Editorial

Health or disease–why does "dark matter" matter more?

Editorial

The expression of protein-coding genes (mRNA) has been the benchmark of pathophysiological studies for decades. However, in recent years, this dogma has been challenged by novel data, namely non-coding RNAs (part of dark matter) and their interactions with binding proteins or other RNAs determine the cellular functions under both healthy and disease states and very intricate regulatory networks underlie complex diseases that require integrated analysis at systems scale. Understanding dark matter provides the opportunity to formulate new paradigms that govern biological systems and devise novel therapies and diagnostic tools. The findings of the ENCODE (Encyclopedia of DNA Elements) project[1] are clear: evidence of wide spread and heterogeneous functionalization in both protein coding and non-protein coding transcribed regions of the human genome means that the attribute 'junk DNA' for the latter regions is no longer appropriate.[2] As a result, the analysis of RNA-Seqtranscriptome profiles obtained under various healthy and pathological conditions should now be directed to 'dark matter' in order to decipher the complex dysregulation patterns which might characterize disease signatures.

In cancer, for instance, the well-known hallmarks that have governed the study of tumorgenesis will have to be redefined, taking into account aspects of dark matter such as the regulatory functions of long non-coding RNAs, as documented in.[3–5] Cancer diagnosis and therapy are in the highest need of novel input from next generation sequencing, which is acquiring a remarkably more clinical focus. More specifically, better diagnosis could come from a richer characterization of tumorigenesis phases by incorporating new oncogenic elements,[6] leading to more accurate differentiation between and within cancers. As for therapy, it is expected that new antitumor drugs designed with increased RNA specificity (i.e. alternative transcript expression) will result in more targeted therapeutic outcomes. Therefore, studying the effects of interactions between non-coding RNAs and other entities such as proteins, metabolites, and other high-throughput genomic products, including RNAs, becomes a crucial research direction. Such data analysis forces the consideration of RNA activity at systems scale and the design and implementation of forms of control of different cancer states. For instance, non-coding RNAs are currently analyzed as decoys exerting control over the mRNA, thus modifying their functions.[7] As such, they influence microRNA regulation in a network of RNA crosstalk focused on competing for mRNA binding, according the known 'sponging'[8] or also 'competing endogenous RNA' (ceRNA) law.[9]

How such mechanisms may work for a wider spectrum of RNA modulation is currently not known. The most recent examples of complex regulation networks are specific to microRNA functional synergistic,[10] microRNA-program mediated,[11] and transcriptionally cross-regulated ceRNA[12] networks, with corresponding results holding at various scales. More refined RNA-binding protein interactions (RBPome) are expected from a deeper use of CLIP (cross-linking immunoprecipitation),[13,14] and thus extra layers of complexity will be added to the current networks. The systems complexity of commonly-known regulation mechanisms underlying gene expression, involving alternative splicing, epigenetics, structural variation, among other factors, is clearly destined to change and novel inference tools will be needed. Both the systems role of non-coding RNAs and their specialization into many possible RNA-RNA associations and integrations (with clinical data) will be central to future studies. The greatest hope is that a better understanding of the web of complexRNA regulatory circuits[15] and the achievement of an extended polymorphism map[14] will offer better opportunities to identify new roles of RNA in disease, say markers or therapeutic targets.[16,17]

A major challenge related to the development of accurate analyses within a systems disease context is the inferring of regulation mechanisms, i.e. RNA-RNA interactions. This is an area of enormous clinical impact in that it depends on the study and understanding of the principles and dynamics of drug actions, in order to understand their characteristic efficacy, resistance, and toxicity. Network approaches provide great inference tools to address this challenge, by optimizing the potential of complex combinatorial treatments through identifying synergistic and antagonistic drug effects. Classical model-based approaches would instead require the specification of high-dimensional and multi-parameter settings, whose robustness and sensitivity properties require massive control to predict systems effects of perturbations (i.e. treatments) at various complexity levels (single versus multiple targeted drugs).

