Role of the long non-coding RNA PVT1 in the dysregulation of the ceRNA-ceRNA network in human breast cancer

F. Conte[1], G. Fiscon[1], and P. Paci[1]

(1) Institute for Systems Analysis and Computer Science "A. Ruberti" (IASI) National Research Council (CNR) of Rome

Abstract

Keywords: Networks analysis, Epigenetics, ceRNA, long non coding PVT1, miR-200

Recent findings have identified competing endogenous RNAs (ceRNAs) as the drivers in many disease conditions, including cancers. The ceRNAs indirectly regulate each other by reducing the amount of microRNAs (miRNAs) available to target messenger RNAs (mRNAs). The ceRNA interactions mediated by miRNAs are modulated by a titration mechanism, i.e. large changes in the ceRNA expression levels either overcome, or relieve, the miRNA repression on competing RNAs; similarly, a very large miRNA overexpression may abolish competition. The ceRNAs are also called miRNA "decoys" or miRNA "sponges" and encompass different RNAs competing with each other to attract miRNAs for interactions: mRNA, long non-coding RNAs (lncRNAs), pseudogenes, or circular RNAs. Recently, we developed a computational method for identifying ceRNA-ceRNA interactions in breast invasive carcinoma. We were interested in unveiling which lncRNAs could exert the ceRNA activity. We found a drastic rewiring in the cross-talks between ceRNAs from the physiological to the pathological condition. The main actor of this dysregulated lncRNA-associated ceRNA network was the lncRNA PVT1, which revealed a net biding preference towards the miR-200 family members in normal breast tissues. Despite its up-regulation in breast cancer tissues, mimicked by the miR-200 family members, PVT1 stops working as ceRNA in the cancerous state. The specific conditions required for a ceRNA landscape to occur are still far from being determined. Here, we emphasized the importance of the relative concentration of the ceRNAs, and their related miRNAs. In particular, we focused on the withdrawal in breast cancer tissues of the PVT1 ceRNA activity and performed a gene expression and sequence analysis of its multiple isoforms. We found that the PVT1 isoform harbouring the binding site for a representative miRNA of the miR-200 family shows a drastic decrease in its relative concentration with respect to the miRNA abundance in breast cancer tissues, providing a plausibility argument to the breakdown of the sponge program orchestrated by the oncogene PVT1.

References

PLOS ONE

https://doi.org/10.1371/journal.pone.0171661

RESEARCH ARTICLE

Role of the long non-coding RNA PVT1 in the dysregulation of the ceRNA-ceRNA network in human breast cancer

Federica Conte[1], Giulia Fiscon[1], Matteo Chiara[2], Teresa Colombo[1], Lorenzo Farina[3], Paola Paci[1]*

1 Institute for Systems Analysis and Computer Science "Antonio Ruberti", National Research Council, Rome, Italy, 2 Department of Biosciences, University of Milan, Milan, Italy, 3 Department of Computer, Control and Management Engineering, "Sapienza" University, Rome, Italy

* paola.paci@iasi.cnr.it

Paci et al. BMC Systems Biology 2014, 8:83
http://www.biomedcentral.com/1752-0509/8/83

BMC Systems Biology

doi:10.1186/1752-0509-8-83

RESEARCH ARTICLE Open Access

Computational analysis identifies a sponge interaction network between long non-coding RNAs and messenger RNAs in human breast cancer

Paola Paci[1*], Teresa Colombo[2] and Lorenzo Farina[3]

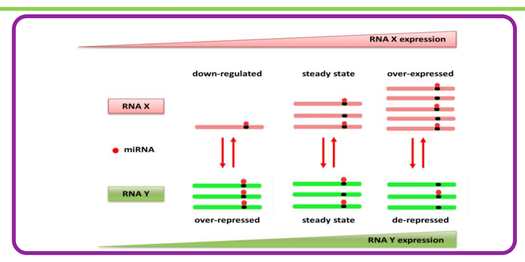
Aim

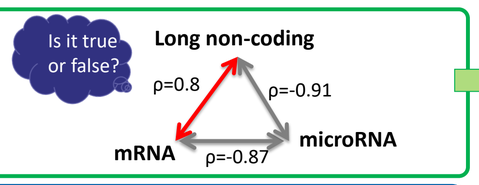
- **Proposing a data-driven approach to explore the ability of lncRNAs to act as ceRNAs protecting mRNAs from miRNA repression in breast invasive carcinoma**
- **Unveiling which lncRNAs could exert the ceRNA activity**
- **Focusing on lncRNA PVT1 activity as sponge modulator of the activity of the miR-200 family members on their targets and on the withdrawal of its decoy service in breast cancer tissues**

