Dark matter RNA: an intelligent scaffold for the dynamic regulation of the nuclear information landscape

Georges St. Laurent1,2,*, Yiannis A. Savva3 and Philipp Kapranov2**

1 Immunovirology – Biogenepse Group, University of Antioquia, Medellin, Colombia
2 St. Laurent Institute, Cambridge, MA, USA
3 Department of Molecular Biology, Cell Biology and Biochemistry, Brown University, Providence, RI, USA

Perhaps no other topic in contemporary genomics has inspired such diverse viewpoints as the 95+ % of the genome, previously known as “junk DNA,” that does not code for proteins. Here, we present a theory in which dark matter RNA plays a role in the generation of a landscape of spatial micro-domains coupled to the information signaling matrix of the nuclear landscape. Within and between these micro-domains, dark matter RNAs additionally function to tether RNA interacting proteins and complexes of many different types, and by doing so, allow for a higher performance of the various processes requiring them at ultra-fast rates. This improves signal to noise characteristics of RNA processing, trafficking, and epigenetic signaling, where competition and differential RNA binding among proteins drives the computational decisions inherent in regulatory events.

Keywords: dark matter RNA, non-coding RNA, vLinc RNA, RNA binding protein, biological signaling, RNA editing, RNA processing, molecular scaffold

INTRODUCTION

The emerging picture of the nucleus portrays a multifaceted environment where RNA processing events occur with accuracy, precision, and high resolution. Since diffusion cannot account for the speed and coordination of the molecular events occurring within its matrix, the nucleus must depend on precisely articulated macromolecular architectures and active transport mechanisms to achieve adequate throughput and signal to noise performance (Lancet et al., 2007; Misteli, 2007). In addition to the execution of baseline processing of RNA, the extensive network of RNA interaction machineries must respond to incoming physiological signaling, such as stress and cues from the physical environment (McKee and Silver, 2007; Sharma and Lou, 2011), by making rapid but precise changes at decision points, while at the same time maintaining robustness of the overall network. In effect, the entire nuclear space is a finely tuned RNA processing machine, designed to maintain accuracy in the dynamic and reversible regulation of myriad of transcriptome processing events simultaneously. Since the expansion of transcriptome processing increases the computational plasticity (Herbert and Rich, 1999) and the information processing capacity of biological networks (Marrick, 2007; St. Laurent and Wahlestedt, 2007), several authors argue that biological complexity itself has RNA complexity at its core (Licatalosi and Darnell, 2010).

Considering only current knowledge of these networks, and without extrapolating to yet undiscovered regulatory intricacies, their performance already gracefully exceeds that of systems biology models and mechanisms. Its diversity of specific functions, and the finely tuned regulation of those functions in response to physiological signals, suggests the existence undiscovered mechanisms and network design principles at work to maintain robustness of the RNA output of a cell. In fact, recent studies of disease mechanisms suggest that humans can tolerate a little loss of signal to noise performance in the nucleus. Healthy physiological function depends on the precision, reliability, and accuracy of the nuclear RNA processing machine, as processing errors in RNA molecules often lead to serious diseases (Garcia-Blanco et al., 2004; Cooper et al., 2009; Venable et al., 2009; Venables et al., 2009, Licatalosi and Darnell, 2010; Ward and Cooper, 2010; Jia et al., 2012).

It is in this context that we would like to consider the genomic “dark matter,” one of the major mysteries of the post-genome era. Perhaps no other topic in contemporary genomics has inspired such diverse viewpoints as the 95+ % of the genome, previously known as “junk DNA,” that does not code for proteins. Reports of pervasive transcription of these vast “dark matter” regions, combined with frequent identification of families of long or very non-coding RNAs (IncRNAs) originating from them, have opened new chapters of both discovery as well as controversy. The observation that the percentage of “dark matter” genomic sequence correlates monotonically with organismal complexity, for every species sequenced to date (Tuft et al., 2007), has inspired theories proposing a central role for these regions in the information processing of complex organisms (Marrick, 2007; St. Laurent and Wahlestedt, 2007). Yet, while an increasing number of specific interactions between IncRNAs and other biological molecules have demonstrated functions for a number of dark matter transcripts (Wang and Chang, 2010), a global concept of function has not yet emerged. In effect, the original reports of pervasive transcription (Kapranov et al., 2002, 2007b; Carninci et al., 2005, 2008; Katayama et al., 2005) of the mammalian genome have faded somewhat, with a focus instead on separately developed lists of IncRNAs detected in specific experiments or filtered by certain properties that hint at functionality (Willingham et al., 2005; Guttman et al., 2009; Khalil et al., 2009).
et al., 2009; Wai et al., 2010; Askarian-Amiri et al., 2011; Khaitan et al., 2011). These lists of lncRNAs usually only cover a few percent of the genome, representing only a small fraction of the original pervasive dark matter and typically, sample intergenic space as introns of known genes are usually assumed to represent pre-mRNAs. For example, recent work has shown the presence of numerous very long transcribed regions of intergenic genomic space not currently covered by the lincRNA annotations (Kapranov et al., 2010) and has shown that introns of mouse genes produce stable RNAs regulated separately from the mature protein-coding RNAs (St. Laurent et al., submitted). Partly due to this uncertainty, some authors have cast doubt on the importance of dark matter transcripts, labeling them transcriptional noise (Brosius, 2005; Struhl, 2007; van Bakel and Hughes, 2009; Robinson, 2010; van Bakel et al., 2010), or even arguing that they largely represent “fragments of known pre-mRNAs” (van Bakel et al., 2010). Even the existence of much of the dark matter RNA implied by the early reports of pervasive transcription has stirred recent controversy (van Bakel et al., 2010). On balance, a common view in the field holds that while there is a collection of lncRNAs with specific interactions and functions, they exist among a larger collection of dark matter transcriptional noise.

