THE EXPRESSION OF LINE1-MET CHIMERIC TRANSCRIPT IDENTIFIES A SUBGROUP OF AGGRESSIVE BREAST CANCERS

THE EXPRESSION OF LINE1-*MET* CHIMERIC TRANSCRIPT IDENTIFIES A SUBGROUP OF AGGRESSIVE BREAST CANCERS

Running title: LINE1-*MET* in breast cancer

Authors: Umberto Miglio[1]*, Enrico Berrino[1]*, Mara Panero[1], Giulio Ferrero[2,3], Lucia Coscujuela[2], Valentina Miano[2], Carmine Dell'Aglio[1], Ivana Sarotto[1], Laura Annaratone[4], Caterina Marchiò[1,4], Paolo M. Comoglio[5], Michele De Bortoli[2], Barbara Pasini[4], Tiziana Venesio[1§] and Anna Sapino[1,4]

*U.M. and E.B. contributed equally to this work.

Affiliations: [1]Unit of Pathology and [5]Molecular Therapeutics and Exploratory, Candiolo Cancer Institute, FPO-IRCCS, Candiolo, Italy; [2]Department of Clinical and Biological Sciences; University of Turin, Orbassano, Italy; [3]Department of Computer Science, University of Turin, Turin, Italy; [4]Department of Medical Sciences, University of Turin, Turin, Italy.

§ ***Corresponding author***: Dr. Tiziana Venesio, Unit of Pathology, Candiolo Cancer Institute, FPO-IRCCS, Strada Provinciale 142, Km 3,95; 10060 Candiolo (Torino), Italy; e-mail: tiziana.venesio@ircc.it ; phone #: +39 011 9933547; fax #: +39 011 9933480

Keywords: LINE-1; L1-*MET*; chimeric transcript; breast cancer; triple negative breast cancer

Abbreviation: Long interspersed nuclear element (LINE-1 or L1); sense promoter (SP); anti-sense promoter (ASP); open reading frame (ORF); splicing donor (SD); splicing acceptor (SA); triple negative breast cancer (TNBC); immunohistochemistry (IHC); formalin-fixed paraffin-embedded (FFPE); disease free survival (DFS).

Category: Cancer Genetics and Epigenetics

Novelty and Impact: The identification of cancer specific L1-derived transcripts, such as L1-*MET*, have recently disclosed new regulating mechanisms of gene transcription and translational. Here we demonstrated that the high expression of L1-*MET* identifies a group of heterogeneous aggressive breast tumours, particularly enriched with triple negative breast cancers (TNBC). These novel findings significantly extend our understanding of L1-derived transcripts in breast cancer settings, suggesting a precise engagement of L1-*MET* in the diagnosis of early TNBCs.

ABSTRACT

Demethylation of the long interspersed nuclear element (LINE-1; L1) antisense promoter can result in transcription of neighbouring sequences as for the L1-*MET* transcript produced by the L1 placed in the second intron of *MET*. To define the role of L1-*MET*, we investigated the sequence and the transcription of L1-*MET* in *vitro* models and heterogeneous breast cancers, previously reported to show other L1-derived transcripts. L1-*MET* expressing cell lines were initially identified *in silico* and investigated for L1-*MET* promoter methylation, cDNA sequence and cell fraction mRNA. The transcriptional level of L1-*MET* and *MET* were then evaluated in breast specimens, including 9 cancer cell lines, 41 carcinomas of different subtypes, and 11 normal tissues. In addition to a L1-*MET* transcript ending at *MET* exon 21, six novel L1-*MET* splice variants were identified. Normal breast tissues were negative for the L1-*MET* expression, whereas the triple-negative breast cancer (TNBC) and the high-grade carcinomas were enriched with the L1-*MET* mRNA (P =0.005 and P=0.018, respectively). In cancer cells and tissues the L1–*MET* expression was associated with its promoter hypomethylation (ρ= -0.8 and -0.9, respectively). No correlation was found between L1-*MET* and *MET* mRNA although L1-*MET* expressing tumours with higher L1-*MET/MET* ratio were negative for the MET protein expression (P=0.006). Besides providing the first identification and detailed description of L1-*MET* in breast cancer, we clearly demonstrate that higher levels of this transcript specifically recognize a subset of more aggressive carcinomas, mainly TNBC. We suggest the possible evaluation of L1-*MET* in the challenging diagnosis of early TNBCs

***Keywords**:* LINE-1; L1-*MET*; chimeric transcript; breast cancer; triple negative breast cancer (TNBC)

INTRODUCTION

Long interspersed nuclear element (LINE-1 or L1) sequences are transposable elements with a "copy and paste" reverse transcriptase-dependent mechanism of spread, representing about 21% of the human genome.[1]Nevertheless, of over 500,000 L1 copies acquired through evolution, only very few (80-100) retain the ability to transpose.[2] A full length retrotranspositional-competent L1 is composed of a 5' untranslated region (5'UTR) encompassing both sense (SP) and anti-sense (ASP) promoter regions, two open reading frames (ORF-1 and ORF-2), and a 3' untranslated sequence (3'UTR) containing a weak polyadenylation signal.[3] ORF-1 encodes a protein with RNA binding and nucleic acid chaperone activity,[4] while the endonuclease and reverse transcriptase originating from ORF-2 are necessary for completion of the L1 cycle.[5,6] L1 transcripts are rarely expressed in differentiated somatic cells due to the silencing prompted by promoter CpG island methylation.[7] However, several studies have reported their reactivation in undifferentiated stem cells[8] and in pathological states such as autoimmune disease[9] and in epithelial cancers.[10,11] In this regard, the L1 expression was found associated with the genomic heterogeneity and instability of breast cancers [10] where it is considered a hallmark of carcinogenesis.[11]

Demethylation of the L1 promoter leads to the activation of both the SP and ASP. The SP controls canonical transcription of L1 ORFs, while the ASP drives illegitimate transcription of neighbouring sequence.[12] The role and features of these chimeric transcripts are not fully understood, but they represent an emerging mechanism for the regulation of gene expression. A new primate-specific L1 open reading frame (ORF0), whose translation is controlled by a promoter (ORF0 promoter) overlapping the ASP, was recently reported in the 5'UTR of the L1 sequence on the opposite strand of the other two L1 ORFs. ORF0, whose peptide activity was related to enhanced L1 mobility, has two splice donor sites (SD1 and SD2) leading to ORF0-proximal exon fusion transcripts.[13]

L1 elements in the human genome can give rise to illegitimate transcripts. *In silico* analyses have revealed up to 911 new putative antisense chimeric transcripts across the genome, but only a

few of these L1-containing genes are implicated in cancer.[14] A fusion transcript has been described between L1 and *MET*, a well-known oncogene, playing as driver in different cancers.[15,16] Interestingly, L1-*MET* was reported to be a negative regulator of canonical *MET* expression *in vitro*.[17] Conversely, some studies described that increased MET protein levels are associated with a higher transcription level of L1-*MET* in colon and liver carcinomas.[18,19] Although the recent identification of the ORF0-derived splice donor sites suggests that alternative forms of L1-*MET* can exist,[13] the complete sequence of the L1-*MET* mRNA has not been established. Moreover, there are no reports concerning the L1-*MET* expression in breast cancers, despite the detection of other L1-derived fusion transcripts.[20]

Here we aimed to comprehensively describe the L1-*MET* transcript and investigate its role in breast cancer. To this purpose, we preliminary characterized the L1-*MET* length, splicing-related mechanisms, and cellular localization *in-silico* using selected cancer cell lines as models. Then, we evaluated the L1-*MET* and *MET* transcripts in different types of breast cancer cells and in normal and malignant breast tissues with different clinic-pathological features.

MATERIALS AND METHODS

In silico characterization of the L1-MET transcript

To check the location of L1 in *MET* gene, the GenBank-derived L1 sequence (GenBank Accession #M80343.1) was aligned to the reference *MET* genomic sequence (GenBank Accession #NG_008996.1). Both *MET* genomic #NG_008996.1 and *MET* mRNA #NM_000245.3 sequences were used to establish the reference numbering of the L1-*MET* transcript.

The transcripts originating from the *MET* locus were investigated using publicly available RNA-seq data from the ENCODE Consortium[21] with the WashU Epigenome Browser.[22] Cell lines were selected for further analysis according to their L1-*MET* expression evaluating *MET* transcription by the CSHL and Caltech RNA-seq datasets. We focused on genomic coverage of the *MET* locus by long poly(A)- and poly(A)+ RNA-seq experiments, and analysed coverage computed on the forward and reverse strands, separately. Computational analyses of RNA-seq data were performed on whole cell, cytosolic, or nuclear contents.