Methods

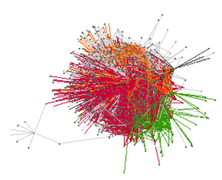
Model hypothesis

1. mRNA-lncRNA pairs must be highly correlated (Pearson correlation ρ > 0.7)
2. The correlation in the selected pairs must be mediated by one or more microRNAs (indirect correlation)
3. The selected mRNA-lncRNA pairs must share one or more binding sites for at least one microRNA mediating their interaction



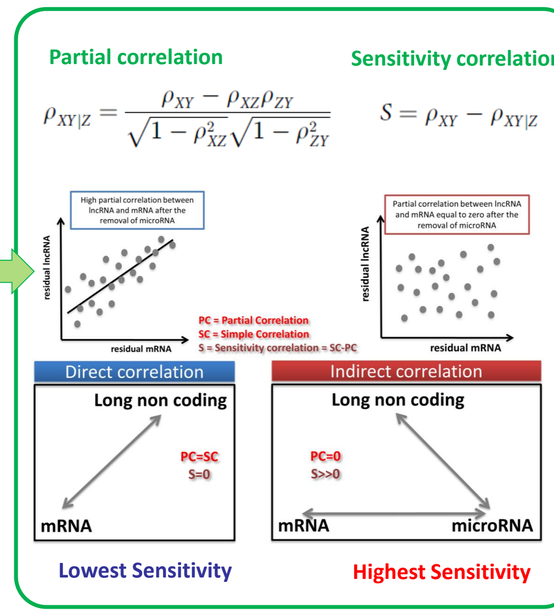


MicroRNA-Mediated Interaction (MMI) network



Linked nodes: mRNA-lncRNA pairs satisfying the 3 hypothesis

Edges: microRNAs sharing the binding site between mRNA-lncRNA pairs



Data collection

Gene expression analysis

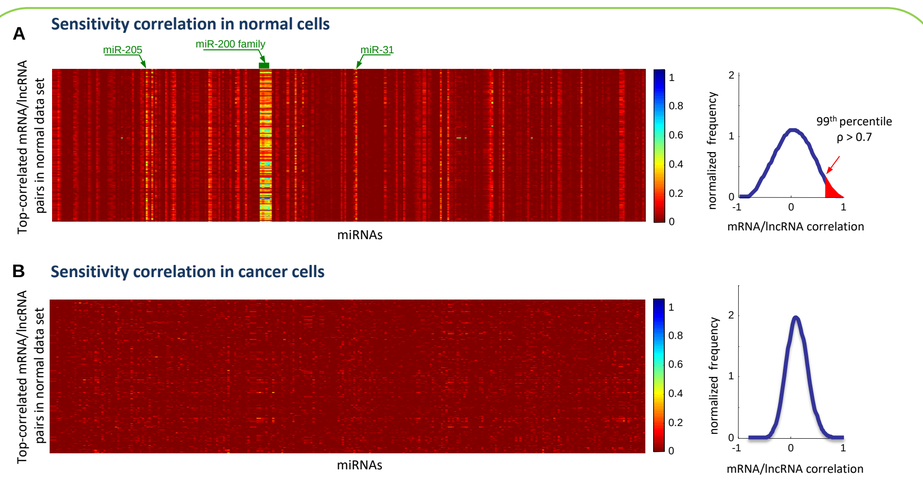
- Collections of tumour and normal expression data (FPKM) from high-throughput RNA- and miRNA-sequencing of breast invasive carcinoma were downloaded from the TCGA data portal.
- The analysis was restricted to 72 individuals for which the complete sets of tumour and matched-normal profiles (for both RNA and miRNA data) were available.
- A total of 10492 mRNAs, 311 miRNAs, and 833 lncRNAs were analysed.

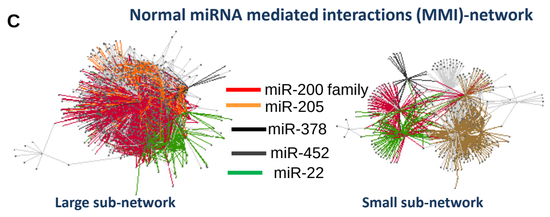
Sequence analysis of PVT1 isoforms

- Mapped read data (bam files) for the 72 patients were downloaded from the TCGA via controlled access.
- For each patient the relative two bam files (one for breast tumour and one normal sample) are used as input for the Cufflinks software in order to assemble transcripts and to estimate the relative abundances (FPKM) of these transcripts.
- The PVT1 locus assembled by Cufflinks was compared with genome annotations provided by Ensembl (release Homo sapiens GRCh37).