As the controversy surrounding the function of “dark matter” RNA continues, a number of recent studies provided more comprehensive datasets, through the implementation of improved methodologies to confirm its existence and, more importantly, to measure its relative mass. A recent investigation designed to capture and measure non-ribosomal signals, revealed surprisingly that dark matter RNAs actually comprise a majority of non-ribosomal non-mitochondrial RNAs in human cells (Kapranov et al., 2010). We also know that the nucleus is rich in dark matter RNA (Cheng et al., 2005). Since the majority of protein-coding RNAs reside in the cytoplasm, the fraction of dark matter RNA is likely to be many folds higher in the nucleus than that of protein-coding RNAs.

Considering the vital importance of maintaining the performance of nuclear processing of all types, the nuclear molecular machineries would not tolerate the accumulation of large amounts of non-functional RNA molecules. Any significant population of such molecules would at best represent a large input of noise into the fine-tuned computational machinery of nuclear processing, not likely to benefit the performance of the nucleus or the cell as a unit. In practical terms, if dark matter had no biological function, the high performance and signal to noise ratios of the nuclear RNA processing machineries would logically conflict with the high levels of dark matter RNA that are currently discovered in human cells (Kapranov et al., 2010). In other words, the currently emerging picture of the nucleus contains a paradox: a nuclear micro-environment simultaneously populated by high concentrations of precision RNA processing machineries, and by an astonishing level of noise from dark matter RNA. How can the nucleus precisely regulate such highly accurate processing events in tens of thousands of transcripts simultaneously, while ignoring the massive amount of inherent noise from dark matter RNA existing in the same nuclear space?

To resolve this apparent paradox, and to provide a mechanism for global function of dark matter RNA, in this article we present a theory in which dark matter RNA plays a role in the generation of a landscape of spatial micro-domains coupled to the information signaling matrix of the nuclear landscape. Within and between these micro-domains, dark matter RNAs additionally function to tether RNA interacting proteins and complexes of many different types, and by doing so, allow for a higher performance of the various processes requiring them at ultra-fast rates. This improves signal to noise characteristics of RNA processing, trafficking, and epigenetic signaling, where competition and differential RNA binding among proteins drives the computational decisions inherent in regulatory events.

**THE SYSTEM WIDE PERFORMANCE CHARACTERISTICS OF NUCLEAR RNA PROCESSING MACHINERIES**

It is estimated that an average human cell contains 300,000 mRNAs (Hastie and Bishop, 1976), each containing on average 10 exons, a start site, and a poly A+ tail. Thus, every such average molecule had to go through at least 18 splicing reactions (selection of splice donor and acceptor sites) plus selections of the start site and the polyadenylation site. In total, a minimum of 6M processing events had to occur to generate this diversity. This does not take into account (i) all subsequent base modifications such as RNA editing, N6-methyladenosine, 5′-cap, (ii) subsequent cleavage events, or (iii) transportation of these RNA molecules to their sites of function or into well demarcated nuclear storage for later use. Nor does it account for the polyA− RNA population that exceeds that of the polyA+ by several folds. Also, if one were to include the ribosomal RNA, that represents ~95% of all cellular RNA (Raz et al., 2011), which is also processed and modified, then the minimal order of the number of cellular processing events needed to accommodate the real complexity of RNAs within a single nucleus is likely to be in the tens of millions.

As a vital step in transcript processing, RNA editing offers further insight into the high level of orchestration of nuclear RNA processing machineries. Adenosine deaminase acting on RNA (ADAR) mediates adenosine to inosine (A-to-I) RNA editing in dRNA molecules, which often results in distinct downstream physiological outcomes for the edited RNAs. ADAR RNA editing frequently targets coding regions of mRNAs that encode ion channels and other components of the synaptic release machinery (Hopfengardner et al., 2003; Seeburg and Hartner, 2003). Intronic non-coding sequences with extensive complementarity to upstream or downstream exons containing the adenosine designated to be edited can form simple exon–intron hairpin structures (Higuchi et al., 1993; Burns et al., 1997; Hanrahan et al., 2000; Wang et al., 2000) or more complex RNA secondary structures such as a pseudoknot (Berrin, 2005). RNA editing in mRNAs often generates protein products that are not encoded by the literal genomic information, since upon translation the ribosomal machinery interprets inosines as guanosines (Basilio et al., 1962) resulting in amino acid substitutions. Various studies in different genetic model organisms suggest that RNA editing of mRNAs can result in profound changes in protein function (Rosenthal and Buxamilla, 2002; Bhalla et al., 2004; Ingleby et al., 2009).