The epigenetic status of the *MET* locus was analysed using publicly available ChIP-Seq data from the ENCODE consortium.[21] Specifically, coverage data for H3K4me3, H3K27ac, H3K36me3, H3K9me3, and H3K27me3 histone modifications were visualised using the WashU Epigenome Browser.[22] The highest coverage value was considered for each cell line when multiple replicates of the same epigenetic modification were present.

In healthy breast tissues *MET* mRNA analysis was performed using data from Genotype-Tissue Expression (GTEx) Project version 7.[23] The expression level of the different *MET* isoforms, normalized as Transcript Per Million (TPM), was downloaded from the project website (https://www.gtexportal.org/home/) containing data of 290 healthy breast RNA-Seq. *MET* expression data were pseudo-counted, log10 transformed, and computed with Ensembl annotations using the isoform ENST00000495962 as reference for L1-*MET*

In primary tumours L1-*MET* mRNA was assessed using processed RNA-Seq data from The Cancer Genome Atlas (TCGA) BRCA cohort which is composed of 1,103 primary tumours. RNA-

Seq data were downloaded from FireBrowse (http://firebrowser.org/) taking into account the track "illuminahiseq_rnaseqv2-RSEM_isoforms_normalized (MD5)". This track reports the RSEM-normalized expression level of each *MET* isoform as annotated in UCSC database. Isoform uc011knj was used as reference for L1-*MET* transcript.

The prediction of an open reading frame originating from the L1-*MET* chimeric transcripts was evaluated using the Translate tool from Expasy – Bioinformatics Resouce Portal (https://web.expasy.org/translate).

Cancer cell lines

A549, HCT116, MDA-MB468, T-47D, ZR-75-1, and HS-578T cells were grown in RPMI supplemented with 10% FBS; MCF-7, MDA-MB231, SKBR3 in high-glucose DMEM with 10% FBS. For BT474 we used RPMI with 20% FBS and 10 ng/mL insulin, for MDA-MB453 Leibovitz's L-15 Medium with 10% FBS in a free gas exchange atmospheric air, whereas HMEC were grown in DMEM/F15 supplemented with 10% FCS, 10 ng/mL insulin, 0.5 µg/mL hydrocortisone and 5 ng/mL EGF. BT474 were obtained from the "Deutsche Sammlung von Mikroorganismen und Zellkulturen" (DMSZ), whereas MDA-MB468, T-47D, HS-578T, and MDA-MB231 were obtained from NCI-60 panel. All the other cell lines were obtained from the American Type Culture Collection. Their genetic identity was confirmed by short tandem repeat profiling (PowerPlex® 16 HS System, Promega, Madison, WI), last repeated in June2018. Cells were periodically tested for mycoplasma contamination using Venor GM kit (Minerva Biolabs, Berlin, Germany).

L1-MET and MET gene analysis in cancer cell lines

DNA was extracted from cells using the Blood & Cell Cultured Mini Kit (Qiagen, Hilden, Germany) whereas RNA was obtained with the RNeasy Mini Kit (Qiagen). Nucleic acids were evaluated using Qubit 2.0 fluorometer (ThermoFisher Scientific, Waltham, MA) and NanoDrop spectrophotometer (ThermoFisher Scientific).

L1-*MET* ASP methylation was achieved by performing quantitative bisulfite pyrosequencing in triplicate using the PSQ 96 pyrosequencing (Qiagen). For each sample, 300 ng of DNA were converted by MethylEdge Bisulfite Conversion System (Promega) and the methylation status assessed as previously reported (Supplementary Table 1).[18] Five µl of converted DNA was amplified in a reaction mix consisting of 1X buffer, 2.5 mM $MgCl_2$, 0.3 mM dNTPs, 0.2 µM of each primer, 1X EvaGreen dye, 0.025U/µL Taq Polymerase (TaKaRa, Clontech, Mountain View, CA), and H_2O to a final volume of 50 µL. End-point PCR amplifications were performed on RotorGene 6000 thermal cycler (Qiagen). Fully methylated and unmethylated DNAs (Promega) were used as positive and negative controls.

One µg RNA extracted from cells was reverse transcribed in a mix containing 120 ng of oligo(dT) and *MET* exon 21 (5'- GGGTGCCAGCATTTTAGCATT -3') primers, using the Reverse Transcription System (Promega). cDNA was quantified with the Qubit ssDNA HS assay kit on the Qubit 2.0 fluorometer (ThermoFisher Scientific).

To investigate the length of L1-*MET* mRNA, a common forward primer (F1) and four different reverse primers on *MET* exons 5, 8, 14, and 21 were used. *MET* transcript was amplified with a forward primer located in exon 2 and a reverse primer in exon 5. Primers and different PCR conditions are reported in Supplementary Table 1. All PCR products were purified with Illustra ExoProStar (GE Heathcare, Buckinghamshire, UK) for subsequent sequencing analyses.

To determine alternative L1-*MET* transcripts, a specific pair of primers (F2 forward and R3 reverse) amplifying the ORF0-related SD1 or SD2 sites was adopted. These variants were characterized by six forward primers mapping to each splice junction and a common R3 reverse primer (Supplementary Table 1). PCR templates were separated by electrophoresis on 3% agarose gel, and the corresponding bands were extracted using PCR clean-up gel extraction (Nucleo Spin Macherey-Nagel; Düren, Germany) and PCR templates were purified by Illustra ExoProStar (GE Heathcare).

L1-*MET* transcript subcellular localisation was evaluated by specific nucleus-cytoplasm separations. After cell centrifugation, nucleus was split from cytosolic fraction by adding 200 µL of Dautry Buffer (10 mM Tris HCl pH 7.8, 140 mM NaCl 1.5 mM MgCl2, 10 mM EDTA 0.5% NP40) and 100 U/ml recombinant RNasin ribonuclease inhibitor (Promega) to the pellet.[24] RNA was extracted fromthe collected cytoplasmic fraction, the nuclear and the total cell pellets using 1mL Trizol Reagent (Invitrogen). Amplifications were carried out using F1 forward and R3 reverse primers (Supplementary Table 1). 14S ribosomal RNA and U2 small nuclear RNA were used as controls.[25,26]

Cycle-sequencing PCR reactions were set up using the Big Dye Version 3.1 Terminator cycle-sequencing kit (ThermoFisher Scientific) with 15 ng of amplified cDNA, 5 pmol of specific primer (Supplementary Table 1) in a final volume of 20 µL as follows: 25 cycles at 96°C for 10 s, 50°C for 5 s, and 60°C for 4 min. The products were purified by Agencourt CleanSEQ (Beckman Coulter, Brea, CA), and sequenced on an automated 16 capillary sequencer (3730 DNA Analyzer, Applied Biosystems, Foster City, CA). The cDNA electropherograms were aligned with the genomic *MET* sequence (GenBank Accession #NG_008996.1) using the Basic Local Alignment Search Tool (BLAST, http://blast.ncbi.nlm.nih.gov/Blast.cgi).

The expression profiles of L1-*MET* transcripts and *MET* gene were evaluated in triplicate by qRT-PCR using 25 ng of cDNA in a reaction mix consisting of 1X buffer, 2.5 mM $MgCl_2$, 0.2 mM dNTPs, 0.2 µM of each primer, 2X EvaGreen dye, 0.04U/µL Taq Polymerase (Promega), and H_2O to a final volume of 25 µL. The L1-*MET* amplification was carried out using the primers F2 and R3, whereas for the alternative transcripts splice-specific primers were used in combination with R3; for *MET* gene *MET* F2 and R3 primers were used (Supplementary Table 1). Relative expression quantification (RQ) was calculated according to the following formula, using *GAPDH* as endogenous control: $RQ = 2^{-[\Delta Ct]}$, where ΔCt = [Ct target gene (tumour sample) − Ct *GAPDH* (tumour sample)].