Results

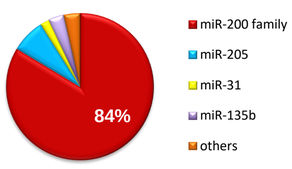
ceRNA-ceRNA interaction network in human breast





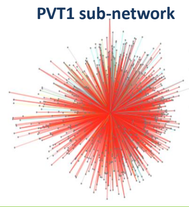
Heat-map representing the sensitivity correlation, computed using: **(A)** the normal expression data for the top-correlated mRNA/lncRNA pairs (N = 87398) in the normal dataset. Light vertical stripes point to a little pool of miRNAs that is responsible for the high correlation between all top-correlated mRNA/lncRNA pairs. Top-correlated mRNA/lncRNA pairs: Pearson correlation values exceeding the 99th percentile of the overall correlation distribution; **(B)** the cancer expression data for the top-correlated mRNA/lncRNA pairs in the normal dataset. **(C)** The normal MMI-network (1738 nodes and 32375 edges).

PVT1 edges in normal MMI-network



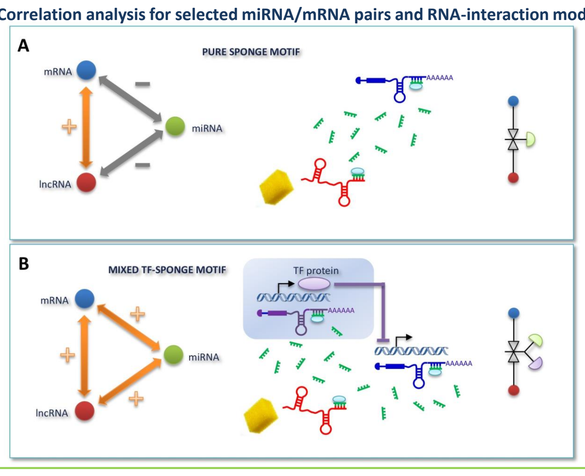
Nearest-neighbors of PVT1

PVT1 sub-network



- PVT1 is the first hub
- PVT1 is linked to important cancer genes
- PVT1 interacts with more than 50% of the all mRNAs in the whole network

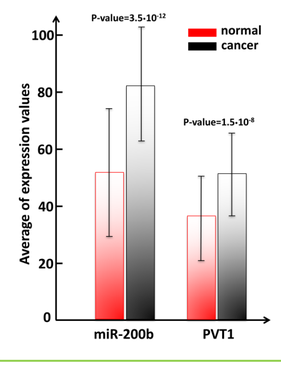
Correlation analysis for selected miRNA/mRNA pairs and RNA-interaction modules



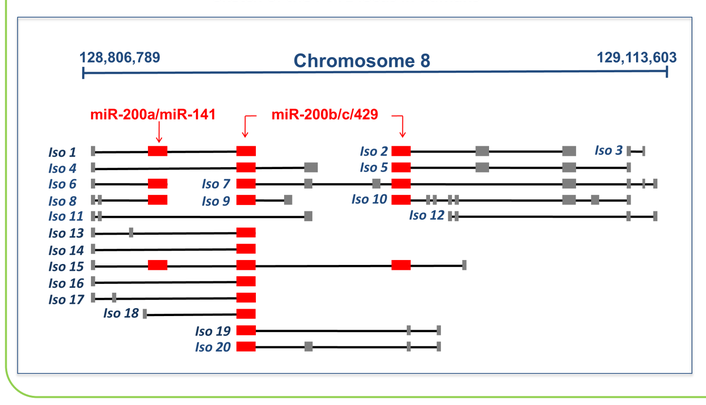
(A) the pure sponge module, where the mRNA/miRNA correlation is negative; **(B) the mixed TF-sponge module**, where the mRNA/miRNA correlation is positive and could be due to the presence of a repressor TF. In both cases, the lncRNA and the mRNA compete for miRNA binding (panels A and B, center). In the panels A and B (right), the valve symbols provide a schematic representation of the regulatory circuit: **(A)** the mRNA (blue circle) concentration is due to the presence of the miRNA (valve) that is in turn regulated by the lncRNA (red circle); **(B)** the mRNA (blue circle) concentration is due to the presence of a repressor TF regulated by the miRNA that is in turn regulated by the lncRNA.

Why PVT1 stops working as sponge regulator of mir-200 family in breast cancer?

