A STORY OF GROWING CONFUSION:
GENES AND THEIR REGULATION

S. J. PROHASKA

*Biomedical Informatics, Arizona State University, PO-Box 878809, Tempe, AZ
85287, USA, and
Center for Evolutionary Functional Genomics, The Biodesign Institute, Arizona
State University, PO-Box 875301, AZ 85287, USA
E-mail: sonja.prohaska@asu.edu*

P. F. STADLER

*Bioinformatics Group, Department of Computer Science, and Interdisciplinary
Center of Bioinformatics, University of Leipzig
Härtelstrasse 16-18, D-04109 Leipzig, Germany, and
RNomics Group, Fraunhofer Institut für Zelltherapie und Immunologie — IZI
Deutscher Platz 5e, D-04103 Leipzig, Germany, and
Institute for Theoretical Chemistry, University of Vienna, Währingerstrasse 17,
A-1090 Wien, Austria, and
Santa Fe Institute, 1399 Hyde Park Rd., Santa Fe, NM 87501, USA
E-mail: studla@bioinf.uni-leipzig.de*

High-throughput experiments have produced convicing evidence for an extensive
contribution of diverse classes of RNAs in the expression of genetic information.
Instead of a simple arrangement of mostly protein-coding genes, the human tran-
scriptome features a complex arrangement of overlapping transcripts, many of
which do not code for proteins at all, while others "sample" exons from several
different "genes". The complexity of the transcriptome and the prevalence of non-
coding transcripts forces us to reconsider both the concept of the "gene" itself and
our understanding of the mechanisms that regulate "gene expression".

1. Introduction

The prevailing picture of genome organization is a rather simple linear ar-
rangement of separated individual genes which are predominantly protein-
coding. Albeit a few exceptional cases of overlapping or interleaving genes
have long been known, this paradigm has been dominating research in
molecular genetics so far. The ENCODE Pilot Project [86], the mouse cDNA
project FANTOM [60,45], and a series of other large scale transcriptome stud-

2

ies [34,65] have amassed a robust body of data and profoundly change the picture in a way that might be difficult to accept: The "obscure exceptions" are in fact the rule! The mammalian transcriptome is characterized by an complex mosaic of overlapping, bi-directional transcripts and a plethora of non-protein coding transcripts arising from the same locus.

This newly discovered complexity is not unique to mammals. Similar high-throughput studies in invertebrate animals [47,28] demonstrate the generality of the mammalian genome organization among higher eukaryotes. Even the yeast *S. cerevisiae*, whose genome has been treated as conceptionally well understood, catches up with the more complex picture and surprises us with a much richer repertoire of transcripts than previously thought [27,14,57].

Both experimental and computational evidence suggest that many of the newly discovered transcripts and processing products are functional, although at present we have very little knowledge of the potentially different types of molecular mechanisms and function integrating the individual transcripts. However, both theoretical considerations and a set of examples [98] hint at a role in a variety of different regulation mechanisms.

The unexpected diversity of the transcriptome and the unexpected complexity of genomic organization also forces us to re-think our understanding of genes and their regulation. We not only have most likely overlooked many of the players by focusing on three dominating and well-separated layers of regulation (transcriptional regulation, signaling networks, and microRNA-based gene silencing), but we also grossly underestimate the complexity of the regulatory circuitry by assuming a hierarchical structure of well-separated layers. Indeed, several systems demonstrate an intimate interplay of distinct regulatory mechanisms. The interleaved usage of protein- and ncRNA-based mechanisms in apoptosis control [11,29] shows that transcriptional networks alone are not sufficient to understand the regulatory processes despite their great success in the area of early development [3,50].

This contribution is organized as follows. In the next section we briefly argue that most of the observed transcription has biological function, and more strictly, at least a large fraction of the non-coding transcripts themselves are functional. We then consider the structure of mammalian transcriptome, concluding that the very notion of the "gene" itself becomes problematic, begging the question what is regulated in "gene regulation". We then refine our claim about transcript function, arguing that many of the "novel" transcripts have functions as RNAs, and that these functions are most likely regulatory. Given that gene regulation occurs in space and

time to pattern a three-dimensional organism, we will briefly discuss how spatial regulation is achieved and integrated into the whole picture.

