin and p120-ctn, aggregation induction and invasion suppression properties, when expressed on its own, matching those of canonical E-cadherin. These data imply that, in the absence of the canonical E-cadherin, its functions may be accomplished by the new isoform herein described. Nevertheless, there are differences between E-cadherin and CDH1a, as two N-terminal-specific peptides from the latter (953.52 and 1722.85 Da) were detectable by mass spectrometry in CDH1a expressing cells. These peptides potentially constitute good targets for antibody design to better disclose CDH1a expression pattern.

Interestingly, the behavior induced by CDH1a expression alone was not mimicked when CDH1a was expressed in a context of endogenous canonical E-cadherin expression. In this setting, canonical protein localization was not impacted to a perceivable extent, but CDH1a forced expression led to moderate aggregation impairment and significantly increased angiogenic potential and invasion levels. Invasion increase was reverted upon CDH1a targeting through specific siRNA. A similar observation has been recently described in breast cancer-derived cell lines (39). The authors showed that co-expression of P-cadherin and canonical E-cadherin in MCF7 cells promoted invasion in vitro, while their counterparts expressing either molecule alone would remain non-invasive. The effect of CDH1a may be comparable in that CDH1a overexpression hampers E-cadherin normal function despite the cells’ epithelial phenotype is maintained. When we overexpressed CDH1a in E-cadherin expressing cells, no increase in the protein levels was observed unless we treated cells with a proteasome inhibitor. This led to a significant augment in total protein levels, implying that total protein levels are controlled by the ubiquitin-proteasome machinery (40,41), over a given expression threshold. We therefore hypothesize that the observed consequences at the functional level in cells co-expressing CDH1a and the canonical form can only occur if the cell blindly degrades both isoforms and not only CDH1a. The co-existence of both molecules, despite their similarities and expression at the right amounts, provides cancer cells with an advantageous invasive phenotype. The means by which this phenotype is achieved is not known, although our experiments indicate that CDH1a is less efficiently processed than E-cadherin. It is possible that potentiality extra peptides retained in the N-terminal of the mature CDH1a determine the effects observed either directly or indirectly through a number of genes with altered expression levels, upon CDH1a overexpression, as revealed by our expression microarray analysis. This experiment identified 50 upregulated and 33 downregulated genes. We performed Ingenuity analysis on the list of significant differentially expressed genes to select the functional classes (defined as network) overrepresented. The highest Ingenuity scoring networks containing the most statistically robust candidates for hypothesis building indicated the Antimicrobial Response, Inflammatory Response and Infectious Disease network as the most significant, despite others showed also putatively interesting results. Importantly, 4 out of 15 genes from the network known to be involved in gastric cancer are present in the list of differentially expressed. Since our previous results were indicative of gain of function effects, we gave greater importance to upregulated genes. We have therefore selected the eight genes with higher fold-change and lower P-values, when comparing MKN28 CDH1a overexpressing cells with controls, for qRT–PCR validation. We have selected also two downregulated genes. Possibly due to the different vectors used to establish cells for the array experiment and qRT–PCR validation we had a low level of concordance between both techniques. This, however, strengthens the results obtained for IFITM1 and IFI27 genes. These were the only genes with upregulation validated by qRT–PCR (besides CDH1) and interestingly they are part of the above-referred network. CDH1a siRNA treatment led to the downregulation of both IFITM1 and IFI27.

IFITM1 is a member of the IFN-inducible transmembrane protein family and its involvement in the migratory and invasive potential of gastric cancer cells is well established, supporting our data (42). It was demonstrated that IFITM1 induces tumor resistance to NK cells in gastric tumor cells, being hypothesized that it behaves like a surface molecule utilized by tumor cells for immune escape and migration, making of IFITM1 a possible therapeutic target for the treatment of gastric cancer. Recently, IFITM1 was implicated in the invasive front of early invasive and advanced HNSCC (43,44) and its knockdown has been shown to significantly inhibit migration and invasion of glioma cells (44).