Execution of this type of modification requires great deal of precision from the RNA processing machinery in terms of

---

*fgene-03-00057* — 2012/4/25 — 17:22 — page 2 — #2
Within minutes of amino acid deprivation or similar cellular events, the fate of such inosine-rich RNA molecules is different from the ones subject to “pinpoint” editing. They can in fact have at least two fates: retention within the nuclear compartment through editing sites generally demonstrate a high level of conservation in up to 50% A-to-I conversions (Nishikura et al., 1991; Polson and Bass, 2006), resulting in up to 50% A-to-I conversions (Nishikura et al., 1991; Polson and Bass, 2006). Choice of such substrates is also controlled as not every adenosine in molecules that are substrates for ADAR is edited at specific sites in specific transcripts, followed by 5′-capping, in response to little understood signals and circumstances (Affymetrix/CSHL ENCODE Project, 2009; Mercier et al., 2010). CTN RNA represents an intriguing example where both of these mechanisms are combined. Within minutes of amino acid deprivation or similar cellular stress, signals transduced into the nucleus result in cleavage of the sequestered CTN RNA, and the release and transport to the cytoplasmic translation machinery of the amino acid transporter for which the cleaved RNA product codes (Prasanth et al., 2005). In some cases, cleavage events themselves produce small RNAs, whose activities feedback into splicing decisions, as in the example of the HBII52 snoRNA, which is cleaved from intronic RNA templates in the SNURF-SNRPN locus, and interacts with the serotonin 2C mRNA to regulate its alternative splicing (Kishore and Stamm, 2006).

Considering all of these regulatory layers together, millions of RNA processing events have to happen with accuracy and precision to support the accuracy and reliability of high volume splicing output. As a result, achieving a correct splicing decision at a given site usually depends on a precise sequence of combinato-rial events, composed of multiple protein and RNA elements, and whose activities feedback into splicing decisions, as in the example of the HBII52 snoRNA, which is cleaved from intronic RNA templates in the SNURF-SNRPN locus, and interacts with the serotonin 2C mRNA to regulate its alternative splicing (Kishore and Stamm, 2006).

Performance related challenges would face any system designed to produce such a wide array of molecular outputs. Yet, even with these challenges, the systems performance of nuclear RNA processing appears to be surprisingly high. A recent investigation of cell-to-cell variability of alternative splicing determined that non-transformed cells maintained very low splicing isoform variability between individual cells, and concluded that mammalian cells minimize fluctuations in mRNA isoform ratios by tightly regulating the splicing machinery (Wisaks et al., 2011). Evidence increasingly supports precise and finely tuned regulation of transcription processing events as a rule in the nucleus. For example, a growing number of reports describe links between perturbations in splicing (Cooper et al., 2009; Ward and Cooper, 2010), or transcript localization (Faghihi et al., 2008) and diseases. This trend underlines the importance of high precision and accuracy in RNA based machineries under healthy physiological conditions.

Thus, using as yet little known organizational principles, nuclear RNA processing machineries not only produce processed and modified RNAs with high efficiency and accuracy, but implement a large scale integration of dynamic physiological signals, which then drive precise regulatory control and plasticity in response to a myriad of signaling events. From this perspective, the nuclear RNA processing machineries, and the dynamic structural environment surrounding them, must orchestrate their tasks with high accuracy and precision. They must recognize and
distinguish processing motifs in RNA structural signals with high sensitivity and yet reject sub-optimal motifs with a high selectivity. Furthermore, the catalytic processes involved in the processing steps must occur with little or no errors. Once formed, the products must enter downstream trafficking pathways, and RNA whose presence is no longer required must be rapidly degraded to avoid introducing noise into earlier steps in the processing pathways due to RNA-waste accumulation. Finally, the entire multilayer system must maintain sufficient plasticity to quickly respond to thousands of potential information signals from outside the nucleus to generate appropriate alterations in processing steps at determined loci in response to changing physiological signals.