PVT1 and miR-200b expression levels in human breast cancer tissues

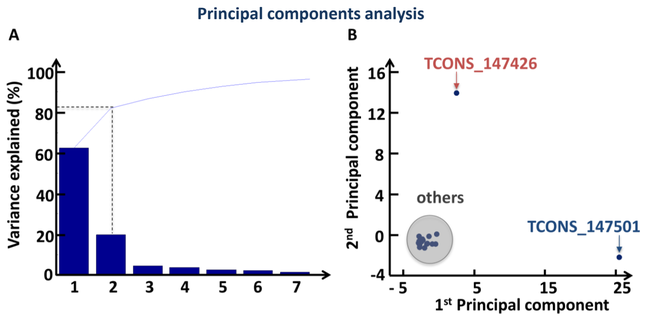


Sketch of the PVT1 locus in humans



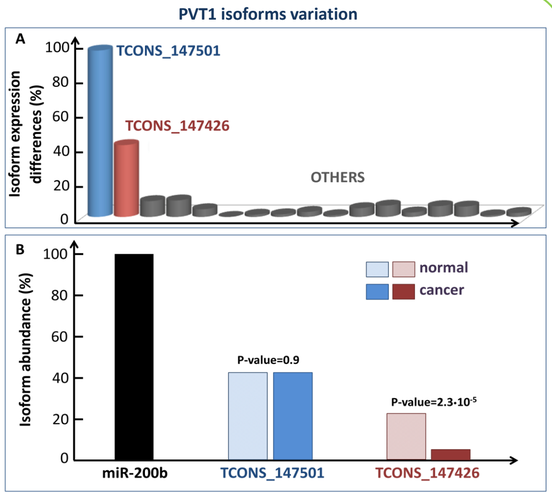
Model of PVT1 genomic locus as reconstructed by Cufflinks spans across a genome interval of over 300 kb (i.e. bases 128,806,789-129,113,603 within the February 2009 human genome build GRCh37/hg19) on the forward strand of chromosome 8. The large PVT1 locus gives rise to 91 different variants according to raw RNA-seq data of TCGA for breast invasive carcinoma. The isoform names correspond to an increasing symbolic numbering and not to the actual nomenclature of the PVT1 variants. Lines represent introns and boxes (red and grey) represent exons. Red boxes correspond to the binding sites for the miR-200 family members. Note that some isoforms lack such binding sites (e.g. Iso11 and Iso12).

Principal components analysis



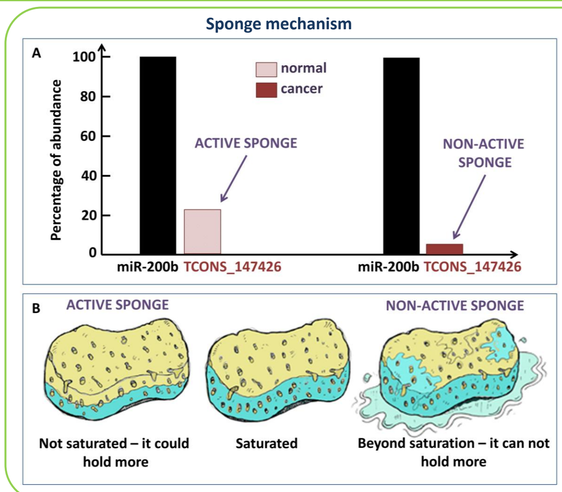
(A) The percent variability explained by each principal component. It is a type of chart, called Pareto chart, that contains both bars and a line graph, where individual values are represented in descending order by bars, and the line represents the cumulative total value. In particular, the y-axis represents the percentage of the data variance explained by each principal component, whereas the x-axis represents the principal components that are able to explain the first 100% of the cumulative distribution. The PCA is performed using the variations of all the isoforms between normal and cancer tissues. Two components explain more than the 80% of the variance of the data. **(B) The scatter plot (score plot)** of the projection of the original data (i.e. the variations of all the isoforms between normal and cancer tissues) onto the first two PCs; the x-axis contains the first PC while the y-axis contains the second PC. In this plot, it is possible to group isoforms in three classes: the isoform missing the binding site for the miR-200 family members (blue isoform, TCONS_147501), the isoform with the seed match for the miR-200b/200c/429 cluster (red isoform, TCONS_147426), and all the others. The first PC, which explains about the 60% of the variance in the original data, is able to separate the variation of the blue isoform from the others; the second PC, which explains about the 20% of the variance in the original data, is able to separate the variation of the red isoform from the others.

PVT1 isoforms variation



(A) The variations between cancer and normal tissues of all the PVT1 isoforms with respect to the variation of the blue isoform lacking the binding site for the miR-200 family members. The red and blue isoforms are the only isoforms that change. **(B)** The ratio between the abundance of the blue and red isoform with respect to the miR-200b in both normal (striped rectangle) and cancer tissues (full boxes).

Sponge mechanism



(A) The ratio between the abundance of the red isoform (with the binding site for the miR-200b/200c/429 cluster) over the abundance of the miR-200b in both normal (striped rectangle) and cancer tissues (full boxes). **(B)** Our hypothesis for the breakdown of the sponge mechanism in breast cancer orchestrated by PVT1 and the miR-200 family members It is like in the case of a real sponge: before saturation the sponge can hold more water, beyond saturation-there is too much water - the sponge can't hold more.

We speculate that in the normal tissues only **the isoform of PVT1 gene harbouring the binding site for the miR-200b/200c/429 cluster** acts as sponge regulator of the miR-200 family members. **In cancer** tissues it **stops working as sponge since its concentration is much lower than the concentration of the miR-200 family members**.