In order to obtain whole cell lysates for western blot analyses, cells were solubilized in boiling EB buffer and protein concentration was determined using a BCA Protein Assay Reagent kit (Pierce Biotechnology, Rockford, IL, USA). After separation (50 μg for each sample)by SDS-PAGE (4–12%, Life Technologies Foster City, CA, USA), proteins were transferred onto a nitrocellulose membrane (TransBlot Turbo, BIO-RAD, Hercules, CA, USA), immunoblotted and ECL Chemiluminescence (BIO-RAD) detected by ChemiDoc (BIO-RAD). For immunoblot we used MET antibody (D-4 sc514148, Santa Cruz, Dallas, TX, USA) and, as endogenous control, HSP90 antibody (sc13119 – Santa Cruz).

L1-MET and MET analysis in breast cancer tissues

Fresh frozen tissues derived from 41 breast cancers were collected, stored at −80°C, and annotated for their clinico-pathological data (Supplementary Table 2**)** from 2010 to 2014 at the Pathology Division of the Department of Medical Sciences, University of Turin. Samples included tissues from different molecular subtypes of breast cancers: 6 luminal A, 11 luminal B, 10 luminal B/HER2, 2 HER2-positive, and 12 triple-negative breast cancers (TNBCs) as defined by the St. Gallen expert consensus conference.[27, 28] Histologically, 31 were ductal carcinomas, 3 lobular and 7 represented other types of breast cancer. Among these tumours, 23 were G3 grading, 16 G2 and only one G1. In our cohort, 16 patients did not show lymph node metastases, whereas 18 were pN1, 3 pN2 and 3 pN3. As for as the staging, 14 were pT1, 22 pT2, 3 pT3 and only one pT4. Moreover, 31 patients (77.5%) showed vascular invasion and 32% developed recurrence (Supplementary Table 2). Eleven fresh normal breast tissue samples were also collected in 2016 at the Unit of Pathology, Candiolo Cancer Institute, and stored in RNA later (ThermoFisher Scientific). Written informed consent was obtained from all patients. For all tissues DNA and RNA were extracted using Maxwell RSC Cell DNA Purification (Promega) and Maxwell RSC miRNA Tissue Kit (Promega); quantifications carried out using Qubit 2.0 fluorometer (ThermoFisher Scientific). All samples were investigated for the methylation status as reported for the cells. One μg RNA template was reverse transcribed using the SuperScript™ IV VILO™ reverse transcriptase (Invitrogen). L1-

MET and *MET* transcripts expression profiles were quantified by qRT-PCR as reported for the cell lines. Relative expression quantification (RQ) was performed according to the following formula, using GAPDH as endogenous control : $RQ = 2^{-[\Delta\Delta Ct]}$, where $\Delta\Delta Ct = [Ct \text{ target gene (tumor sample)} - Ct \text{ GAPDH (tumor sample)}] - [Ct \text{ target gene (calibrator samples)} - Ct \text{ GAPDH (calibrator samples)}]$ As for *MET* transcription analysis, normal breast tissues were used as calibrator, whereas low expressing breast cancer specimens were used as calibrators for L1-*MET* transcript.

Immunohistochemistry (IHC) was performed on 3-µm thick sections, collected from formalin-fixed paraffin-embedded (FFPE) tissues, using the CONFIRM anti-MET primary antibody (clone SP44, dilution 1:100, Ventana Medical Systems, Tucson, USA) on Ventana BENCHMARK® XT instrument using UltraView DAB kit (Ventana Medical Systems). Normal colon tissue was used as positive control for each slide. The membranous and/or cytoplasmic positivity was semi quantitatively evaluated by two independent pathologists (AS and IS), using a score ranging from 0 to 3+.

Statistical analyses

All statistical analyses were carried out using MedCalc statistical software 13.0.6 (MedCalc Software bvba, Ostend, Belgium) whereas the heatmap was obtained using the R software (R Foundation for Statistical Computing, Vienna, Austria). To reduce outlier effects, gene expression data were Log transformed. The Spearman's rank correlation test (ρ) was used to determine the correlation between L1-*MET* methylation and expression levels. In order to evaluate significant differences between groups, the Mann-Whitney test was used whereas the receiver operative characteristics (ROC) curves and Youden's index were performed to identify a cut-off value to predict the probability of recurrence. Disease free survival (DFS) was defined as the time from diagnosis to the recurrence of disease or to date of last follow-up. Survival analysis was evaluated with the Kaplan–Meier method, and groups were compared with the log-rank test. P values of <0.05, with 95% confidence intervals (CI), were considered statistically significant.

RESULTS

In silico characterisation of the L1-MET transcript

In-silico analyses confirmed the presence of the ORF0-containing L1 element, as reported by Denli *et al.*,[13] within the second intron of *MET,* between nucleotides g.51092C (c.1201-13172C) and g.57084C (c.1201-7191C) (Figure 1A). The first methionine of ORF0 mapped to nucleotide g.56654A (c.1201-7610A) of *MET* and its first nucleotide was taken as +1 of the L1-*MET* reference numbering (Supplementary Figure 1A).

In silico RNA-seq analysis, performed to select cell lines expressing different L1-*MET* levels, identified high L1-*MET* signals in breast cancer-derived MCF-7 and lung cancer-derived A549 cells; on the contrary, no L1-*MET* signals were reported in colon cancer-derived HCT116 and non-transformed breast HMEC cells. Accordingly, L1-*MET* RNA was detected in both the nuclear and cytosolic fractions of MCF-7 and A549 cells (Figure 1B).

To verify whether the L1-*MET* transcription was related to epigenetic modification of *MET*, we explored the epigenetic status of the *MET* locus in publicly available ChIP-seq datasets. Consistent with the observed level of the L1-*MET* transcription, there was an enrichment of active promoter histone marks (H3K27ac and H3K4me3) at the L1-*MET* transcription start sites (TSS) in both MCF-7 and A549 cells, but not in HCT116 and HMEC cells (Supplementary Figure 1B).

In vitro characterization of the L1-MET transcript

As expected from *in silico* results, the presence of the L1-*MET* transcript was confirmed by RT-PCR in MCF-7 and A549 cells but not in HCT116 and HMEC (Figure 2A). Pyrosequencing analysis of the four CpG sites in the L1 ASP, confirmed stronger hypomethylation in MCF-7 and A549 cells and hypermethylation in HCT116 and HMEC cells (Supplementary Figure 2A). Further cDNA sequencing analyses confirmed that the identified L1 sequence was not included in the canonical *MET* transcript (primers *MET*- F2 and R5 mapping in exon 2 and 5 of *MET* gene; Supplementary Table 1).

The length of the L1-*MET* transcript was assessed using different reverse primers located on *MET* exon 5, 8, 14, and 21.A novel 2894 bp transcript, encompassing the region between the L1 5'UTR and *MET* exon 21 was identified in both MCF-7 and A549 cells (Figure 2 B and C).

Amplification of the L1-*MET* cDNA showed multiple bands suggesting the occurrence of different L1-*MET* variants (Figure 2A and B), possibly arising from ORF0 splice donor sites SD1 or SD2. SD1 and SD2 are common SD sites with adjacent GT dinucleotides, located at +107 and +192 in respect to L1-*MET* numbering (Supplementary Figure 1A). Sequencing analysis supported the presence of six splice variants (SVs) formed by combinations of SD1, SD2, and SD3 located at +790, and three different splice acceptor (SA) sites: SA1 at nucleotide +633, SA2 at +712, and SA3 corresponding to the canonical *MET* exon 3 acceptor site (data not shown). The SA sites and SD3 were specific for the L1-*MET* transcript located in the intronic *MET* sequence and did not belong to L1. Each splice variant (SV1-6) was sequenced using primers overlapping the splice sites (Figure 3A). All the splice combinations are reported in detail in Figure 3B. The quantitative expression of each transcript measured by qRT-PCR revealed that MCF-7 cells preferentially retained SV2, whereas A549 cells predominantly displayed SV1 and SV2 (Figure 3C). Purified nuclear and cytoplasmic RNAs showed that L1-*MET* was present in both nuclear and cytoplasmic compartments in both MCF-7 and A549 cells (Supplementary Figure 2B and 2C).

Potential L1-*MET* open reading frames were assessed by Expasy tool evaluating all the discovered splicing variants, but no ORFs were identified.