In conclusion, integrative inference methodologies and tools are able to address the dark matter in depth (range of expression levels) and breadth (types of variation). These tools require the development of new computational paradigms that are mostly hybrid approaches of known methods. They should be designed to address the nature of associations between bio-types, correlative or causal, the power of multiple testing and validation procedures, the application of ensemble optimization techniques to handle unknown complexity, and the integration of various types of data in the same model framework.

Acknowledgements

None.

Enrico Capobianco, Nicholas Tsinoremas
Center for Computational Science, Miller School of Medicine, University of Miami, USA

Correspondence: Nicholas Tsinoremas, Center for Computational Science, University of Miami, Miami, FL 33146, Miller School of Medicine, University of Miami, Miami, FL 33136, USA, Email ntsinoremas@gmail.com

Conflict of interest

Author declares that there is no conflict of interest.

References

1. ENCODE Project Consortium, Bernstein BE, Birney E, et al. An integrated encyclopedia of DNA elements in the human genome. *Nature*. 2012;489:57–74.

2. Clark MB, Amaral PP, Schlesinger FJ, et al. The reality of pervasive transcription. *PLoS Biol*. 2011;9(7):e1000625.

3. Gutschner T, Diederichs S. The hallmarks of cancer. a long non-coding RNA point of view. *RNA Biol*. 2012;9(6):703–719.

4. Cheetham SW, Gruhl F, Mattick JS, et al. Long noncoding RNAs and the genetics of cancer. *British Journal of Cancer*. 2013;108:2419–2425.

5. St Laurent G, Shtokalo D, Dong B, et al. VlincRNAs controlled by retroviral elements are a hallmark of pluripotency and cancer. *Genome Biol*. 2013;14(7):R73.

6. Huarte M, Rinn JL. Large non-coding RNAs: missing links in cancer? *Hum Mol Genet*. 2010;19(R2):R152–R161.

7. Poliseno L, Salmena L, Zhang J, et al. A coding-independent function of gene and pseudogene mRNAs regulates tumour biology. *Nature*. 2010;465(7301):1033–1038.

8. Marques AC, Tan J, Ponting CP. Wrangling for microRNAs provokes much crosstalk. *Genome Biol*. 2011;12(11):132.

9. Salmena L, Poliseno L, Tay Y, et al. A ceRNA hypothesis: the Rosetta Stone of a hidden RNA language? *Cell*. 2011;146(3):353–358.

10. Xu J, Li CX, Li YS, et al. MiRNA-miRNa synergistic network: construction via co-regulating functional modules and disease miRNA topological features. *Nucleic Acids Res*. 2011;39(3):825–836.

11. Sumazin P, Yang X, Chiu HS, et al. An extensive MicroRNA-mediated network of RNA-RNA interactions regulates established oncogenic pathways in glioblastoma. *Cell*. 2011;147(2):370–381.

12. Ala U, Karreth FA, Bosia C, et al. Integrated transcriptional and competitive endogenous RNA networks are cross-regulated in permissive molecular environments. *Proc Natl Acad Sci U S A*. 2013;110(18):7154–7159.

13. Rinn JL, Ule J. Oming in on RNA-protein interactions. *Genome Biology*. 2014;15:401.

14. Cech TR, Steitz JA. The noncoding RNA revolution-trashing old rules to forge new ones. *Cell*. 2014;157(1):77–94.

15. Geisler S, Coller J. RNA in unexpected places: long non-coding RNA functions in diverse cellular contexts. *Nat Rev Mol Cell Biol*. 2013;14(11):699–712.

16. Ning S, Wang P, Ye J, et al. A Global map for dissecting phenotypic variants in human lincRNAs. *Eur J Hum Genet*. 2013;21(10):1128–1133.

17. Esteller M. Non-coding RNAs in human disease. *Nat Rev Genet*. 2011;12(12):861–874.

18. Prensner JR, Chinnaiyan AM. The emergence of lncRNAs in cancer biology. *Cancer Discov*. 2011;1(5):391–407.