IFI27 was identified in breast carcinoma cell lines (45) and belongs to a family of small, interferon-alpha (IFN-α)–inducible genes. Suomela et al. (46) proposed IFI27 as a novel marker of epithelial proliferation and cancer due to its upregulation in cutaneous squamous cell cancers as confirmed later (47).

A PGSEA, which allows the analysis of gene expression data to determine deregulation of gene signatures or ‘molecular concepts’, suggested that CDH1a overexpression in canonical E-cadherin expressing cells could impact the TGF-β pathway.

We observed that when compared with control cell lines, CDH1a overexpressing cells have upregulation of the TGF-β pathway, which has a well-recognized dual role both in tumor initiation and progression (48) and in tumor suppression (49,50). The mechanisms of tumor promotion by TGF-β include increased angiogenesis. The impact of CDH1a overexpression on the latter was seen through a marked increase in angiogenesis in the chick embryo CAM model. It is therefore possible that this induction of angiogenesis is triggered by TGF-β pathway upregulation, although the mechanism through which this pathway is activated remained elusive.
unique. A protein may carry specific domains that by themselves, or through their interactions, are toxic. Other proteins show high intrinsically disorder, which is likely to be associated with their deleterious effect. Yet others are tightly regulated, and are likely to perturb cellular homeostasis when the dynamics of their expression is disrupted. At the moment we cannot put forward which applies for CDH1a although our results seem to favor the last hypothesis.

In summary, our work describes, for the first time, tissue-specific coding and non-coding CDH1 transcripts arising from new exons encoded by intron 2 sequences, and highlights a possible novel mechanism underlying epithelial cancer cell invasion and angiogenesis. It is possible that the unique CDH1a peptides described and the potential effectors IFITM1 and IFI27 could become therapeutic targets and that isofrom-specific antibodies could be designed to target CDH1a-mediated cancer invasion and angiogenesis.

**MATERIALS AND METHODS**

**Biological samples and cell lines**

In order to perform transcript quantification, we have purchased commercial total RNA from several normal tissues: spleen (Ambion), stomach (Ambion), colon (Stratagene) and breast (Stratagene). Regarding thyroid and PBLs, we have pooled several samples from normal controls available at IPA-TIMUP. We have quantified CDH1a and the canonical transcript in RNA samples from MKN28, MKN45, GP202, SNU1, SNU638, KATOIII, AGS, IPA220 and NCI-N87 gastric cancer cell lines, available at the IPATIMUP’s repository. The cell lines used for transduction and transfection experiments were: (i) the CHO cell line, as a cadherin-free cell line, and (ii) the human gastric-cancer-derived MKN28, as a model with normal and functional E-cadherin. Normal stomach and spleen tissues used for immunohistochemistry were obtained from paraffin blocks from the Hospital S. João tissue bank (Porto, Portugal) and VU University Medical Center (Amsterdam, The Netherlands). Chicken fertilized eggs for the angiogenesis assay were acquired from commercial sources (Granja Santa Isabel, Spain).

**Bioinformatics**

Using the Ensembl database [http://www.ensembl.org, version 64, September 2011 (54)] and the UCSC Genome Browser [http://genome.ucsc.edu, NCBI36/hg18 (50)], we have collected data on several genomic elements predicted/annotated within CDH1’s intron 2 genomic locus: (i) CpG islands (49,54); (ii) EF (First Exon Finder) elements (54,55); (iii) Alu repeat elements (54,56); (iv) CTCF (CCCTC-binding factor) sites (34,54); (v) DNase I Hypersensitive Sites (49,56); (vi) EST (Expression Sequence Tags) from the database dbEST (54,57); and (vii) CAGE (Cap Analysis Gene Expression) tags from the ENCODE project (58,59). CAGE data collected refer to those obtained using the ‘normal’ lymphoblastoid cell line GM12878 and the leukaemia cell line K562. RNA populations analysed include total RNA and polyA-negative RNA. Subcellular compartments analysed...
include nucleus, cytosol, nucleoplasm, polysome and chromatin.

Using the annotated Expression Sequence Tags (EST) present in the UCSC Genome Browser (http://genome.ucsc.edu/cgi-bin/hgTracks), we have designed primers to amplify specifically new areas of transcription.