2. Function of Noise?

The dramatic expansion of transcript repertoire has been argued to be at least largely functional [52,55,95,7]. However, noise is unavoidable in any physical system. Therefore, the novel non-coding transcripts could be interpreted as *"transcriptional noise"* resulting from accidental initiation of transcription or the accumulation of malformed RNAs as a by-product of transcription of the truly functional mRNAs. This explanation is highly unlikely for a variety of reasons.

The amount of observed non-coding transcripts, at least half of the transcriptome, would imply a very unfavorable signal to noise ratio. The effort for the cell to sort the wheat from the chaff on transcript level should be a sufficient basis for a selection to reduce over time. The structure and organization of the majority of the non-coding transcripts is also striking: they are well-defined and do not seem to differ from the previously known ones in terms of their processing (e.g. capping, splicing, and polyadenylation) [86] or in terms of regulated expression [66]. Their transcription start sites, as identified by CAGE/ditags, are very similar to known start sites judged by the presence of proximal regulatory elements, as identified via ChIP-on-Chip experiments [86]. Consequently, many of the novel transcripts show differential expression [8,66] in differentiation and disease [51]. Also phylogenetic considerations support their retention: in many cases there is at least syntenic conservation (i.e., transcripts arising from corresponding positions relative to homologous neighboring features, even though their sequences might not be well-conserved) [35,70,46]); in many cases there is also extensive sequence conservation. In fact, many of the ultraconserved sequence elements in mammals give rise to ncRNAs of hitherto unknown function [6].

Given a set of transcripts that show a well-defined genomic location and pattern of expression but a lack of functional annotation, one might be tempted to handle the "act of transcription" as the transcripts' function while the RNA molecules themselves are meaningless. It is well known that *transcriptional interference* [79,53] may play an important role when transcripts from the same locus compete for transcription. It has been demonstrated, for instance, that perturbation of an anti-sense RNA can alter the expression of the sense messenger RNAs [36]. Furthermore, it is yet undecided whether chromatin has to be in open conformation to al-

4

low for transcription or whether the act of transcription opens the closed chromatin conformation [23]. As in the case of transcriptional interference, the regulation of chromatin structure may explain non-coding transcription without a function for the transcript itself. There are, however, several lines of evidence suggesting strongly that the transcripts are not irrelevant. In addition to transcriptional interference, a recent review[39] discusses four distinct mechanisms for anti-sense function: RNA masking (in which the anti-sense transcripts regulates alternative splicing), RNAi (in which RNA double strand formation triggers RNA degradation), RNA editing of double stranded regions, and the involvement in chromatin remodeling.

Most of the transcripts show evidence of post-processing. Typical non-coding RNAs appear to be spliced, polyadenlyated, and exported just like regular mRNAs. Furthermore, small RNAs $\leq$ 200nt in length [35] have been shown to be processed from longer precursors. It is unclear why the cell would carry the burden of processing when it could make use of several efficient degradation mechanisms [32] to get ride of non-functional RNAs. While this is certainly the fate of *some* "cryptic" pol-II transcripts [97], it does not seem to apply to the many stable mRNA-like ncRNAs.

At least a fraction of the long non-coding primary transcripts are precursors of small functional ncRNAs. Starting with microRNAs in the year 2000, several newly discovered classes of small non-coding RNAs, among them piRNAs [40,1], 21-U RNAs [72], and the promoter and terminator associated paRNAs and taRNAs [35] add to the diversity of ncRNAs. For each of these classes there is at least circumstantial evidence that they are functional. Classical genetic studies have stumbled across only a small number of small ncRNAs, among them the microRNAs *lin-4*, *let-7*, or *iab-4*, for several reasons: they are hard to find in mutagenesis experiments because of their size, many functions (e.g. in antiviral defense or specific stress response) do not produce a discernible phenotype under most lab conditions, and ncRNAs are often redundant because they appear as multicopy genes.