While the mystery of how all of this occurs within the space of the nucleus remains unresolved, it depends on the choreography of combinatorial interactions between transcripts and hundreds of RNA binding proteins (RBPs). RBPs interact with complex combinations of primary and secondary structure signals in RNA, and function in both cooperative and competitive types of architectures, as documented in splicing regulation (Darnell, 2006; Sharma and Black, 2006; Ule and Darnell, 2007; Liao et al., 2008; Hallegger et al., 2010), and more recently in chromatin signaling (Tsi et al., 2010; Zhao et al., 2010). The first RBPs to interact with a given nascent transcript can influence the subsequent folding steps of that RNA, and thereby change the downstream distribution of protein interactions for that RNA. The process of differential recognition of nascent RNA information signals by the correct RBPs must occur at a pace complementary to that of transcription as well as subsequent processing or chromatin signaling. To maintain plasticity for the accurate transduction of environmental signals, the RNA–protein interaction landscape must somehow achieve an extraordinary coupling between computational, catalytic, and structural elements.

**EMERGING FEATURES UNDERLYING THE HIGH PERFORMANCE OF NUCLEAR RNA PROCESSING MACHINERIES**

As investigations continue to reveal the depth and performance of nuclear RNA processing functions, the challenge for systems biologists grows more daunting. Current systems biology modeling cannot account for the precision, accuracy, or signal to noise ratios achieved by RNA processing machineries. Nevertheless, the transcriptome–proteome interface in the nucleus contains a vast store of dynamical information. To consistently make use of this information, the nuclear systems network architecture must have a number of key design features, including maintenance of reversibility, temporal coherence (the timing and velocity of information processing between network layers), and the ability to resolve logical conflicts over the spatial extent of the networks that comprise the system. To help explain how nuclear RNA processing networks harness the power of that information, a number of concepts have emerged.

**DYNAMIC SCAFFOLDING MAXIMIZES INFORMATION FLOW**

Biological molecules in the nuclear space exist in a constrained environment where diffusion occurs relatively slowly (Albert et al., 2012). Thus, biochemical kinetics can represent a hurdle for adequate performance of complex multistep processing pathways, as the entire interdependent system must minimize bottlenecks and flow imbalances. In order to overcome these physical limitations, RNA processing machineries must rely on highly articulated spatial domains, where local environments transduce information efficiently. The concept of global scaffolding can create these performance-enhancing interaction topologies. For example, an RNA scaffold can increase the local concentration of an RBP, such as Nova 1, and a corresponding increase in the signal to noise performance of Nova’s influence on splice site selection within that particular spatial domain (Figure 1A). Increasing the local concentration of these factors also permits improved interaction kinetics with less ΔG, making the interactions more reversible (Figure 1B and also below).

Structural features in the nucleus that enhance transcriptional control and RNA processing contain a large amount of spatially and temporally coded information content. The nucleus for example appears to depend on RNA secondary structure signals for its effective formation, as the absence of these RNA secondary structures resulted in complete disarray of the nucleolus (Peng and Karpen, 2007). The reversibility of such events in nuclear architecture means that elements responsible for their formation also encode sufficient information to detect external signals and respond with disassembly and transport of components to other spatial domains or downstream processing pathways (Spector and Lamond, 2011).

**COMPETITION AND COMPUTATION AT THE TRANSCRIPTOME–PROTEOME INTERFACE**

With their unique combination of primary, secondary, and tertiary structure, RNA offers a multiplicity of ways to code for biological information. At the core of this system is a language of RNA–protein, and RNA– RNA/DNA recognition implemented by RNA’s unique ability to couple analog and digital signals (St. Laurent and Wallestedt, 2007). The efficient transduction of that information often depends on its timely recognition by the appropriate RBPs present in the immediate vicinity of an elongating primary transcript. As the transcript emerges from Pol II, it begins to fold. That folding is also influenced by the RBPs that are supposed to interact with it. They influence which of many folding paths the RNA can take. If the correct RBPs are not right there to quickly associate with the RNA, then the RNA could take another folding path, which would in turn lead to a different set of downstream events, as in the case of Nova splicing proteins and their influence on upstream or downstream splice site choice. So the presence or absence of a given distribution of RBPs in the vicinity of a nascent RNA chain will influence a series of “memory states” that then modulate other processing events downstream. With such a large space of potential RNA–protein interactions, and the requirement for dynamic reversibility of many of their associated signaling events, the system faces a major challenge to achieve an adequate signal to noise ratio for effective function.

Active competition for recognition site on nascent RNA signals directly addresses these problems. Splicing regulation makes abundant use of competing RBPs to enhance the sensitivity, specificity,
and regulatory control of splice site decision commitment (Ule and Darnell, 2006; Chen and Manley, 2009). Examples include PTB protein which antagonizes Nova (Polydorides et al., 2000) at overlapping recognition sites, establishing a sensitive switch between two splicing choices. Interesting examples from splicesome quality control also demonstrate the importance of reversibility, such as involvement of ATPase Prp19p in both forward and discard splicing pathways (Koedatdhangal et al., 2010).