The L1-MET and MET transcripts in breast cancer

We first compared by qRT-PCR the relative expression of the fragment encompassing all the splice variants (primers F2 and R3 on *MET* exon 3, Supplementary Table 1) in a set of 9 breast cancer cell lines which were representative of the different molecular subtypes: MCF-7, ZR-75-I, and T47D cells for Luminal A, BT474 for Luminal B-HER2, SKBR3 and MDA-MB453 for HER2-enriched, and MDA-MB468, MDA-MB231, and HS578T for triple negative. Cells variably

expressed L1-*MET* (with BT474 and MCF-7 cells showing a higher level compared to the others), whose level was directly correlated with the hypomethylation status of the L1 ASP (ρ=-0.8) (Figure 4A). SD1 derived transcripts were the prevalent isoforms (Supplementary Figure 3A). The canonical *MET* transcript, evaluated using *MET*-F2 and -*R3* primers, was also variably expressed, but it did not correlate with the L1-*MET* mRNA level (ρ= 0.6) (Figure 4B).

Breast cancer cells were also checked by western-blot for their MET protein expression which resulted to be high in MDAMB-231, MDAMB-468, and HS578T cells, weak in BT474, and negative in the other analysed cells (Figure 4C).

In breast cancer tissues and normal mammary gland L1-*MET* expression was preliminary assessed by *in silico* analyses using TCGA RNA-seq (breast cancer tissues) and GTEx (normal mammary gland) datasets. Both L1-*MET* and *MET* were heterogeneously expressed in breast cancers (Supplementary Figure 4A). However, *MET* mRNA was also variably detected in normalmammary gland, where no L1-*MET* signal was found (Supplementary Figure 4B).

In order to confirm the *in silico* results, L1-*MET* and *MET* transcripts were evaluated in a cohort of 41 breast carcinomas belonging to different molecular subtypes and 11 normal mammary gland derived samples. L1-*MET* and *MET* transcripts were variably detected in the majority of breast cancers (32/41: 78% and 31/41: 76%, respectively) with enrichment of both in the TNBCs compared to the Luminal subtypes (P=0.005 and P=0.03, respectively) (Figure 5). However, no correlation was found between L1-*MET* and *MET* mRNA (ρ=0.45) and higher levels of L1-*MET*, but not of *MET*, were significantly associated with poorly differentiated (G3) carcinomas (P=0.018) (Figure 5). In line with the *in silico* results, no L1-*MET* expression was found in normal mammary gland samples, while *MET* transcript was detected at low levels. Moreover, the L1-*MET* expression correlated with the promoter hypomethylation status (ρ=-0.9) in all samples. Finally, as in the breast cancer cells, the SD1-dependent isoforms were prevalent in the eight samples with the higher L1-*MET* expression (LogRQ >2) (Supplementary Figure 3B and 3C).

MET protein was then evaluated by IHC in breast cancer tissue confirming that overall the positive samples (2+ and 3+) had higher *MET* transcription than the MET negative ones (0 and 1+) (P=0.04). Moreover, the TNBC was the subtype enriched with MET expressing cases (Figure 5). The ratio between L1-*MET* and *MET* transcripts was used to evaluate the potential effect of L1-*MET* on MET protein. A high ratio was associated with the absence of MET expression (P=0.006) (Figure 5). Moreover, MET protein was negative in 5 samples (7, 12, 24, 26, and 27) showing L1-*MET* but not *MET* mRNA (Figure 5).

We next investigated the relationship between L1-*MET* expression and prognosis. Due to the short time of follow up of the patients, we focused on disease-free survival (DFS). Twenty-six patients did not develop recurrences and 13 patients relapsed, with a mean DFS of 42 months (range 1-71) and 20 months (range 2-52), respectively. Based on the onset of recurrence, we determined a L1-*MET* cut-off value of 1.21 using a receiver operating characteristics (ROC) curve with an area under the curve (AUC) of 0.68. According to this cut-off, two groups with low and high L1-*MET* gene expression were defined. The patients with low levels of L1-*MET* had better disease free survival, with a trend to significance (P=0.066) (Figure 6).

DISCUSSION

We here provide a comprehensive characterisation of L1-*MET* transcripts, including their epigenetic modification, length, splice-related mechanisms, and subcellular localisation in breast cancer cells. We also report, for the first time, that L1-*MET* transcripts are present in breast cancers but not in the normal mammary gland, with higher expression in triple-negative and more aggressive carcinomas.

In silico investigation of RNA-seq data confirmed the reported L1-*MET* and *MET* expression profiles of MCF-7 (L1-*MET* positive/*MET* negative) and HCT116 (L1-*MET* negative/*MET* positive) cell lines,[17] and we found that *MET* gene-expressing A549 cells[29] are L1-*MET* positive as well, whereas HMECs are L1-*MET* negative/*MET* negative. Therefore, these cancer cells represent a heterogeneous set suitable for the L1-*MET* transcript characterisation. *In silico* epigenetic analyses proved the presence of positive histone marks (H3K27 acetylation and H3K4 tri-methylation) strongly associated with euchromatin and active promoters (and therefore activated gene expression) at the L1-*MET* locus in L1-*MET*-expressing cell lines only.[30] Our *in vitro* experiments demonstrated hypomethylation of CpG islands at the L1-ASP of cell lines expressing the alternative transcript but hypermethylation in those not transcribing L1-*MET*. Our data further corroborate the role of the L1-ASP demethylation in controlling L1-*MET* transcription.[17-19,31] The lack of L1-*MET* in HCT116 cells can be explained by the microsatellite instability-related hypermethylation phenotype in this colon cancer cell line, whereas in HMEC breast cells L1-*MET* absence could possibly be attributed to transposable element silencing.

To date, the detailed sequence of the L1-*MET* transcript has been poorly defined. One study described a transcript ending at *MET* exon 5,[32] whereas others have reported the presence of an amplified fragment ending at exon 3.[17-19,31] Here we provided the first evidence of a L1-*MET* transcript ending at exon 21 of the host *MET* gene.

The recent discovery of the ORF0 sequence in the antisense orientation of L1 revealed the presence of two splice donor sites,[13] suggesting the existence of different variants. The same donor

but different acceptors were reported in an older study.[33] We demonstrated that L1-*MET* acceptor sites specifically belonged to the *MET* sequence and that the internal donors of ORF0 generated six L1-*MET* variants involving three acceptors, two located in *MET* intron 2 (SA1 and the downstream SA2) and one belonging to *MET* exon 3 (SA3). L1-*MET* splicing variants were previously reported, but not characterised, in samples of oesophageal carcinoma,[34] and in NTera2D1 teratocarcinoma cells.[32] To the best of our knowledge, this is the first report concerning breast cancer cells and tissues where the predominant variants arise from the SD1.

The L1-*MET* transcript was previously detected in colorectal and hepatocellular carcinomas.[18,19] In the present study, we focused on breast cancer in which the expression of *MET* host gene has been variably detected,[35] and L1-chimeric transcripts other than L1-*MET* have been reported.[20] Our results show that L1-*MET* chimeric transcripts can be heterogeneously detected in breast cancers but not in normal mammary gland, thus supporting a cancer-specific involvement of L1-*MET* in this type of tumour.

In our study L1-*MET* was not associated with *MET* transcript expression. This could be partially explained by the different promoter regulating mechanisms. Indeed, we found a correlation between the L1-*MET* level and its promoter hypomethylation in both cancer cells and tissues, while *MET* transcription has been reported not to be directly regulated by methylation.[17,36]

At present the function of L1-*MET* is unknown, but poly(A) (+) RNA-seq data and our *in-vitro* analyses revealed the presence of L1-*MET* RNA in both nucleus and cytoplasm, implying that the L1-*MET* transcript can be processed as a canonical RNA. Moreover, L1-*MET* was shown to originate a protein when transfected in cancer cell lines[31] and a positive correlation between L1-*MET* expression and MET protein was also reported *in vitro*.[18,19] As far as breast cancer cells, our *in vitro* analysis showed that the L1-*MET* expression can be independent of the MET protein presence.

Since evidences concerning the potential interaction between L1-*MET* transcription and MET translation are lacking for cancer tissues, we evaluated MET immunostaining in our cohort of

breast cancer specimens. In our hands, samples positive for L1-*MET*, but negative for *MET* transcription, did not express MET protein, supporting furtherly that L1-*MET* mRNA does not encode for a protein. Notably, a decreased MET immunostaining was observed in those samples where the transcription level of L1-*MET* was higher than *MET*. According to these findings, we suggest for L1-*MET* a role as long non-coding RNA (lncRNA). *MET* translation could be regulated by L1-*MET* which, retaining the same *MET* 3'UTR and containing *MET*-specific miRNA response elements (MREs),[37-39] might act as a competing endogenous RNA (ceRNA).[40]

Although the number of cases was too low to reach any definitive conclusion, in our cohort the breast cancer patients with shorter DFS had higher L1-*MET* transcription, which, in turn, was significantly associated with tumours with an aggressive biology, such as the triple-negative phenotype and the higher tumour grade.

