

## A next generation sequencing-based approach to identify piRNAs in breast cancer cells

C. Cantarella<sup>1</sup>✉, C. Stellato<sup>1</sup>, G. Giurato<sup>1</sup>, M.R. De Filippo<sup>2</sup>, R. Tarallo<sup>1</sup>, A. Weisz<sup>3</sup>

<sup>1</sup>Laboratory of Molecular Medicine and Genomics, University of Salerno, Baronissi, Italy

<sup>2</sup>Fondazione IRCCS SDN, Naples, Italy

<sup>3</sup>Division of Molecular Pathology and Medical Genomics, 'SS. Giovanni di Dio e Ruggi d'Aragona' Hospital, University of Salerno, Salerno, Italy

### Motivations

There are RNA species produced by eukaryotes, classified as non-coding RNAs, which are not involved in translation, but play a major role in regulation of gene expression at the transcriptional, post-transcriptional and translational level. The small non-coding RNA (sncRNA) molecules, along with the Argonaute family of proteins, have been identified as key players in various forms of sequence-specific gene silencing, including RNA interference (RNAi), translational repression and heterochromatin formation. Next-generation sequencing methods have allowed researchers to quickly sequence and profile sncRNA populations. Recently, a new class of non-coding small RNAs has been found, named Piwi-interacting RNAs (piRNAs), which interact with Piwi and are produced by a Dicer-independent mechanism. These non-coding small RNAs regulate a series of small RNA-mediated mechanisms that modulate a large variety of biological processes, such as silencing of selfish DNA elements, development, genome stability, and DNA integrity maintenance. It was reported that piRNA-pathway disorders increase the repeats of retrotransposon and cause DNA damage, both common occurrence in tumorigenesis of germline and somatic cells. Currently, piRNAs have been identified in human cancer cells, regulated by the Hili protein and involved in carcinogenesis. Another report describes how piRNAs are aberrantly expressed in human cancer cells. Other small molecules playing a central role in many RNAi mechanism are: small-interfering RNAs (siRNA) and transcription initiation RNAs (tiRNAs). While it has not been definitively demonstrated that they play a causal role in human disease, this does not necessarily preclude the existence of disease-mediating siRNAs or tiRNAs. Recent studies showed as endogenous siRNAs are derived from naturally occurring double-stranded RNAs (dsRNAs) and have roles in the regulation of gene expression in mouse oocytes. On the other hand, tiRNAs

are nuclear localized 18nt RNAs derived from sequences immediately downstream of RNA polymerase II (RNAPII) transcription start sites. Several reports have shown that tiRNAs are intimately correlated with gene expression, RNA polymerase II binding and behaviours, and epigenetic marks associated with transcription initiation, but not elongation.

### Methods

Small non coding RNAs sequences were obtained by sequencing human cancer cells RNA with an Illumina Genome Analyzer. In order to identify piRNAs expressed in the samples, we used the miRanalyzer tool with the reference databanks RNAdb and Rfam. Two databanks were filtered to reduce redundancies and integrated with data available in public repositories. Programs developed in house (using Perl and PHP languages) have been applied to identify piRNAs not yet annotated, called putative "ping-pong piRNAs". Software is based on the following experimental evidences: there are primary (p-piRNAs) and secondary (s-piRNAs) piRNAs; the initial cleavage site is located within the base-paired region 10nt downstream of the 5' terminal U of the p-piRNA and the resulting s-piRNAs are therefore distinguished by an A at position 10. Some s-piRNAs are expected to be reverse complementary to the original p-piRNA precursor transcript and may themselves be able to direct cleavage of these to recreate the original p-piRNA. During piRNA-guided cleavage of target transcripts a 19-mer product arises. Although 19-mers do appear to associate with Piwi proteins, the fact that they are not stabilized by 3' end methylation argues against their function as piRNAs. Their size and the apparent lack of the 3' end modification would allow them to load into Argonaute proteins and function in RNAi-like silencing as short interfering RNAs. Results have been compared using the recently developed program piRNAPredictor, which uses a k-mer scheme to identify piRNA se-

quences, relying on the training sets of non-piRNA and piRNA sequences of five model species.

### Results

Starting from available tracks of human piRNAs, a non-redundant databank has been assembled removing sequences with ambiguous nucleotides and sequences mapping within the same locus of the human genome but annotated with different accession number. The databank has been used to identify piRNAs molecules in two distinct sequencing experiments, with different coverages. Some piRNAs are differentially expressed in samples respect to controls, revealing their possible involvement in breast cancer.

The programs developed to identify piRNAs not yet annotated, produced by "ping-pong model", have been tested both with reference piRNA sequences and experimental data. Analysis of sequences reads highlights the presence of the three molecules produced during piRNA biogenesis: primary, secondary and 19-mer piRNA.

### Acknowledgements

Research supported by: Fondazione con il Sud; Italian Association for Cancer Research; Italian Ministry for Education, University and Research; Regione Campania; University of Salerno; Fondazione Umberto Veronesi.