Competition may also drive accurate computation in the small RNA regulatory pathways, with duplex regions competing for recognition by ADEK vs Drosha/Dicer, with contrasting outcomes depending on which protein prevails (Nishioka, 2006). Similarly, many epigenetic signaling events may be mediated by competitive interactions between IncRNAs and protein components of signal-machineries (Lee, 2011). All of these regulatory mechanisms require effective concentrations of interacting proteins to achieve adequate signal to noise ratios. The Lin28–let7 miRNA interaction provides an interesting example of specificity that would be difficult to achieve with low protein concentrations (Nam et al., 2011; Piskounova et al., 2011).

REVERSIBILITY AND FEEDBACK LOOPS

Erasure of information presents a challenge for any complex system (Lloyd, 2001). In biological systems, thermodynamic constraints make the cost of information innately high, and yet its value can oscillate from vital to worthless or even harmful in seconds once the message or a signal encoded in it is transduced. The dynamics of this “volatile market” reality make erasure of biological information a high priority in any system, but especially in the nucleus where many network pathways converge. While DNA retains the permanent information, a large majority of the dynamical information exists within the transcriptome, as combinatorial accumulations of RNA–protein and RNA–RNA/DNA interactions.

Not surprisingly, reversibility is a key feature of information coding at the transcriptome–proteome interface. The conformational flexibility of RNAs, especially ncRNAs whose secondary structures are not constrained by coding regions, and the dynamic changes in their structure that can occur in response to protein binding and environmental signals provide not only increased symbolic information density, but contribute to the reversibility of RNA–protein interactions. Proteins that bind RNA also tend to contain natively unstructured regions. This could be the basis for structural articulation (i.e., the incorporation of information containing motifs and elements into nuclear scaffold structures) that improves precise temporal and spatial choreography of RNA processing machineries. For example, interactions between RNA and their cognate proteins often involve natively unstructured regions in the protein, and similarly flexible structures in the RNA (Leslie and Varani, 2001). These regions of evolutionarily coded local disorder contribute useful properties for information processing. Precisely orienting them within an articulated regional structure increases their sensitivity, specificity, and reversibility, thereby contributing directly to the throughput and precision of nuclear machineries. When these regions form a stepwise interaction with their RNA target, the entropy of the complex is decreased, thereby producing an “entropic spring” effect, which enhances reversibility when the interaction is no longer required (Tompa and Comb, 2004). Together with reversibility of individual interactions within RNA–protein interaction networks, frequent feedback loops support the reversibility of these networks. These features operate cooperatively to facilitate the timely erasure of information, and the finely tuned response of RNA processing machineries to changes in signaling and various environmental conditions.

Thus, a central part of our argument maintains that the performance and throughput of nuclear RNA processing machineries...
requires the functional coupling of well-articulated spatial and temporal landscapes in order to maximize the flow of biological information through the components of RNA processing networks.

**THE DARK MATTER INTELLIGENT SCAFFOLD IMPLICATIONS OF THE PREPONDERANCE OF CELLULAR DARK MATTER RNA IN MAMMALIAN CELLS**

Several years ago, John Mattick presented the concept of the nucleus as an "RNA machine" (Amaral et al., 2008), arguing that much of the information processing in the nucleus occurs through RNA intermediates, and that ncRNA overcomes the prohibitive regulatory overhead associated with saturated protein-protein regulatory interactions in this environment (Gagen and Mattick, 2005). From the point of view of information theory, this implies that the RNA content of the nucleus functions in a manner roughly equivalent to an information channel, and that the "channel capacity" (information throughput) of this system depends on the available degrees of freedom of the combined population of RNA molecules contained in the system. Consequently, RNA quality control and degradation machinery must actively pursue the elimination of non-functional RNAs that would represent noise to the "RNA machine." Yet, with the recent discovery that dark matter RNA makes up the majority of cellular RNA by mass (Kapranov et al., 2010; and an even greater majority in the nucleus), it appears that this enigmatic class of RNA does not represent noise in the RNA based information channel of the nucleus, and instead likely comprises an integral part of the information channel itself.

The information containing structural features of dark matter RNAs and their ability to interact with the nuclear proteome appear similar to their coding counterparts. If the primary sequence patterns and secondary structure motifs that determine protein interactions occur with similar densities in both classes of RNAs, then they must both exist in the nucleus in complex with proteins. If dark matter RNA represented noise or spurious transcription, their predominant mass would compromise signal to noise ratios of the target transcript, measured as recently demonstrated for XIST and transcription factor YY1 (Jeon and Lee, 2011). Alternatively, proteins or RNAs can use co-transcriptional targeting where the transcript is tethered during transcription by RNA polymerase, similar to the mechanism of the TAR RNA targeting by HIV TAT protein (Brady and Kashanchi, 2003). Transcriptional targeting may also occur with the short ncRNAs transcribed from the 5' ends of many human genes (Wei et al., 1998; Kapranov et al., 2007a; Kanhere and Jenner, 2012). Dark matter RNAs have all three of these mechanisms available to mediate their interactions with DNA and other RNAs, providing a large combinatorial basis for the formation of flexible complexes that drive spatial and computational integration.