In conclusion, our findings provide the first identification and detailed description of L1-*MET* transcripts in breast cancer, particularly in TNBC. These data, together with the complete absence of L1-*MET* expression in normal tissues, could suggest a possible and useful engagement of this transcript in the challenging morphological diagnosis of early triple negative lesions.

Financial disclosure and conflict of interest

This work was partly supported by funds from FPRC 5xMille Ministero Salute 2011, FPRC 5xMille Ministero Salute 2014, AIRC 5xMille MCO Extension Program ref. 9970, Italian Ministry of Health, Ricerca Corrente 2017 to A.S and by AIRC IG ref.15600 to M.D.B; L.A. is recipient of Fondazione Umberto Veronesi Post-Doctoral Fellowship.

The authors declare non conflict of interest.

REFERENCES

1. International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. Nature 2001; 409:860–921

2. Brouha B, Schustak J, Badge RM, Lutz-Prigge S, Farley AH, Moran JV, Kazazian HH Jr. Hot L1s account for the bulk of retrotransposition in the human population. Proc Natl Acad Sci USA 2003;100:5280–5285

3. Swergold GD. Identification, characterization, and cell specificity of a human LINE- 1 promoter. Mol Cell Biol 1990;10:6718–6729

4. Kolosha VO, Martin SL. High-affinity, Non-sequence-specific RNA Binding by the Open Reading Frame 1 (ORF1) Protein from Long Interspersed Nuclear Element 1 (LINE-1). J Biol Chem 2003;278:8112–8117

5. Feng Q, Moran JV, Kazazian HH Jr, Boeke JD. Human L1 retrotransposon encodes a conserved endonuclease required for retro-transposition. Cell 1996;87:905–916

6. Luan DD, Korman MH, Jakubczak JL, Eickbush TH. Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition. Cell 1993;72:595–605

7. Yoder JA, Walsh CP, Bestor TH. Cytosine methylation and the ecology of intragenomic parasites. Trends Genet 1997;13:335-40

8. Robbez ML, Rowe HM. Retrotransposons shape species specific embryonic stem cell gene expression. Retrovirology 2015;12:45

9. Mavragani CP, Sagalovskiy I, Guo Q, Nezos A, Kapsogeorgou EK, Lu P, Liang Zhou J, Kirou KA, Seshan SV, Moutsopoulos HM, Crow MK. Expression of Long Interspersed

Nuclear Element 1 Retroelements and Induction of Type I Interferon in Patients with Systemic Autoimmune Disease. Arthritis Rheumatol 2016;68:2686-2696

10. Chen L, Dahlstrom JE, Chandra A, Board P, Rangasamy D. Prognostic value of LINE-1 retrotransposon expression and its subcellular localization in breast cancer. Breast Cancer Res Treat 2012;136:129-42

11. Rodić N, Sharma R, Sharma R, Zampella J, Dai L, Taylor MS, Hruban RH, Iacobuzio-Donahue CA, Maitra A, Torbenson MS, Goggins M, Shih IeM, Duffield AS, Montgomery EA, Gabrielson E, Netto GJ, Lotan TL, De Marzo AM, Westra W, Binder ZA, Orr BA, Gallia GL, Eberhart CG, Boeke JD, Harris CR, Burns KH. Long interspersed element-1 protein expression is a hallmark of many human cancers. Am J Pathol 2014;184:1280-6

12. Speek M. Promoter of human L1 retrotransposon drives transcription of adjacent cellular genes. Mol Cell Biol 2001;21:1973-85

13. Denli AM, Narvaiza I, Kerman BE, Pena M, Benner C, Marchetto MC, Diedrich JK, Aslanian A, Ma J, Moresco JJ, Moore L, Hunter T, Saghatelian A, Gage FH. Primate-specific ORF0 contributes to retrotransposon-mediated diversity. Cell 2015;163:583-93

14. Criscione SW, Theodosakis N, Micevic G, Cornish TC, Burns KH, Neretti N, Rodić N. Genome-wide characterization of human L1 antisense promoter-driven transcripts. BMC Genomics 2016;17:463-78

15. Prat M, Narsimhan RP, Crepaldi T, Nicotra MR, Natali PG, Comoglio PM. The receptor encoder by the human c-MET oncogene is espressed in hepatocytes, epithelial cells and solid tumours. Int J Cancer 1991;49:323-8

16. Boccaccio C, Comoglio PM. MET, a driver of invasive growth and cancer clonal evolution under therapeutic pressure. Curr Opin Cell Biol 2014;31:98-105

17. Weber B, Kimhi S, Howard G, Eden A, Lyko F. Demethylation of a LINE-1 antisense promoter in the cMet locus impairs Met signalling through induction of illegitimate transcription. Oncogene 2010;29:5775-84

18. Hur K, Cejas P, Feliu J, Eden A, Lyko F. Hypomethylation of long interspersed nuclear element-1 (LINE-1) leads to activation of proto-oncogenes in human colorectal cancer metastasis. Gut 2014;63:635-46

19. Zhu C, Utsunomiya T, Ikemoto T, Yamada S, Morine Y, Imura S, Arakawa Y, Takasu C, Ishikawa D, Imoto I, Shimada M. Hypomethylation of long interspersed nuclear element-1 (LINE-1) is associated with poor prognosis via activation of c-MET in hepatocellular carcinoma. Ann Surg Oncol 2014;21:S729-35

20. Cruickshanks HA and Tufarelli C. Isolation of cancer-specific chimeric transcripts induced by hypomethylation of the LINE-1 antisense promoter. Genomics 2009;96:397-406

21. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. Nature 2012;489:57-74

22. Zhou X, Maricque B, Xie M, Li D, Sundaram V, Martin EA, Koebbe BC, Nielsen C, Hirst M, Farnham P, Kuhn RM, Zhu J, Smirnov I, Kent WJ, Haussler D, Madden PA, Costello JF, Wang T. The Human Epigenome Browser at Washington University. Nat Methods 2011;8:989-90

23. Carithers LJ, Ardlie K, Barcus M, Branton PA, Britton A, Buia SA, Compton CC, DeLuca DS, Peter-Demchok J, Gelfand ET, Guan P, Korzeniewski GE, Lockhart NC, Rabiner CA, Rao AK, Robinson KL, Roche NV, Sawyer SJ, Segrè AV, Shive CE, Smith AM, Sobin LH, Undale AH, Valentino KM, Vaught J, Young TR, Moore HM; GTEx Consortium. A Novel Approach to High-Quality Postmortem Tissue Procurement: The GTEx Project. Biopreserv Biobank 2015;13:311-319

24. Weil D, Boutain S, Audibert A, Dautry F. Mature mRNAs accumulated in the nucleus are neither the molecules in transit to the cytoplasm nor constitute a stockpile for gene expression. RNA 2000;6:962-75

25. Miano V, Ferrero G, Reineri S, Caizzi L, Annaratone L, Ricci L, Cutrupi S, Castellano I, Cordero F, De Bortoli M. Luminal long non-coding RNAs regulated by estrogen receptor alpha in a ligand-independent manner show functional roles in breast cancer. Oncotarget 2016;7:3201-16

26. Jin X, Sun T, Zhao C, Zheng Y, Zhang Y, Cai W, He Q, Taira K, Zhang L, Zhou D. Strand antagonism in RNAi: an explanation of differences in potency between intracellularly expressed siRNA and shRNA. Nucleic Acids Res 2012; 40:1797-806

27. Goldhirsch A, Winer EP, Coates AS, Gelber RD, Piccart-Gebhart M, Thürlimann B, Senn HJ. Personalizing the treatment of women with early breast cancer: highlights of the St Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2013. Ann Oncol 2013;24:2206-23

28. Coates AS, Winer EP, Goldhirsch A, Gelber RD, Gnant M, Piccart-Gebhart M, Thürlimann B, Senn HJ. Tailoring therapies--improving the management of early breast cancer: St Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2015. Ann Oncol. 2015;26:1533-46

29. Nagahori T, Dohi M, Matsumoto K, Saitoh K, Honda ZI, Nakamura T, Yamamoto K. Interferon-gamma upregulates the c-Met/hepatocyte growth factor receptor expression in alveolar epithelial cells. Am J Respir Cell Mol Biol 1999;21:490-7