To assess CDH1a and CDH1j sequence conservation, the nucleotide level GERP score was downloaded from the UCSC genome browser, using the start and end coordinates of the novel portion of the transcript identified (i.e. excluding the known protein-coding portion of CDH1), as shown in Supplementary Material, File S1. The data were then assessed both at the single nucleotide level (first column in the table), as well as using a 10 bp sliding window (second column in the table and shown also as graph). Overall conservation was indicated by verifying the average, median, minimum and max GERP score observed across the entire region. Local conservation for CDH1j was investigated by searching for blocks which presented more than 10 nt with GERP score >1, highlighted in yellow in the table (Supplementary Material, File S1). The table also contains a UCSC screenshot showing GERP score, transcription score based on ENCODE RNA-Seq data and ENCODE DNAase I hypersensitivity regions.

RNA isolation and cDNA synthesis
RNA Isolation was performed using the Tripure Isolation Reagent (Roche). RNA quality was verified using the NanoDrop ND-1000 (Thermo Scientific) for confirmation of acceptable 260/280 nm absorbance ratio as well as determining sample concentrations. Subsequent first-strand cDNA synthesis was done using ~1 μg of total RNA with Superscript II Reverse Transcriptase and random hexamer primers (Invitrogen) following the company protocol.

Quantitative RT–PCR
Expression levels were assessed using TaqMan Gene Expression Assays. Further details are available at Supplementary Material and Methods.

Rapid Amplification of cDNA Ends (RACE)
We have acquired and used stomach and spleen FirstChoice RLM-RACE Kit (Ambion) according to the manufacturer’s instructions. Briefly, we have submitted RACE-ready cDNA to a first round of amplification using a 5' RACE outer primer and an outer exon 4 CDH1 primer, followed by a second round using another set of primers. Products were cloned and sequenced. More details are available at Supplementary Material and Methods. All sequences thus obtained were analysed and compiled: the spleen FirstChoice RLM-RACE Kit revealed six distinct types of sequences, two of which were very similar hence summed in one (seq4); the stomach FirstChoice RLM-RACE Kit revealed eight distinct types of sequences. Data on the coordinates of each sequence [after performing Blast analysis against the human genome (54)] are presented on Supplementary Material, Table S1.

Quantitative-SnapShot (Q-SnapShot)
CDH1a quantification was performed by Q-SnapShot (10). Further details are available at Supplementary Material and Methods.

Primer-walking PCR
This strategy was used to characterize the full sequence of transcripts studied. The primers used for the sequential amplification steps are available upon request.

Plasmids construction and virus production
The human E-cadherin nucleotide sequence as reported in the literature (2) was cut from the pcDNA3 plasmid described earlier (61) and cloned into a pLenti6/V5 Directional TOPO® (Invitrogen) giving rise to the CDH1 expression vector. The sequence of exons 1 and 2 was later replaced by exon 1a. These vectors, along with the empty vector, were used to transduce CHO cells. Viruses were produced following the company instructions. pIRES vectors were subsequently constructed using the above-mentioned inserts. Further details are available at Supplementary Material and Methods.

Cell culture, transduction, brefeldin-A and MG132 treatment
CHO and MKN28 cells were grown using standard conditions. These conditions as well as the brefeldin-A and MG132 treatments are described in detail at Supplementary Material and Methods.

Cell sorting and flow cytometry
Stable CHO and MKN28 cells were sorted in a Coulter Epics XL-MCL flow cytometer (Beckman Coulter) using the anti-E-cadherin HECD1 (Zymed Laboratories) and endogenous GFP expression, respectively, in order to obtain a homogeneous population of expressing cells. GFP flow cytometry was performed for MKN28 cells to monitor transfection levels across the course of experiments. Detailed protocol is available at Supplementary Material and Methods.

Antibodies, immunofluorescence, immunohistochemistry and microscopy
Antibodies and detailed conditions are available at Supplementary Material and Methods.