As these RNA accumulate into spatial micro-domains surrounding one or more genomic loci, they establish a region of nuclear space under their influence, which in turn attracts a variety of molecules. The RNAs can interact with many proteins, and other large and small RNAs, often with relatively low affinities, which results in a temporally and spatially distributed macromolecular landscape around that locus (see Figure 2). Since these molecules function primarily in transcription and epigenetic regulation, the dark matter guided landscapes would facilitate the structural and computational operations of both systems, as well as catalyze crosstalk between them. In this manner dark matter RNAs can effectively establish finely tuned concentration gradients of epigenetic signaling and RNA processing proteins (and small RNAs) for efficient operation of these systems. An intriguing example of this has recently been described as a "molecular cage" for PRC1 complexes. The "molecular cage" apparently uses a combination of methylated H3K27 moieties and low affinity binding sites on nascent lncRNAs to increase the local concentration of PRC1 for chromatin signaling (Beisel and Para, 2011).

The intelligent scaffold mechanism facilitates the accumulation of higher concentrations of RBPs (and small RNAs) within chromatin regions, as well as the micro-partitioning of these regions at an optional resolution for RNA processing, epigenetic signaling, and transcript expression regulation. Macromolecules within these micro-domains can dissociate from their low affinity binding sites in these dark matter rich micro-regions, as they find higher affinity sites in nascent strands emerging from RNA Pol II transcription. Abundant sites of alternative localization in dark matter equates with more effective differential recognition of RNA motifs by competing RBPs, and increased reversibility of signal transduction in regulatory events. The key here is that signal to noise ratio is not driven only by the size of ∆G, but by the ratio of ∆G "protein A" to ∆G "protein B" or the ratio of ∆G site1 of protein A on the "target" nascent strand RNA molecule to ∆G site2 of the same protein A on the "repository" dark matter RNA molecule. This is shown as "Biological Information Content" on Figure 2. If both ∆Gs are large compared to their difference, then the signal is low and the noise is high. A recent experiment that used RNAi knockdown to reduce the expression levels of splicing regulator SRSF1 confirmed the importance of high concentrations of RNA processing proteins to maintain adequate signal to noise ratios. Lowered concentrations of SRSF1 markedly increased the variance of splicing isoform ratios of the target transcript, measured in populations of single cells (Waks et al., 2011).
FIGURE 2 | The information content of a hypothetical dark matter RNA.

Combinations of primary sequence and secondary structure form high affinity interaction sites with high information content (left side). These specific interactions have a large ΔG. At the other end of the spectrum, the same RNA can have a large number of relatively non-specific interaction sites that nevertheless have biological information content and functional significance. Their absence would result in subtle loss of signal to noise characteristics across many affected pathways that could be unrelated to the RNA.

Adjacent micro-partitions could favor higher concentrations of some proteins over others, due to the heterogeneous distributions of low affinity binding sites along the lengths of dark matter RNA molecules in each micro-partition. The result, depicted in Figure 3, shows varying levels of sequestration of RNA processing components, depending on the systems performance requirements of each component. Overall, higher concentrations of effector components equate with faster kinetics and more finely tunable regulation, which in turn improve signal to noise ratios and system performance.

The temporal and spatial dynamics of intelligent scaffolds permit integration of signals from many levels of biological information processing. Changes in the intelligent scaffolding environment of a three-dimensional chromatin micro-region can impact the dynamics of transcriptional folding, processing, localization, and degradation of transcripts as well as chromatin signaling (see Figure 4). For example, dark matter cleavage events can quickly change the structure of the micro-domain by sweeping away large numbers of proteins, RNAs, and scaffold, and at the same time generate small RNAs, or expose regions of RNA complementarity to small RNAs, as described in the recent theory of competing endogenous RNAs (ceRNAs) by the Pandolfi group (Salmena et al., 2012). Cleavage of very long dark matter RNAs, for example those coming from the vlinc regions (Kapranov et al., 2010), could occur even with their RBPs still attached. Cleaved RNAs could then function as IncRNAs. Small RNAs could also interact with sites in tethered vlincs, thereby acting as a sink, or by blocking sites that would otherwise be occupied by other signaling molecules. Under some circumstances, combinations of these other events could serve as the signal to trigger cleavage of the vlincs, which could then form a rapid feed-forward circuit as the cascade of cleavage continues in the entire micro-domain.

CONCLUSION: THE FOREST ENRICHES THE FUNCTIONALITY OF THE TREES

While specific interactions drive the bulk of molecular information processing in biological systems, in the RNA based regulatory networks of the nucleus the performance characteristics of specific interactions are determined by the surrounding micro-environment. The dark matter RNA plays a key role in implementing the dynamic responsiveness of that surrounding micro-environment. Considering its importance, the concept of functions for dark matter RNAs should embrace a continuum, from those that arise from highly specific interactions, to those at the other end of the spectrum that involve lower affinity and less specificity, but nevertheless contribute to the synergistic attributes of the surrounding micro-environment. Those attributes permit the specific interactions, and facilitate their coordination and integration.