30. Liang G, Lin JC, Wei V, Yoo C, Cheng JC, Nguyen CT, Weisenberger DJ, Egger G, Takai D, Gonzales FA, Jones PA. Distinct localization of histone H3 acetylation and H3-K4

methylation to the transcription start sites in the human genome. Proc Natl Acad Sci USA 2004;101:7357-62

31. Wolff EM, Byun HM, Han HF, Sharma S, Nichols PW, Siegmund KD, Yang AS, Jones PA, Liang G. Hypomethylation of a LINE-1 promoter activates an alternate transcript of the MET oncogene in bladders with cancer. PLoS Genet 2010;6:e1000

32. Mätlik K, Redik K, Speek M. L1 antisense promoter drives tissue-specific transcription of human genes. J Biomed Biotechnol 2006; 71:753-68

33. Nigumann P, Redik K, Mätlik K, Speek M. Many human genes are transcribed from the antisense promoter of L1 retrotransposon. Genomics 2002;79:628-634

34. Lin L, Wang Z, Prescott MS, van Dekken H, Thomas DG, Giordano TJ, Chang AC, Orringer MB, Gruber SB, Moran JV, Glover TW, Beer DG. Multiple forms of genetic instability within a 2-Mb chromosomal segment of 3q26.3-q27 are associated with development of esophageal adenocarcinoma. Genes Chromosomes Cancer 2006;45:319-31

35. Ho-Yen CM, Green AR, Rakha EA, Brentnall AR, Ellis IO, Kermorgant S, Jones JL. C-Met in invasive breast cancer: is there a relationship with the basal-like subtype? Cancer 2014;120:163-71

36. Scarpino S, Di Napoli A, Rapazzotti-Onelli M, Pilozzi E, Ruco L. Papillary carcinoma of the thyroid: methylation is not involved in the regulation of MET expression. Br J Cancer. 2004; 91:703-6

37. Yan D, Dong Xda E, Chen X, Wang L, Lu C, Wang J, Qu J, Tu L. MicroRNA-1/206 targets c-Met and inhibits rhabdomyosarcoma development. J Biol Chem 2009;284:29596-604

38. Luo W, Huang B, Li Z, Li H, Sun L, Zhang Q, Qiu X, Wang E. MicroRNA-449a is downregulated in non-small cell lung cancer and inhibits migration and invasion by targeting c-Met. PLoS One 2013;8:e64759

39. Sun C, Liu Z, Li S, Yang C, Xue R, Xi Y, Wang L, Wang S, He Q, Huang J, Xie S, Jiang W, Li D. Down-regulation of c-Met and Bcl2 by microRNA-206, activates apoptosis, and inhibits tumour cell proliferation, migration and colony formation. Oncotarget 2015;6:25533-74

40. Poliseno L, Salmena L, Zhang J, Carver B, Haveman WJ, Pandolfi PP. A coding-independent function of gene and pseudogene mRNAs regulates tumour biology. Nature 2010;465:1033-8

FIGURE LEGENDS

Figure 1. A) Graphical representation of the L1 element located within intron 2 of *MET*; L1 is inserted in the opposite orientation of *MET* with the sense promoter (SP) and the 5'-UTR encompassing the recently identified ORF0 sequence; ORF contains the antisense promoter (ASP) leading to the development of a chimeric L1-*MET* transcript. B) Poly(A)+ RNA-seq experiments from the ENCODE Consortium using the WashU Epigenome Browser; genomic coverage of the *MET* locus was carried out for whole cell, nucleus, and cytosol of MCF-7 and A549 cell lines; whole cell datasets were available only for HCT116 and HMEC cell lines; L1PA2 indicates the genomic location of the L1 element within *MET*; peaks represent all *MET*-transcribed sequences; MCF-7 and A549 cells display RNA-seq signals, located in the intron 2 of *MET* (squares) and corresponding to the L1-*MET* mRNA, which is absent in HCT116 and HMEC cell lines.

Figure 2. The L1-*MET* transcript analysis. A) L1-*MET* transcript amplification obtained with primer F1 and R5. Bands with different molecular weights were detected in MCF-7 and A549 cells, whereas no amplification was found in HCT116 and HMEC cells (expected size range from 553 bp to 794 bp). B) L1-*MET* transcript length with increasing size (exon 5 to exon 21) in MCF-7 and A549 cells; amplifications were obtained using the forward primer (F1) and different reverse primers located on *MET* exons 5, 8, 14, and 21; expected sizes: exon 5 (553-794 bp), exon 8 (943 bp), exon 14 (1894 bp), exon 21 (2894 bp). C) Graphical representation of the L1-*MET* transcript arising from the L1 element located within intron 2 of *MET*; the antisense promoter sequence (ASP) of the L1 ORF0 leading to the development of different size chimeric L1-*MET* transcripts; MW, molecular weight.

Figure 3. The L1-*MET* alternative transcript analysis. A) L1-*MET* splice variant amplifications in MCF-7 and A549 cells; amplifications were performed using specific forward primers located in each splice junctions (expected size: SV1 271 bp, SV2 193 bp, SV3 117 bp, SV4 274 bp, SV5 194 bp, SV6 122 bp). B) Scheme of the six L1-*MET* splice variants; the non-canonical *MET* transcripts induced by L1 ASP activation are reported in red boxes; SD1, SD2 and SD3 represent the three

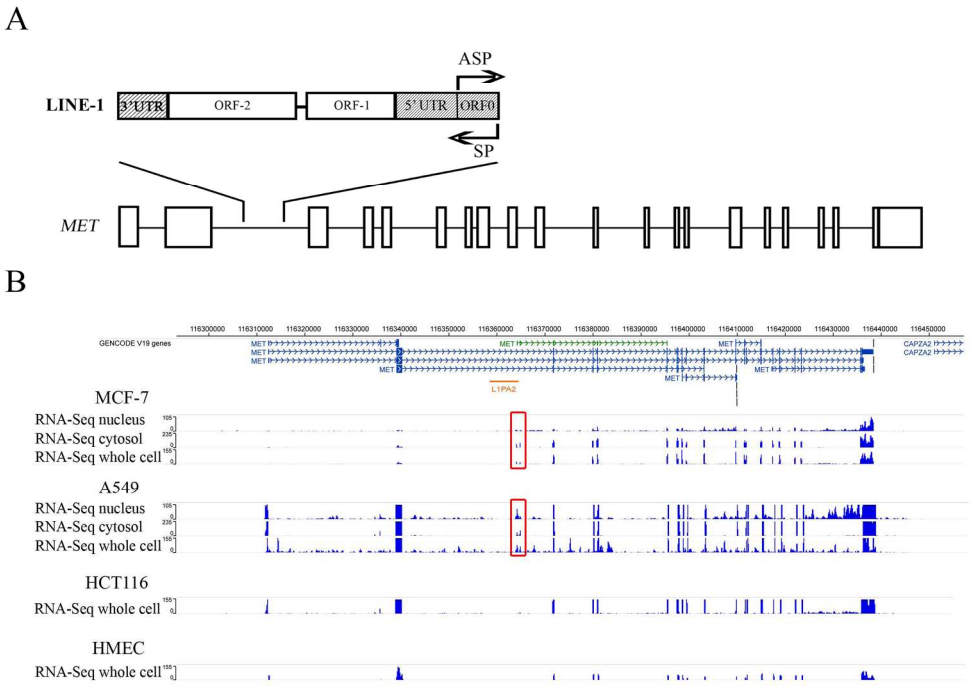
splice donor sites, whereas SA1, SA2, and SA3 are the splice acceptor sites involved in SV development; on the right, the corresponding sequences with the donor-acceptor splice junctions, highlighted in red squares, are reported. C) SV expression profiles with standard deviation (error bars) in MCF-7 (blue) and A549 (red) cancer cells; SV1 is predominant in A549 cells, whereas the MCF-7 cells show higher expression of the SV2 isoform. SV, splicing variant; MW, molecular weight; SD, splicing donor; SA, splicing acceptor.

Figure 4. L1-*MET* and *MET* expression in breast cancer cell lines. A) Correlation plot of the L1-*MET* transcript (reported as logarithm of the relative quantification) and the L1-ASP methylation level (reported as percentage of methylation). B) Correlation plot of the L1-*MET* and *MET* mRNA level in breast cancer cell lines. C) Western blot analysis of the MET protein expression in breast cancer cell lines. ASP, anti-sense promoter; RQ, relative quantification.