A sizable fraction of both large and small ncRNAs shows weak but significant sequence conservation. Bioinformatics-based surveys of genomic DNA show that there is a large number of loci which show unexpected conservation of RNA secondary structure [91,92,63,89,93]. Stabilizing selection of RNA structure strongly suggests specific function(s) at the RNA level.

Summarizing these argument, we conclude that the view of predominantly non-functional transcripts is not a parsimonious explanation of the available data. Of course, we cannot (yet) ascribe a *specific* function to the majority of transcripts, a fact that does not imply that they are not

functional. Conversely, we do not claim that all transcripts or transcription is biologically functional — we only argue that non-coding transcripts are functional often enough that we cannot treat them as odd special cases.

3. What Are Genes Anyway?

The data briefly reviewed in the previous section draw a picture of the mammalian transcriptome that forces us to rethink the notion of well-separated individual genes. The human genome no longer presents itself as linear arrangement of mostly protein coding genes (together with a few exceptional RNAs), littered by more the 90% of useless junk DNA that make up huge "gene deserts" that nicely separate functional regions. Instead, at least the non-repetitive fraction of the genome is transcribed almost in its entirety; regions with high density of protein coding genes form complex clusters of multiple overlapping transcripts in both sense and anti-sense direction, Fig. 1. These primary transcripts are then processed into a plethora of different mature products of diverse functions and localizations. Only a small fraction of the transcripts gives rise to "classical" protein-coding mRNAs.

With overlapping transcripts emerging as the rule rather than the exception, and with functional non-protein-coding RNAs arising as a major class of transcripts we are forced to rethink the concept of the *gene* itself.
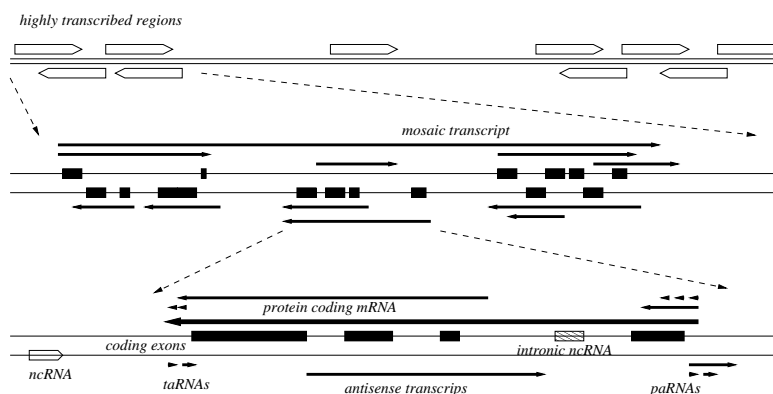

Figure 1.   Sketch of the post-ENCODE view of a mammalian transcriptome (adapted from [35]). Highly transcribed regions consist of a complex mosaic overlapping transcripts (arrows) in both reading-directions. These transcripts link together the locations of several protein coding genes (coding exon indicated by black rectangles). Conversely, multiple transcription products, many of which are non-coding, are processed from the same locus as a protein coding mRNA.

6

This implication has of course been realized in the field [22,24], but so far no satisfactory solution has arisen that really encompasses the non-protein-coding transcription products.

Conceptually, one would like to use the term *gene* for a unit in both functional and structural terms. We argue that function is primarily associated with the transcripts and their processing products rather than with the DNA locus, hence it becomes futile to assign a single function or a single functional product to a DNA locus. In human, for instance, both the non-coding co-activator SRA, whose function has been well studied experimentally, and a protein SRAP, which is conserved at least throughout chordates [42], are produced as overlapping sense-transcripts from the same locus. A gene concept advocated by Gerstein *et al.* [22] is based on the assumption that the "important" function is encoded in the protein-coding exon(s), if present at all. Abandoning this "proteinocentric" point of view, we argue that functional units on the level of the transcriptome or proteome and structural units, most naturally intervals on the DNA, are typically in a many-to-many relation. This implies, for instance, that one and the same mutation may have several distinct functional consequences. A recent example of this type [31] thus might as well be a common phenomenon.