Slow aggregation assay and matrigel invasion assay
Cell aggregation assays were performed by coating the wells of a 96-well plate with 50 μl of an agar solution with subsequent cell seeding. Aggregation was evaluated at 24, 48 and 72 h under an inverted microscope.

For CHO and MKN28 cells’ invasion assays, Matrigel invasion chambers (BD Biosciences) were used according to the manufacturer conditions. Details are available at Supplementary Material and Methods.
**SDS–PAGE, western blotting and immunoprecipitation**

Total protein lysates have been used for all experiments. Standard conditions have been employed as described in detail in the Supplementary Material and Methods.

**Proteomic analysis**

After E-cadherin enrichment by immunoprecipitation, proteins were separated by SDS–PAGE using 3 μg of total protein. Following SDS–PAGE separation, proteins were stained, excised and MS and MS/MS peptide mass spectra were acquired with a MALDI-TOF/TOF 4700 Proteomics Analyzer. Detailed description is available at Supplementary Material and Methods.

**Chicken embryo in vivo angiogenesis assay**

The chicken embryo CAM model was used to evaluate angiogenic response of MKN28 parental and engineered cell lines. The number of new vessels growing radially towards the ring area was counted in a blind fashion. Further details are available at Supplementary Material and Methods.

**CDH1a siRNA treatment**

A custom CDH1a-specific siRNA was designed (IDT). Briefly, cells were seeded into six-well plates and grown in standard conditions until transfection with 100 nm of CDH1a siRNA with Lipofectamin 2000 following manufacturer’s instructions. After 60 h, we proceeded to invasion assays as described earlier. Remaining cells were used for RNA extraction, CDH1a, IFITM1 and IFI27 quantification by real-time PCR. A non-silencing siRNA (ThermoScientific) was used as negative control.

**Gene expression array**

Biotinylated cRNA targets were synthesized from each sample and hybridized to Affymetrix oligonucleotide chips according to the manufacturers’ instruction (Affymetrix Inc.). Gene-Chips® Human Genome U133 Plus 2.0 containing 54 000 probe sets (47 000 transcripts and variants including 38 500 well-characterized human genes) were used. The data discussed in this publication have been deposited in NCBI’s Gene Expression Omnibus (62) and are accessible through GEO Series accession number GSE32540 (http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE32540). Details concerning the analysis of data are available at Supplementary Material and Methods.

**Statistical analysis**

The R statistical program was used for the construction of standard boxplots, plotted with whiskers and outliers calculated with a maximum of 1.5 IQR. Due to the non-normal distribution of the data, a non-parametric test, in particular the Wilcoxon rank-sum test was used to calculate the significance. The P-values obtained were further corrected by using the Bonferroni Correction due to the multiple testing performed.

Corrected P-values <0.05 were considered statistically significant.

**SUPPLEMENTARY MATERIAL**

Supplementary Material is available at HMG online.

**ACKNOWLEDGMENTS**

The authors acknowledge Paula Silva and Adriana Rocha, Bárbara Gomes and Professor Fátima Carneiro for the support provided with tissue processing, immunohistochemistry and analysis immunohistochemistry experiments, respectively. The authors thank Beatriz Carvalho and Nicole van Grieken for the frozen normal spleen samples provided.

Conflict of Interest statement. None declared.

**FUNDING**

The work was supported by The Portuguese Foundation for Science and Technology (FCT) [Projects: PTDC/SAU-GMG/72168/2006; FCOMP-01-0124-FEDER-015779 (Ref. FCT PTDC/SAU-GMG/110785/2009); PhD grants: SFRH/BD/41223/2007 to H.P., SFRH/BD/46462/2008 to R.C., SFRH/BD/44074/2008 to J.C., SFRH/BD/32984/2006 to P.O.J.]. Salary support to CO from POPH—QREN/Type 4.2, European Social Fund and Portuguese Ministry of Science and Technology (MCTES); H.P. was a recipient of an EMBO short-term fellowship ASTF 239.00-2010.

**REFERENCES**


promotes squamous cell carcinomas in stratified epithelia. Cancer Cell, 12, 313–327.