Figure 5. Graphical representation of the gene expression of L1-*MET* and *MET* in comparison with tumour grading, L1-*MET*/*MET* ratio, and MET protein expression. In the upper line the colour representation of the tumour grading: G1 (light green), G2 (dark green), G3 (red); patient with in situ breast cancer in white. In the second and third lines the heatmaps of L1-*MET* and *MET* gene expression in the breast cancer samples according to the IHC classification: red is for high expression and green for a low expression. The ratio of the gene expression between L1-*MET* and *MET* (L1-*MET*/*MET*) is described as a histogram with bars of different colours depending on the increasing expression of L1-*MET* (violet) or *MET* (blue). In the lower line the colour representation of the immunohistochemistry results for MET protein expression: 0 (light green), 1+ (dark green), 2+ (brown) and 3+ (red). TNBC, triple negative breast cancer; IHC, immunohistochemistry

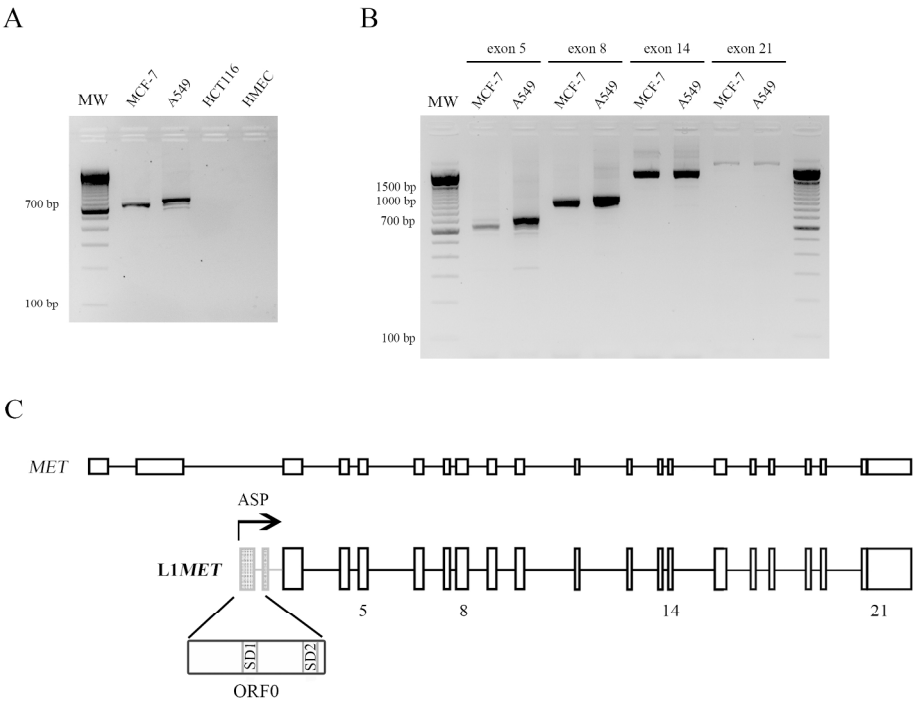
Figure 6. DFS of the breast cancer patients according to L1-*MET* expression; Kaplan-Meier curves of high-expressing L1-*MET* patient samples in dashed and low-expressing L1-*MET* patient samples in solid; cut-off value for discriminating the patients derived from ROC analysis (1.21).

DFS, disease free survival



A) Graphical representation of the L1 element located within intron 2 of *MET*; L1 is inserted in the opposite orientation of *MET* with the sense promoter (SP) and the 5'-UTR encompassing the recently identified ORF0 sequence; ORF contains the antisense promoter (ASP) leading to the development of a chimeric L1-*MET* transcript. B) Poly(A)+ RNA-seq experiments from the ENCODE Consortium using the WashU Epigenome Browser; genomic coverage of the *MET* locus was carried out for whole cell, nucleus, and cytosol of MCF-7 and A549 cell lines; whole cell datasets were available only for HCT116 and HMEC cell lines; L1PA2 indicates the genomic location of the L1 element within *MET*; peaks represent all MET-transcribed sequences; MCF-7 and A549 cells display RNA-seq signals, located in the intron 2 of *MET* (squares) and corresponding to the L1-*MET* mRNA, which is absent in HCT116 and HMEC cell lines.

219x156mm (300 x 300 DPI)



The L1-*MET* transcript analysis. A) L1-MET transcript amplification obtained with primer F1 and R5. Bands with different molecular weights were detected in MCF-7 and A549 cells, whereas no amplification was found in HCT116 and HMEC cells (expected size range from 553 bp to 794 bp). B) L1-*MET* transcript length with increasing size (exon 5 to exon 21) in MCF-7 and A549 cells; amplifications were obtained using the forward primer (F1) and different reverse primers located on *MET* exons 5, 8, 14, and 21; expected sizes: exon 5 (553-794 bp), exon 8 (943 bp), exon 14 (1894 bp), exon 21 (2894 bp). C) Graphical representation of the L1-*MET* transcript arising from the L1 element located within intron 2 of *MET*; the antisense promoter sequence (ASP) of the L1 ORF0 leading to the development of different size chimeric L1-*MET* transcripts; MW, molecular weight.

227x163mm (300 x 300 DPI)

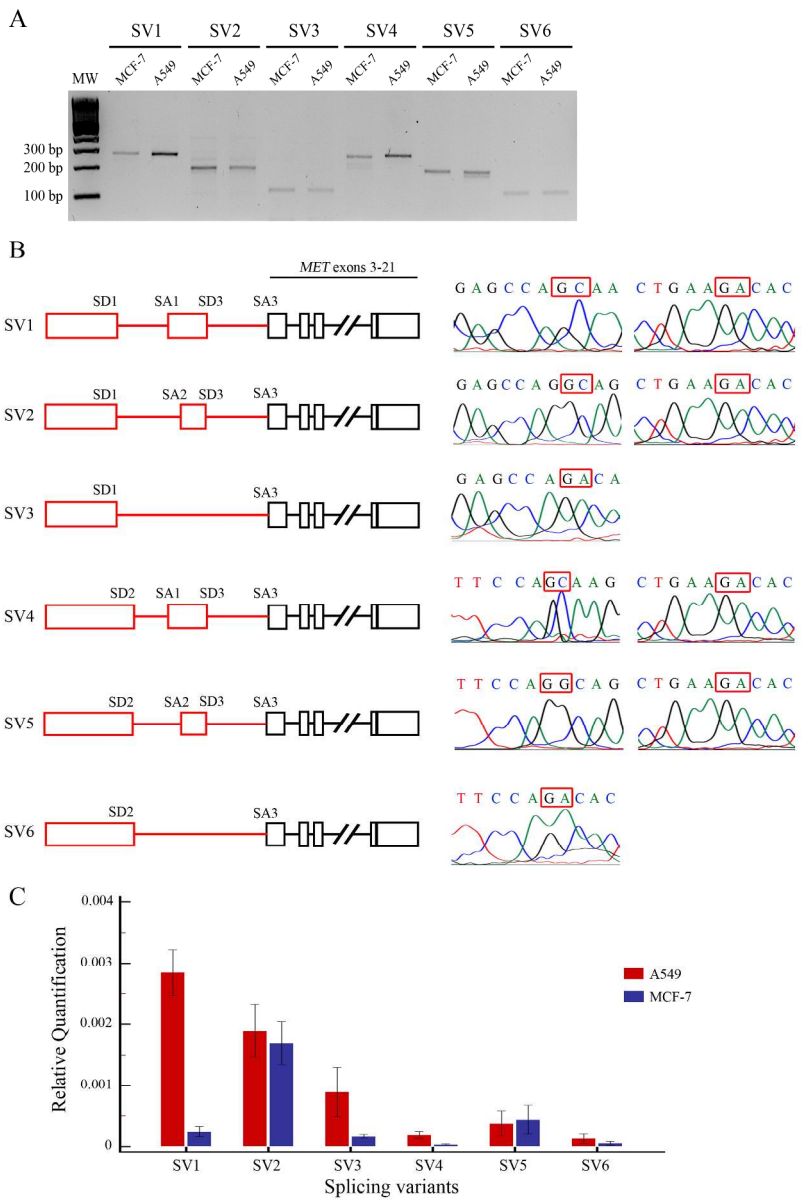


Figure 3. The L1-*MET* alternative transcript analysis. A) L1-*MET* splice variant amplifications in MCF-7 and A549 cells; amplifications were performed using specific forward primers located in each splice junctions (expected size: SV1 271 bp, SV2 193 bp, SV3 117 bp, SV4 274 bp, SV5 194 bp, SV6 122 bp). B) Scheme of the six L1-*MET* splice variants; the non-canonical *MET* transcripts induced by L1 ASP activation are reported in red boxes; SD1, SD2 and SD3 represent the three splice donor sites, whereas SA1, SA2, and SA3 are the splice acceptor sites involved in SV development; on the right, the corresponding sequences with the donor-acceptor splice junctions, highlighted in red squares, are reported. C) SV expression profiles with standard deviation (error bars) in MCF-7 (blue) and A549 (red) cancer cells; SV1 is predominant in A549 cells, whereas the MCF-7 cells show higher expression of the SV2 isoform. SV, splicing variant; MW, molecular weight; SD, splicing donor; SA, splicing acceptor.

