

## Exploring the dark matter of the human genome by oligonucleotide based molecules

Concetta Avitabile<sup>1</sup>, Luca Domenico D'Andrea<sup>1</sup>, Alessandra Romanelli<sup>2\*</sup>

<sup>1</sup> Institute of Biostructure and Bioimaging, CNR, via Mezzocannone 16, 80134 Naples (Italy)

<sup>2</sup> University of Naples "Federico II", Department of Pharmacy, via Mezzocannone 16, 80134 Naples (Italy)

e-mail: alessandra.romanelli @unina.it

\*: corresponding author.

Keywords: non-coding RNA, oligonucleotides, PNA, LNA, RNA- RNA and RNA-protein interactions.

The sequencing of the human genome, initiated with ENCODE project, substantially revolutionized the concept of gene; a catalogue of human transcripts has been obtained and functional elements, defined as discrete segments of the genome possessing a reproducible biochemical signature or encoding for a defined product, have been identified[1]. Very large genetic screens, defined genome wide association studies, further widened the knowledge on the organization and cross-talk between genes, questioning the Mendelian inheritance theory and suggesting for example that is not a single gene, but a large number of them responsible for a biological trait, like the height or a disease. Interestingly, only a small percentage of our genome (about 2-3%) codes for proteins, while the remaining 97-98% , the so-called "dark matter" of the genome, is represented by non-coding RNAs (ncRNAs). Typically, ncRNAs are classified based on their length in short (19-25 nucleotides) and long (200 or more nucleotides); the first to be identified were the short interfering RNAs (siRNAs) and microRNAs (miRNAs), both derived from the Dicer mediated cleavage of long double stranded RNAs and involved in the regulation of gene expression by different mechanisms [2]. Few information are available on the role of long non coding RNAs (lncRNAs) and, to complicate the situation, more and more lncRNA sequences are discovered every day. Some lncRNA sequences are reported to host the transcripts of small ncRNAs, some are known to control the expression of genes located on the same chromosome (those which act *in cis*), while other regulate the expression of genes located at independent loci (those which act *in trans*)[3]. Increasing experimental evidences support the hypothesis that a number of non-coding RNAs is involved in the organization and regulation of human genes; it has been hypothesized that genomic variants associated to diseases affect regulatory properties of non-coding RNAs in mammalian cells[4]. The network of interactions between coding and non-coding genes seems quite puzzling.

But how to unravel the knot? Which tools have been used so far and are available to discover the biological role of ncRNAs? Bioinformatics tools represent a valid resource in the search of potential interactors of ncRNAs, yet the experimental validation is always needed. If we analyze the experimental techniques

employed to illuminate the function of ncRNAs, we will notice that many of the tools employed to this aim rely on oligonucleotides or oligonucleotide analogues. The examples which follow show some applications of oligonucleotide based molecules as inhibitors of the function of ncRNAs, in loss of function studies, as handles for capturing ncRNAs and ncRNA complexes with DNA/RNA or proteins and finally as probes for their imaging.

Loss of function studies are often performed using synthetic small interfering RNAs (siRNAs), double stranded RNAs, 21-23 nt long that induce post-transcriptional gene silencing in a sequence specific fashion. These tools initially applied to silence genes in the nematode *C.elegans* and in the fruit fly *D.melanogaster*, are currently employed for the sequence specific knockdown of genes in mammalian cells and are also investigated as drugs for the treatment of a number of diseases including AMD and also viral diseases[5]. siRNAs are employed in loss of function screens aimed at identifying the function of lncRNAs involved in the regulation of different processes such as adipogenesis[6] and cancer[7, 8]. A large number of chemical modifications has been introduced into siRNAs for different scopes, for example to change the thermostability profile of siRNA duplexes, which affects the rate at which siRNA duplexes unwind in the RISC loading and RNA release processes, to increase the intracellular siRNA stability, to reduce siRNA immunogenicity[9]. Recently antisense oligonucleotides, composed of a DNA oligomer flanked at both ends by LNA nucleotides, defined LNA gapmers, have been employed to discover the function of lncRNAs, such as MALAT1 [10].

In many cases the function of ncRNAs is mediated by the interaction of ncRNAs with oligonucleotide sequences and/or with proteins. Many methods have been developed to identify the genes and proteins interacting with ncRNAs and also the details regarding the domains (of a protein or a gene) involved in the interaction. Techniques for the identification of mRNA targets of miRNAs have been deeply explored[11]; immunoprecipitation of the miRNA-mRNA pairs in complex with the RISC components Argonaute or TNRC6 followed by deep sequencing has been employed to detect the target of miR-1 and miR-124[12, 13]. Recent advances in this technique are based on the UV based cross-linking of RNA to RNA associated proteins, which precedes the immunoprecipitation and sequencing steps (HITS-CLIP). The crosslinking step followed by nuclease digestion allows for a more stringent purification of the RNA-protein complexes, ending up in the identification of the protein interacting with the RNA and also on details regarding the RNA sequence responsible for the interaction. Very efficient cross-linking occurs when the photoactivable nucleoside 4-thiouridine is used (PAR-CLIP)[14]. An alternative strategy for the identification of miRNA targets is based on 3'-biotinylated miRNAs; upon transfection, the tagged miRNA is incorporated into the RISC complex, where the association to the target mRNA occurs[15]. The miRNA-mRNA hybrids are purified on streptavidin beads and analyzed by qRT-PCR. Isolation of miRNA-mRNA complexes has also been achieved using biotinylated oligonucleotides complementary to the mRNA[16]; in this case the capture

