



# Genomic “dark matter”: implications for understanding human disease mechanisms, diagnostics, and cures

**Philipp Kapranov\* and Georges St. Laurent\***

St. Laurent Institute, Cambridge, MA, USA

\*Correspondence: philippk08@gmail.com; georgest98@yahoo.com

## WHAT IS GENOMIC “DARK MATTER?”

The realization that protein-coding genes use only a tiny fraction of the three billion base pairs that make up the human genome has given birth to perhaps the largest and most persistent question in modern genetics: of what use, if any, is the vast non-coding sequences that we all carry in each of our cells. Is it really non-functional “junk” DNA as referred to by some, or does it provide the basis for the blueprint for organismal complexity and cellular information processing, as argued by others? The massive expanse of the non-coding portion of the genome, combined with the current technological and analytical limitations inherent to its functional analysis, has resulted in a mass of conflicting ideas and conclusions. Collectively this has created an aura of mystery and doubt surrounding it, leading to the label of genomic “dark matter.” In a manner analogous to the “dark matter” of the universe, it is something that we can neither easily detect nor understand, but that nonetheless exists and is open to careful experimental queries.

## DOES IT HAVE FUNCTION?

Classical approaches, such as sequence conservation and mutagenesis, have been unable to address the question of the functionality in the non-coding realm. The simplest explanation for these observations is that the non-coding portion of the genome lacks function, but is instead a neutral passenger on the evolutionary journey. However, this leaves us with the intellectually unsatisfying conclusion that most of our genome exists for no reason. Function resonates well with an important aspect of the mysteries surrounding genomic “dark matter” – most of it is used to produce RNA. Moreover, this “dark matter” RNA is not just a minor fraction of the cell’s RNA, but rather makes up a majority of it (not counting the ribosomal or mitochondrial RNAs). RNA production is a sign of a functional DNA sequence and even more so if such RNA is abundant. While encouraging, our knowledge of the “dark matter” RNA is still rather limited in large measure due to the fact that most RNA analysis endeavors so far have focused on polyA + RNA, while detection of “dark matter” transcripts requires total RNA presumably because they either tend to be polyA – or somehow lost during polyA-selection process. Thus, this realm is ripe for discoveries, with brain and embryonic tissues likely (based on analysis of protein-coding mRNAs) harboring some of the richest reservoirs of novel “dark matter” transcripts. It is also worth noting that exons of well-characterized protein-coding transcripts can be found in unusual arrangements linking for example very distant genes, potentially due to trans-splicing, adding to the repertoire of un-annotated transcripts whose function we only now begin to understand.

## HOW COULD IT FUNCTION?

It is generally assumed that “dark matter” RNAs do not code for proteins and function via regulation of expression of other loci. Two very general themes of non-coding RNA-mediated regulation became prominent in the recent decade: modulation of chromatin state via association with chromatin modulator complexes and, the production of small RNAs from longer non-coding RNAs to regulate various layers of transcript expression. Discovery of RNA-mediated non-Mendelian inheritance of an epigenetic change in mammals uncovered a new tantalizing possibility for RNA function. It is worth noting that the basic assumption has been challenged recently by evidence suggesting existence of plethora of short peptides produced by the “dark matter” RNAs, even though they could conceivably be products of non-functional translation of bona fide non-coding RNAs. However, if the last decade has taught us anything, it is that a given locus can produce a variety of different RNAs, that can be thought of as a “transcriptional forest” as coined by the FANTOM consortium researchers, and such RNAs could well have different functions in a cell.

## DOES IT BOOST HUMAN NERVOUS SYSTEM COMPLEXITY?

The genomes of humans and flies have approximately the same information complexity – on the order of ~20K protein-coding genes, just a couple fold higher than that of yeast. Some additional reservoir of complexity should exist. The human nervous system contains widespread expression of “dark matter” RNAs, distributed in highly articulated intracellular and cell specific patterns. Over this decade, investigations have revealed a stream of more and more striking functions in the nervous system, including the recent demonstration that LINE1 transposons result in somatic diversity within the neurons of individual humans.

## COULD IT HAVE IMPLICATIONS FOR HUMAN HEALTH?

Even if one were to assume the worst-case scenario where most of the “dark matter” RNA is not functional at all, we now know that it is highly cell-type specific and this opens a wide area of additional diagnostic biomarkers based on these RNAs. Indeed, first reports showing that profiling “dark matter” RNAs offers superior diagnostic and prognostic information compared to protein-coding genes are now starting to appear, particularly in the cancer field which is leading others in development of these RNAs as biomarkers. However, if some or all, of the “dark matter” RNA is indeed functional then we can only imagine the plethora of secrets that are locked within that can shed light on basic mechanisms of development, homeostasis, and disease that still await exploration. For example, it is very curious that a (very) long

non-coding RNA could be highly restricted to a particular type of cancer and begs a question of why that would be the case if it had no functional role in the disease. Illuminating these mechanisms may help approach the contemporary challenge of understanding molecular and cellular biology from a new, holistic perspective, potentially revealing novel key aspects of cellular systems integration pathways. And, this is not necessarily limited to human cells, as other organisms, including those medically important to our health also likely possess their "dark matter" RNAs and its mysteries. However, one cannot over-emphasize the fact that we are the very beginning of the process of understanding what kind of RNAs are produced by a cell and what functionality they may

have. Still, the possibility that almost entirely unexplored treasure-trove of biological information is buried within our reach is too tantalizing to ignore.

*Received: 19 April 2012; accepted: 09 May 2012; published online: 29 May 2012.*

*Citation: Kapranov P and St. Laurent G (2012) Genomic "dark matter": implications for understanding human disease mechanisms, diagnostics, and cures. *Front. Gene.* 3:95. doi: 10.3389/fgene.2012.00095*

*This article was submitted to *Frontiers in Non-Coding RNA*, a specialty of *Frontiers in Genetics*.*

*Copyright © 2012 Kapranov and St. Laurent. This is an open-access article distributed under the terms of the Creative Commons Attribution Non Commercial License, which permits non-commercial use, distribution, and reproduction in other forums, provided the original authors and source are credited.*