209x297mm (300 x 300 DPI)

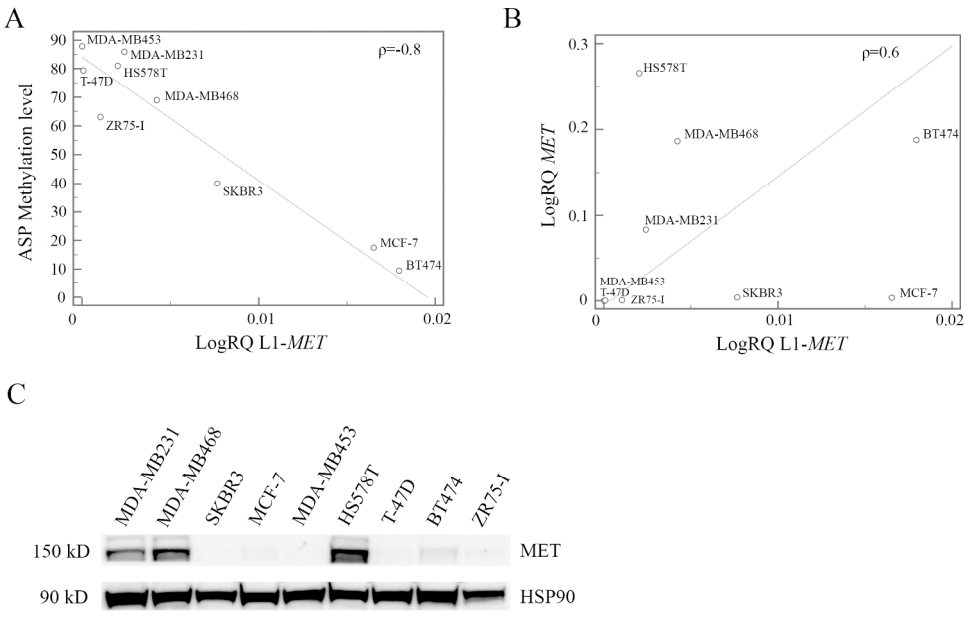


Figure 4. L1-*MET* and *MET* expression in breast cancer cell lines. A) Correlation plot of the L1-*MET* transcript (reported as logarithm of the relative quantification) and the L1-ASP methylation level (reported as percentage of methylation). B) Correlation plot of the L1-*MET* and *MET* mRNA level in breast cancer cell lines. C) Western blot analysis of the MET protein expression in breast cancer cell lines. ASP, anti-sense promoter; RQ, relative quantification.

217x147mm (300 x 300 DPI)

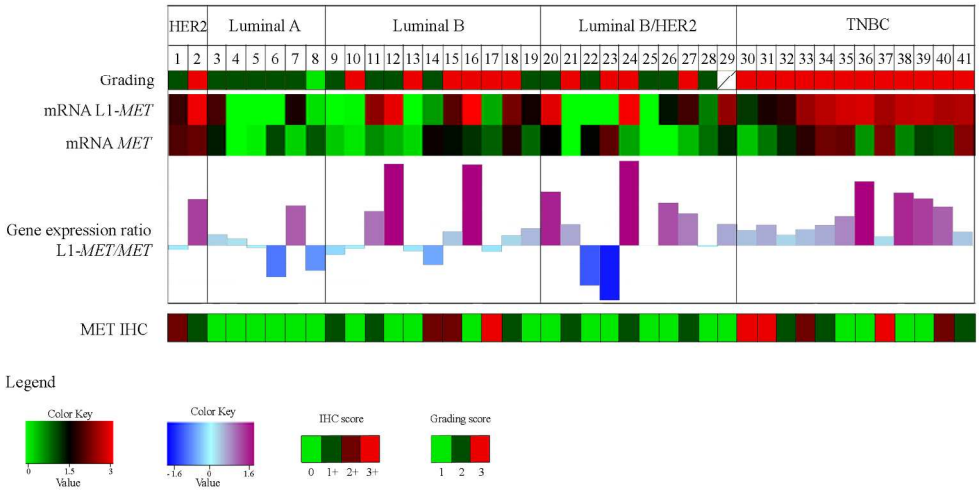
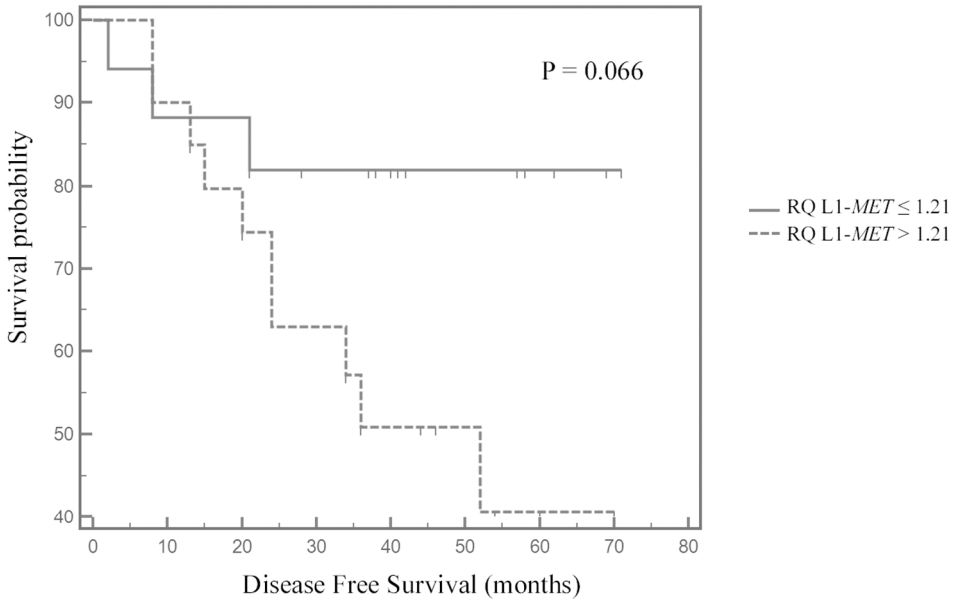


Figure 5. Graphical representation of the gene expression of L1-*MET* and *MET* in comparison with tumour grading, L1-*MET*/*MET* ratio, and MET protein expression. In the upper line the colour representation of the tumour grading: G1 (light green), G2 (dark green), G3 (red); patient with in situ breast cancer in white. In the second and third lines the heatmaps of L1-*MET* and *MET* gene expression in the breast cancer samples according to the IHC classification: red is for high expression and green for a low expression. The ratio of the gene expression between L1-*MET* and *MET* (L1-*MET*/*MET*) is described as a histogram with bars of different colours depending on the increasing expression of L1-*MET* (violet) or *MET* (blue). In the lower line the colour representation of the immunohistochemistry results for MET protein expression: 0 (light green), 1+ (dark green), 2+ (brown) and 3+ (red).TNBC, triple negative breast cancer; IHC, immunohistochemistry.

210x119mm (299 x 299 DPI)



DFS of the breast cancer patients according to L1-*MET* expression; Kaplan-Meier curves of high-expressing L1-*MET* patient samples in dashed and low-expressing L1-*MET* patient samples in solid; cut-off value for discriminating the patients derived from ROC analysis (1.21).‼ ┼ DFS, disease free survival‼ ┼

118x76mm (300 x 300 DPI)