Another open question is whether "nearby" regulatory elements (such as promoters, cis-regulatory regions, or elements responsible for positional effects) should be considered part of a "gene". While appealing from a functional point of view (since a gene product's spatial and temporal expression pattern plays a role in its biological function), such a definition makes it nearly impossible in practise to annotate a "complete gene" on the genomic DNA. Furthermore, thinking of a "gene" as single DNA interval might not satisfy our needs to annotate the template of a functional unit together with its dispersed regulatory context.

4. Regulation by RNA

Although the overwhelming majority of both the experimentally discovered non-protein-coding transcripts and the computationally predicted structured RNA elements have resisted functional annotation attempts, it appears safe to assume that a large fraction of them has some *regulatory* function:

(1) Catalytic functions are expected to be rare. While some of the "classical" ncRNAs have catalytic functions (the ribosome after all is essentially a ribozyme, and snRNAs are actively involved in splicing), such an active

role appears to be very rare outside the world of viruses and viroids. A genome-wide survey resulted in a single example, namely a HDV-like sequence in the human CPEB3 gene [73]. Furthermore, proteins are in general more efficient and versatile catalysts than natural and engineered ribozymes (although this limitation might be overcome in theory) [5], and non-coding transcripts in particular are often evolutionarily very young [99]. It would thus be very surprising if a large set of recently-evolved ribozymes were discovered.

(2) RNA is an ideal molecule for regulatory functions from a theoretical perspective. It is capable of "digital" information processing by specifically interacting with other nucleic acids under the rules of base pairing [51]. On the other hand, it can form elaborate three-dimensional structures which can be recognized by proteins with high specificity. Indeed, several known ncRNAs, including RNAse P RNA, microRNAs, and snoRNAs, do not exert their function in isolation but as part of specific ribonucloprotein complexes. The role of the RNA component is to provide specificity to a target, while the generic function (e.g. creating a base modification as in the case of snoRNAs) is performed by the protein component. In summary, ncRNAs could serve as specific though exchangeable links between nucleic acids and proteins.

(3) Recent attempts to identify classes of RNAs that are seemingly unrelated in sequence but share common structures have not only recovered the expected groups of tRNAs, box H/ACA snoRNAs, and microRNAs, but also provided strong evidence for the existence of previously undescribed classes of structured RNAs [94,71]. The structural similarities within these groups are at least indicative of a common function.

(4) A computational analysis of the binding energies of predicted structured RNA candidates [90] with mRNAs that are not transcribed from the same locus shows that these candidate ncRNAs are enriched in sequences that can bind particularly tightly to near-complementary mRNA targets [85], Fig. 2. Although these data provide only statistical evidence at present, they give a strong indication that direct RNA-RNA interaction is not a rare phenomenon.

(5) Synthetic "modifier RNAs" have been used as experimental techniques for changing the gene expression patterns independent of the RNAi pathway [10,54,62]. One possible mechanism is the modification of (m)RNA structure as a consequence of the duplex formation between modifier and its target (m)RNA, which in turn can dramatically affect the binding affinity of the (m)RNA and a protein [25].
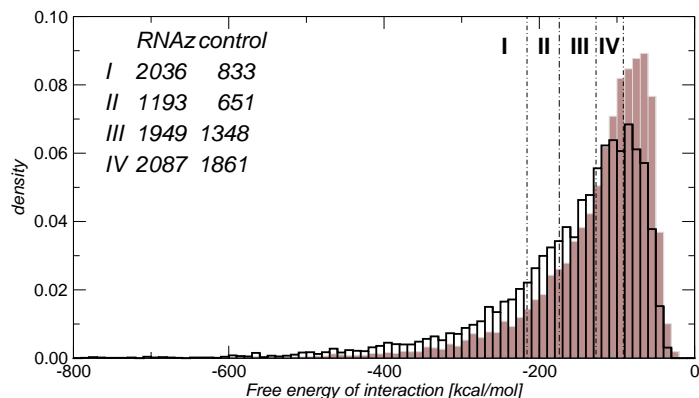
8