oligonucleotide is immobilized onto streptavidin beads and then it is incubated with cell lysates, previously treated with formaldehyde. miRNA profiling is achieved by qRT-PCR.

The discovery of the function of the Xist represents a successful example in which oligonucleotide analogues are employed in experiments aimed at identifying functional regions of lncRNAs. Xist is a long non coding RNA devoted to the silencing of the X chromosome during the development of female mammalian. Initial work to map the functional domains of Xist has been performed using antisense Peptide Nucleic Acids conjugated to the carrier peptide transportan. It was found that PNAs complementary to a distinct repeat region in the first exon of Xist inhibited the binding of Xist to the X chromosome and the formation of the inactive X chromosome (Xi). [17]. In a successive study, Locked Nucleic Acids (LNAs) targeting Xist repeat C were found able to displace Xist from the inactive chromosome (Xi), suggesting the role of the repeat C in anchoring Xist to the chromosome[18].

Other techniques applied for the functional characterization of ncRNAs take advantage of the interaction of ncRNAs with chromatin [19]. Long non coding RNAs have been demonstrated to be involved in the regulation of chromatin states in different biological processes, including imprinting and developmental gene expression. The Chromatin isolation by RNA purification (ChIRP) method allows the mapping of lncRNA occupancy genome-wide at high resolution[20]. This technique exploits tiling antisense biotinylated oligonucleotides (20 mer), designed to cover all the RNA fragments that may be sheared in the process of chromatin preparation, to capture lncRNA-protein complexes crosslinked to chromatin. Complexes are isolated by streptavidin beads; RNA and DNA can be extracted and identified by sequencing. By ChIRP the genomic binding sites of the lncRNA HOTAIR, a lncRNA devoted to recruit the Polycomb Repressive Complex 2 (PRC2) and silence the expression of target genes, have been identified and downstream effects on PRC2 detected. Implementation of ChIRP with mass spectrometry analyses recently reported by Ci et al. enabled the comprehension of the mechanism by which Xist-protein particles assemble in a modular and developmentally controlled fashion and control the spreading of chromatin and the silencing of transcription[21]. Similar strategies aimed at capturing RNA-chromatin complexes are the RNA antisense purification (RAP) and the capture hybridization analysis of RNA targets (Chart), which basically differ from ChIRP in the design of oligonucleotide probes, cross-linking and chromatin shearing methods. Very recently, proteins directly associated to Xist and involved in the transcriptional silencing on the X chromosome have been identified by a methodology named RAP-MS (RNA Antisense Purification coupled to Mass Spectrometry); long biotinylated oligonucleotides antisense to Xist were initially incubated with cells, previously irradiated by UV; Xist complexes were then purified in denaturing conditions and finally proteins which directly interact with Xist were identified by quantitative mass spectrometry[22].

Studies aimed at determining the tissue and subcellular localization of ncRNAs are usually carried out by fluorescence. Of particular interest is the application of probes whose fluorescence is on only upon binding

to the target. One example is represented by the light-up probe thiazole orange (TO), inserted into a molecular beacon composed of Peptide Nucleic Acids, which has been employed for the imaging of the lncRNA CCAT1 in cell and in colorectal cancer biopsies[23]. In a different application a fluorescent hairpin DNA, designed to target a selected miRNA, was immobilized on a golden core-shell nano-probe coated with polydopamine (PDA). The fluorescence of the DNA is quenched until bound to the PGA-gold core; upon recognition of the miRNA, the DNA dissociates from the nanoparticle and the fluorescence signal is on. [24]. A large number of studies relies on the Fluorescence In Situ Hybridization (FISH) technique; the expression of low abundance bacterial small non-coding RNAs from single bacterial cells has been achieved by flow cytometry-FISH using LNA probes[25].

With these tools in our hands we have just began to have some groundings in the ncRNA world, addressing questions related to subcellular localization of ncRNAs, identity of the protein and gene interactors and also high resolution definition of the ncRNA sites devoted to the interactions. A large number of tools is available for the determination of the secondary and three-dimensional structure of ncRNAs, but the comprehension of the roles at the base of the interactions of ncRNAs with proteins and genes is still a challenge. The next and most important goal of the research on ncRNAs will be that of using the ensemble of functional and structural information to control cellular processes.