Evaluating dark matter RNAs in this fashion provides a context and explanation for the relatively low level of conservation of these RNAs, as many informational elements either do not require conservation, or require only functional conservation. As demonstrated for a growing list of ncRNAs, functionality does not require
FIGURE 3 | Distribution of different domains with varying ΔG affinities (Y-axis) for different RBPs (balls of different colors) along the length of a hypothetical series of micro-domains each composed of different (very) long non-coding RNAs (X-axis). The heterogeneous landscape of dark matter RNAs creates pockets of enrichment of these different RBPs.

FIGURE 4 | A nascent RNA strand being synthesized by RNA Pol II interacts with RBPs (balls of various colors), small RNAs and lincRNAs. Interaction between all these molecules is made possible by close proximity of these molecules in the nuclear micro-domains. RBPs and small RNAs are bound to (linc)RNAs with relatively low specificity and the latter present them in exact architectural and temporal environment to the nascent strand that possesses specific motifs for the former. Nuclease action cleaves the non-coding RNA template and thus changes the structure of the scaffold complex. This in turn can change the kinetics of the interaction between the RBPs and the small RNAs bound to it and the sequence of their presentation and interaction with the nascent strand of RNA.

conservation, at least not in the same way that is known to occur for protein-coding sequences (Pang et al., 2006). The theory predicts increasing concentrations of dark matter complexed with RNA interacting proteins in complex organisms, and helps explain the direct correlation of organismal complexity with the genomic percentage of non-coding regions in all genomes sequenced to date (Tait et al., 2007). It also suggests expansion of regions of RNA interacting regions in proteomes of organisms as evolutionary complexity increases.

The dark matter intelligent scaffold concept focuses on the level of coupling between computation and spatial articulation. The theory holds that large increases in biological complexity required ever increasing levels of coupling between computation and structure, as a key driver of that complexity, and ultimately a measure of organismal fitness. Dark matter RNA was recruited to perform this function, to dynamically bridge these two ostensibly orthogonal dimensions, because its flexible structural and computation features endow it with special qualities to serve as a molecular intermediate in the coding, processing, and distribution of information.

ACKNOWLEDGMENTS
We wish to thank Erica Dumais for expert artistic assistance with figure preparation and Juan Pablo Lizarazo for logistical support.
Amaral, P. P., Dinger, M. E., Mer- | Albert, B., Leger-Silvestre, I., Nor- | REFERENCES

Albert, R., Laster-Silvestre, I., Nem- | mand, C., and Gadal, O. (2012). Nuclear | referenced in this work. Nature 457, 280–293.

April 2012 | Volume 3 | Article 57 | www.frontiersin.org

Bailey, T. L., Bansal, M., Baxter, L., Davis, M. J., Wilming, L. G., Aidinis, K., Bajic, V. B., Brenner, S. E., Bat- | rhodesian RNA editing. Trends Genet. 21, 178–188.

Beisel, K. W., Berento, T., Bons, H., | Chalk, A. M., Christoffels, A., Clutton- | graphite: a two-tier strategy for pre-mRNA splicing quality control. RNA 17, 383–395.

Beisel, K. W., Berento, T., Bons, H., | Chalk, A. M., Christoffels, A., Clutton- | graphite: a two-tier strategy for pre-mRNA splicing quality control. RNA 17, 383–395.

Beisel, K. W., Berento, T., Bons, H., | Chalk, A. M., Christoffels, A., Clutton- | graphite: a two-tier strategy for pre-mRNA splicing quality control. RNA 17, 383–395.

Beisel, K. W., Berento, T., Bons, H., | Chalk, A. M., Christoffels, A., Clutton- | graphite: a two-tier strategy for pre-mRNA splicing quality control. RNA 17, 383–395.

Beisel, K. W., Berento, T., Bons, H., | Chalk, A. M., Christoffels, A., Clutton- | graphite: a two-tier strategy for pre-mRNA splicing quality control. RNA 17, 383–395.

Beisel, K. W., Berento, T., Bons, H., | Chalk, A. M., Christoffels, A., Clutton- | graphite: a two-tier strategy for pre-mRNA splicing quality control. RNA 17, 383–395.

Beisel, K. W., Berento, T., Bons, H., | Chalk, A. M., Christoffels, A., Clutton- | graphite: a two-tier strategy for pre-mRNA splicing quality control. RNA 17, 383–395.

Beisel, K. W., Berento, T., Bons, H., | Chalk, A. M., Christoffels, A., Clutton- | graphite: a two-tier strategy for pre-mRNA splicing quality control. RNA 17, 383–395.

Beisel, K. W., Berento, T., Bons, H., | Chalk, A. M., Christoffels, A., Clutton- | graphite: a two-tier strategy for pre-mRNA splicing quality control. RNA 17, 383–395.