Figure 2.   Distribution of interaction free energies of predicted ncRNAs with their best mRNA targets. The density of the interaction free energy distribution of predicted interactions of `RNAz` hits and mRNA is shown in black, the corresponding distribution of randomized control data-set in which the sequences of the `RNAz`-predictions were shuffled is shown in grep. Dotted lines indicate the 0.05, 0.10, 0.25, 0.50 quantiles of the randomized distribution. The insets lists the absolute number of true and randomized interaction in the four energy classes defined by the four quantiles. Figure adapted from [85].

In summary, the available evidence points towards the existence of a large and very diverse pool of functional RNAs whose roles are predominantly regulatory. On the other hand, there is no evidence at present that these functions are performed by means of a uniform mechanism.

5. Elusive Networks: Many More Layers

Ten years ago, gene regulation was essentially understood as the result of two well-separated layers: (1) Signaling pathways would transmit the signal to the nucleus via direct protein-protein interactions and protein modifications (e.g. phosphorylation); (2) Transcription factors would then bind to sequence-specific DNA motifs and recruit the transcription complex to start sites, causing the specific transcription of target mRNA(s). This dichotomy is highlighted by a large number of publications exclusively dealing with *transcription networks* [30,83] and dedicated data bases such as `TRANSPATH` [75] for cell signaling. Taken together, these two subsystems form the *gene regulatory networks* which resulted in a major advance in our understanding of developmental regulation [15].

While our knowledge on signaling and transcription factor binding rapidly grew over time, the general concept stayed the same until two more

"layers" of ubiquitous importance were discovered: epigenetic regulation and post-transcriptional regulation. Again, these additional types of gene regulation were perceived as separate layers, coming with their own set of regulatory molecules and mechanisms.

While regulatory RNAs became known first for their post-transcriptional effects, different classes of ncRNAs are instrumental at most or all layers of the regulatory network [12,64]. Beyond their effect on translation as microRNAs or siRNAs, they could serve as "transport shuttles" for bound proteins, activate and inhibit transcription [43,26], bind to transcription factors [9], act as transcriptional enhancers [18], take part in structuring the nucleus [67], trigger epigenetic activation and inactivation of chromosomal regions (e.g. Xist [59]), DNA elimination during development of the somatic macronucleus in tetrahymena [58], and may even serve as "backup copies" that transmit information to the next generation independent of genomic DNA [44]. While some of these mechanism are observed in diverse eukaryotes or even throughout all kingdoms of life, we have to-date only seen a few individual examples of others, such as the TRE RNAs [77,74] which, tethered to their site of origin, function by anchoring histone-modifying enzymes to their target, or the evf-2 RNA [18], which acts as a transcriptional enhancer. Furthermore, several surprising effects have been observed *in vitro* in reaction to artificial small RNAs, such as "RNAa", i.e., transcriptional *activation* by means of siRNAs [43] and the up-regulation of protein production in response to small RNAs that modify mRNA structure and thereby alter mRNA-protein binding properties [62,54,25].

The structure, content, and diversity of the transcriptome of higher eukaryotes thus forces us to consider a mechanistically much more diverse system in which a substantial amount of the information is stored and processed by the interaction of many different regulatory mechanisms. The direct interaction of regulators from different mechanistic "levels" enables the synchronization of their joint action. For instance, the integration of small ncRNAs and transcription factors in feed-forward loops, which appears to be a common phenomenon, can tighten regulation by combination of both, (slow) transcriptional and (fast) post-transcriptional regulation [80]. The large number and the diversity of distinct functional units encoded by the same DNA locus further suggest dependencies which most likely facilitate additional regulatory mechanisms.