1. An integrated encyclopedia of DNA elements in the human genome. *Nature* 489(7414), 57-74 (2012).
2. Carthew RW, Sontheimer EJ. Origins and mechanisms of mirnas and sirnas. *Cell* 136(4), 642-655 (2009).
3. Kung JT, Colognori D, Lee JT. Long noncoding rnas: Past, present, and future. *Genetics* 193(3), 651-669 (2013).
4. Kapranov P, St Laurent G, Raz T *et al.* The majority of total nuclear-encoded non-ribosomal rna in a human cell is 'darkmatter' un-annotated rna (vol 8, pg 149, 2010). *Bmc Biol* 9, (2011).
5. Avitabile C, Cimmino A, Romanelli A. Oligonucleotide analogues as modulators of the expression and function of noncoding rnas (ncrnas): Emerging therapeutics applications. *J Med Chem* 57(24), 10220-10240 (2014).
6. Sun L, Goff LA, Trapnell C *et al.* Long noncoding rnas regulate adipogenesis. *Proc Natl Acad Sci U S A* 110(9), 3387-3392 (2013).
7. Zhang L, Zhou XF, Pan GF, Zhao JP. Enhanced expression of long non-coding rna zxf1 promoted the invasion and metastasis in lung adenocarcinoma. *Biomed Pharmacother* 68(4), 401-407 (2014).
8. Yang F, Huo XS, Yuan SX *et al.* Repression of the long noncoding rna-let by histone deacetylase 3 contributes to hypoxia-mediated metastasis. *Mol Cell* 49(6), 1083-1096 (2013).
9. Bramsen JB, Kjems J. Development of therapeutic-grade small interfering rnas by chemical engineering. *Front Genet* 3, 154 (2012).
10. Michalik KM, You X, Manavski Y *et al.* Long noncoding rna malat1 regulates endothelial cell function and vessel growth. *Circ Res* 114(9), 1389-1397 (2014).
11. Thomson DW, Bracken CP, Goodall GJ. Experimental strategies for microrna target identification. *Nucleic Acids Res* 39(16), 6845-6853 (2011).

12. Hendrickson DG, Hogan DJ, Herschlag D, Ferrell JE, Brown PO. Systematic identification of mrnas recruited to argonaute 2 by specific micrnas and corresponding changes in transcript abundance. *PLoS One* 3(5), e2126 (2008).
13. Ritchie W, Flamant S, Rasko JE. Predicting microrna targets and functions: Traps for the unwary. *Nat Methods* 6(6), 397-398 (2009).
14. Hafner M, Landthaler M, Burger L *et al.* Transcriptome-wide identification of rna-binding protein and microrna target sites by par-clip. *Cell* 141(1), 129-141 (2010).
15. Orom UA, Lund AH. Isolation of microrna targets using biotinylated synthetic micrnas. *Methods* 43(2), 162-165 (2007).
16. Hassan T, Smith SG, Gaughan K *et al.* Isolation and identification of cell-specific micrnas targeting a messenger rna using a biotinylated anti-sense oligonucleotide capture affinity technique. *Nucleic Acids Res* 41(6), e71 (2013).
17. Beletskii A, Hong YK, Pehrson J, Egholm M, Strauss WM. Pna interference mapping demonstrates functional domains in the noncoding rna xist. *Proc Natl Acad Sci U S A* 98(16), 9215-9220 (2001).
18. Sarma K, Levasseur P, Aristarkhov A, Lee JT. Locked nucleic acids (lnas) reveal sequence requirements and kinetics of xist rna localization to the x chromosome. *Proc Natl Acad Sci U S A* 107(51), 22196-22201 (2010).
19. Chu C, Spitale RC, Chang HY. Technologies to probe functions and mechanisms of long noncoding rnas. *Nat Struct Mol Biol* 22(1), 29-35 (2015).
20. Chu C, Qu K, Zhong FL, Artandi SE, Chang HY. Genomic maps of long noncoding rna occupancy reveal principles of rna-chromatin interactions. *Mol Cell* 44(4), 667-678 (2011).
21. Chu C, Zhang QC, Da Rocha ST *et al.* Systematic discovery of xist rna binding proteins. *Cell* 161(2), 404-416 (2015).
22. Mchugh CA, Chen CK, Chow A *et al.* The xist lncrna interacts directly with sharp to silence transcription through hdac3. *Nature* 521(7551), 232-236 (2015).
23. Kam Y, Rubinstein A, Naik S *et al.* Detection of a long non-coding rna (ccat1) in living cells and human adenocarcinoma of colon tissues using fit-pna molecular beacons. *Cancer Lett* 352(1), 90-96 (2014).
24. Choi CK, Li J, Wei K *et al.* A gold@polydopamine core-shell nanoprobe for long-term intracellular detection of micrnas in differentiating stem cells. *J Am Chem Soc* 137(23), 7337-7346 (2015).
25. Robertson KL, Vora GJ. Locked nucleic acid and flow cytometry-fluorescence in situ hybridization for the detection of bacterial small noncoding rnas. *Appl Environ Microbiol* 78(1), 14-20 (2012).