Beisel, K. W., Berento, T., Bons, H., | Chalk, A. M., Christoffels, A., Clutton- | graphite: a two-tier strategy for pre-mRNA splicing quality control. RNA 17, 383–395.

Beisel, K. W., Berento, T., Bons, H., | Chalk, A. M., Christoffels, A., Clutton- | graphite: a two-tier strategy for pre-mRNA splicing quality control. RNA 17, 383–395.

Beisel, K. W., Berento, T., Bons, H., | Chalk, A. M., Christoffels, A., Clutton- | graphite: a two-tier strategy for pre-mRNA splicing quality control. RNA 17, 383–395.

Beisel, K. W., Berento, T., Bons, H., | Chalk, A. M., Christoffels, A., Clutton- | graphite: a two-tier strategy for pre-mRNA splicing quality control. RNA 17, 383–395.
St. Laurent et al. Dark matter RNA
J. D., Kapranov, P., and Triche, T. J. (2010). Large intergenic noncoding RNAs associated with Foxn1 tamr gene family of tumours. Proc. Am. Assoc. Cancer Res. Abstract #487.

Waks, Z., Klein, A. M., and Shih, P. A. (2011). Cell-to-cell variability of alternative RNA splicing. Mol. Syst. Biol. 7, 206.

Wang, E. T., Sandberg, R., Lau, S., Khubeetamu, J., Zhang, L., Mert, C., Kingdom, S. F., Schroll, G. P., and Biegel, C. R. (2008). Alternative isoform regulation in human tissue transcriptionomes. Nature 455, 470-476.

Wang, E. C., and Chang, H. Y. (2011). Molecular mechanisms of long noncoding RNAs. Mol. Cell 43, 904-914.

Wang, Q., O’Donovan, P. L., Chen, C. X., Xiu, D. S., Murray, J. M., and Nishikura, K. (2000). Altered G protein-coupling functions of RNA editing isoform and splicing variant serotonin2C receptors. J. Neurochem. 74, 1290-1300.

Wang, Q., Zhang, Z., Blackwell, K., and Carmichael, G. G. (2005). Vignettes bind to promiscuously A-to-I edited RNAs and are involved in the formation of heterochromatin. Curr. Biol. 15, 586-591.

Ward, A. J., and Cooper, T. A. (2010). The pathology of splicing. J. Pathol. 220, 152-163.

Wei, F., Garber, M. E., Fung, S. M., Fischel, W. H., and Jones, K. A. (1998). A novel CCR5-associated C type cyclin interacts directly with HIV-1 Tat and mediates its high affinity, loop-specific binding to TAR RNA. Cell 94, 453-462.

Wessjohann, A. T., Dreh, A. P., Batzler, S., Peters, E. C., Wro, B. G., Ara-Blanc, P., Hugenholtz, J. B., and Schulte, P. G. (2005). A strategy for probing the function of noncoding RNAs finds a repressor of Nfat. Science 309, 1570-1573.

Witten, J. T., and Ule, J. (2011). Understanding splicing regulation through RNA splicing maps. Nature 477, 89-97.

Witcosky, W. M., Flanagan, J. J., Malicki, S. Z., Simpson, S. L., and Clemens, J. C. (2004). Alternative splicing of Drosophila Dscam generates axon guidance receptors that exhibit isoform-specific homophilic binding. Cell 118, 619-633.

Xiao, X., and Lee, J. H. (2010). Systems analysis of alternative splicing and its regulation. Wiley Interdiscip. Rev. Syst. Biol. Med. 2, 506-515.

Yang, J. H., Luo, X., Nie, Y., Su, Y., Zhao, Q., Klein, A. M., and Rich, R. A. (2004). Widespread inosine-containing mRNA in lymphocytes regulated by ADAR1 in response to inflammation. J. Immunol. 179, 15-23.

Yang, J. H., Nie, Y., Zhao, Q., Su, Y., Pypart, M., Su, H., and Rich, R. A. (2005). Intermolecular localization of differentially regulated RNA-specific adenosine deaminase isoforms in inflammation. J. Biol. Chem. 279, 45833-45842.

Zhao, J., Ohsumi, T. K., Kung, J. T., Ogawa, Y., Grau, D. J., Sarma, K., Song, J. J., Kingston, R. E., Borowsky, M., and Lee, J. T. (2010). Genome-wide identification of polycomb-associated RNAs by RIP-seq. Mol. Cell 40, 939-953.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 19 February 2012; accepted: 28 March 2012; published online: 26 April 2012.

St. Laurent, G.; Savva, Ya and Kapranov, P. (2012) Dark matter RNA: an intelligent scaffold for the dynamic regulation of the nuclear information landscape. Front. Gene. 3:57. doi: 10.3389/fgene.2012.00057

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

Copyright © 2012 St. Laurent, Savva and Kapranov. 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.