The assumption that a sizable fraction of players and mechanism in cellular regulation so far has escaped discovery and the expectation that these additional molecules strongly interact with known parts of the reg-

ulatory network have far-reaching consequences for our ability to reconstruct regulatory networks from data. In fact, existing methods [48] invariably assume that the vertices of a reasonably autark sub-system are known and complete, so that only the interaction rule need to be inferred. The changing view of the transcriptome, however, implies several important research questions, with potentially dramatic consequences for systems biology: What is the relation between the true and reconstructed networks when vertices are missing in the reconstruction? In other words, to what extent do our current network models represent true interaction and physical regulation? How can we infer missing players in network reconstruction, e.g. from discrepancies between data and incomplete networks? Can such studies actually aid the targeted search for missing components?

6. Space Matters

Molecular interactions partly represented by regulatory networks actually occur in space and require directly interacting partners to physically meet. Integration of the molecular players and their regulation into space and time eventually governs the formation of a three-dimensional organism.

Regulatory molecules are typically not homogeneously distributed in a living cell. In fact, generation of asymmetry where only symmetry has been before is the key to differentiation and an initial divergence in spatial distribution of expression products. It is relevant to all levels of organization from large multicellular organisms to small subcellular compartments. The origin of asymmetry might be found in the dynamic embedding of one-dimensional (e.g. linearly ordered genes on the genome) and two-dimensional (e.g. surfaces of membranes) objects in a three-dimensional space and their physically determined relations. During sporulation of *Bacillus subtilis*, it has been traced back to chromosomal spatial asymmetry and the relative position of two loci (oriC and spoIIAB) on the genome [17] and concentration differences of surface and volume associated factors (spoIIE and spoIIAA) [2]. However, the picture is still incomplete.

Viewing asymmetry as a sorting event, maintenance of an established pattern requires additional efforts as sequestration and/or localization and turns them into important regulatory mechanisms. Incongruity of mRNA and protein distribution have revealed the relevance of translational inhibition (e.g. bcd inhibits cad translation) and cytoskeleton-dependent mRNA localization (e.g. Staufen localizes bcd mRNA) in *Drosophila* embryogenesis [16]. Translational inhibition by miRNAs, mRNA localization [61] and

regulated local translation have been demonstrated to be of considerable importance in the differentiation of dendrites[49,13]. More attention has been payed to subcellular localization of proteins as reflected by the LOCATE Database [19]. The mechanism of intra- and inter-cellular trafficking by protein targeting via signal peptides, direct protein synthesis into the target compartment and usage of the secretory pathway have been described decades ago.

Once a set of regulators is asymmetrically distributed, sometimes in form of a gradient, combinatorial patters of their spatial expression can define new compartments or segments and cause the expression of segment-specific factors [96,4]. These relative positions of spatial expression patterns are biologically meaningful, distinguish biological structures and reveal rules of hierarchical (protein) network organization [37] [78]. Combinatorial molecular phenotypes have been successfully mapped with a technique termed multi-epitope-ligand cartography (MELC or MELK) [78]. Alterations of expression patterns relative to each other have been shown to characterize diseases (e.g. inflammatory skin disease [78]) and are probably instrumental in major evolutionary innovations (e.g. fin-limb transition [56]) indicating clearly that spatial (and temporal) organization of expression products crucially contributes to function. One implication is that subfunctionalization of duplicated genes may also occur in space (e.g. colinear expression of Hox gene paralogs along the anterior-posterior axis).

All possible layers of gene regulation are potentially involved in creating spatial organization. Among the best-know examples are transcriptional enhancers which bind a number of regulators cooperatively [41]. The resulting expression pattern is the evaluation of the expression map of these regulators. A gene can be under the control of several enhancers, each of which represents a sub-pattern of the complete spatial and temporal expression pattern [69,81,82,21] as in the example of eve stripe formation, Fig. 3. Evaluation of the expression map might allow different sets of regulators to cause the same spatial distribution of the regulated gene. Post-transcriptional control as well es a mixture of both can give rise to specific expression patterns [29] but as transcription only occurs in the nucleus, post-transcriptional regulation enables a fine-grained, sub-cellular localization. It has been proposed that splice variants with alternative UTRs probably combine the coding sequence with different elements for post-transcriptional regulation and, consequently, localization [84]. Epigenetic regulation predominantly connects to localized expression on a multicellular level. Fully differentiated cells will only utilize a fraction of all genes for the rest of their lifetime.
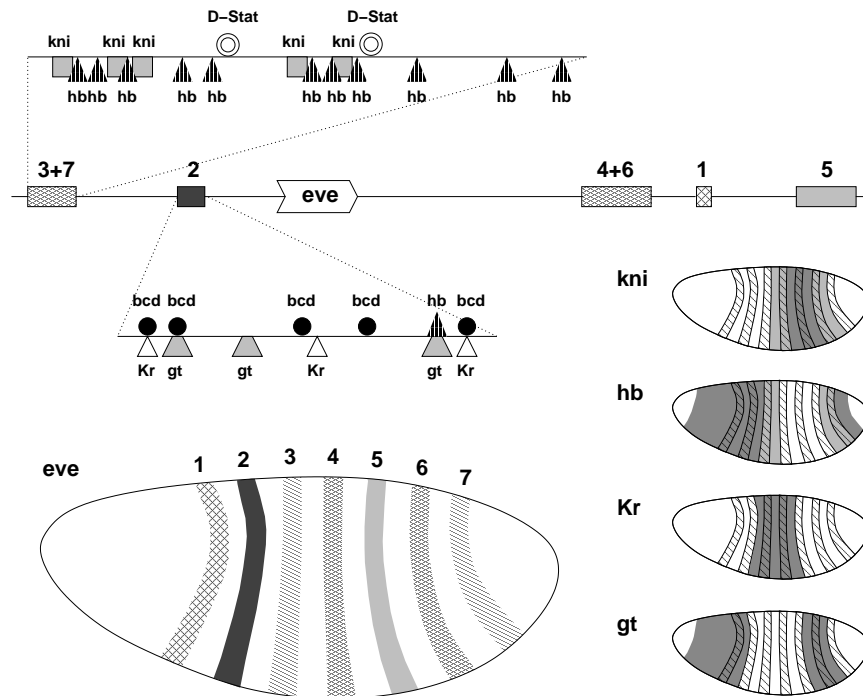
12

Figure 3.   Regulation of the seven even-skipped stripes in early Drosophila development. Five regulatory regions located upstream and downstream of the eve gene, stripe 3+7 enhancer, stripe 2 enhancer, stripe 4+6 enhancer, stripe 1 enhancer and stripe 5 enhancer, each regulate a sub-pattern of the eve expression pattern in the cellularizing syncytial blasdoderm stage of the embryo.   These enhancers are composed of binding sites for transcriptional activators (above the line) and repressors (below the line).   While the activators (e.g. bcd, cad and D-Stat) are rather ubiquitously distributed, the spatially located repressors (e.g. kni, hb, Kr and gt – expression patterns on the right side) define the anterior and posterior borders of eve stripes [81,82,21].   The current picture offers no direct control for the observed even width and spacing of eve stripes.

To repress de-differentiation (and the risk of cancer) early developmental genes and pluripotency-associated genes can be permanently silenced by DNA methylation [68].

Even though much exemplary knowledge about spatial patterns of expression has be gathered over past years, large-scale data sets and comprehensive statistical analyses are not yet available. High-throughput *in-situ* hybridization methods now allow the collection of spatial information (e.g. gene expression patters during the early stages of *Drosophila* development (e.g. BDGP [88]) and search for similar patterns on the basis of image com-

parison (e.g. `FlyExpress` [38]).

Differences in expression of transcripts, RNAs and proteins are due to environmental factors, position effects and the genetic program of the cell. The differences in expression of regulators themselves regulate the spatial distribution of others together with ATP dependent localization mechanisms.

7. Concluding Remarks

In the few years since the completion of the human genome sequence [87], a diverse set of high-throughput experiments has amassed data that profoundly change our understanding of molecular biology. Not only are there an order of magnitude fewer protein coding genes than predicted a decade ago, but the transcriptional output of eukaryotic genome is unexpectedly diverse and covers an unexpectedly large fraction of the genomic DNA. The data confront us with a surprisingly diverse collection of processed RNAs with — as we argue here — diverse, predominantly regulatory, functions which they perform by means of many different molecular mechanisms.

The sheer extent of this regulatory "layer" of ncRNAs implies that ncRNA-based regulation mechanisms should be seen as paradigms of gene regulation rather than as the odd exceptions to a "proteinocentric" picture of molecular biology. Indeed, RNA-related regulation does not appear to be well-separated from the essentially protein-based transcriptional and signaling networks but to incorporate them into a tightly woven web of complex interactions which cannot be understood comprehensively by artificially focusing on particular molecules or interaction mechanisms. We argue that this complexity needs to be integrated into our current modeling approaches [76] for gene regulatory networks.

While the long-known discrepancy between mRNA and protein expression levels has been largely ignored in the interpretation of gene expression experiments, the large number of transcripts and processing products arising from the same genomic locus emphasizes that such a simplification is even conceptually untenable. From the influence of nuclear organization on transcriptional regulation to the spatial expression patterns of RNAs and proteins on a multicellular level, spatial organization will gain importance as data are accumulating. The formation of spatially differentiated expression patterns is ultimately the cause of cell differentiation, tissue formation, and hence the fundamental molecular process in development.

With pattern formation being inherently a dynamical phenomenon, it

14

becomes imperative to consider in more details dynamical aspects of regulation. For instance, it is well know that transcriptional regulation is much slower than post-transcriptional gene silencing by means of microRNAs. We cannot not expect that detailed kinetic data will be available for large fractions of the human transcriptome in the near future (although this is feasible for organism such as yeast because of the much small genome size and simpler transcriptome organization). A much more detailed mechanistic understanding of the interactions and at least an approximately complete picture of the types players and their interaction rules therefore is an indispensable prerequisite for modeling in Systems Biology. In the same context, the consequences of fluctuations in transcript numbers, and in particular of the variation across individuals, can only be understood based on a comprehensive map of the regulatory networks. One would expect that the need for developmental robustness in the presence of substantial noise at the molecular level has shaped the interplay of regulation mechanisms of evolutionary time-scales [33,20].

The changing paradigms in gene regulation imply the need for a corresponding experimental and theoretical research agenda, with topics ranging from large scale tracing of processing pathways, high-throughput functional assays for non-coding transcripts, to computational approaches for unbiased transcript detection, network reconstruction from incomplete data, and comprehensive modelling of spatio-temporal experession patterns.

Acknowledgments

References

1. A. Aravin, D. Gaidatzis, S. Pfeffer, M. Lagos-Quintana, P. Landgraf, N. Iovino, P. Morris, M. J. Brownstein, S. Kuramochi-Miyagawa, T. Nakano, M. Chien, J. J. Russo, J. Ju, R. Sheridan, C. Sander, M. Zavolan, and T. Tuschl. A novel class of small RNAs bind to MILI protein in mouse testes. *Nature*, 442:203–207, 2006.
2. I. Barák and A. J. Wilkinson. Where asymmetry in gene expression originates. *Mol Microbiol*, 57(3):611–620, Aug 2005.
3. S. Ben-Tabou de Leon and E. H. Davidson. Gene regulation: gene control network in development. *Annu Rev Biophys Biomol Struct.*, 36:191–212, 2007.
4. S. Bondos. Variations on a theme: Hox and Wnt combinatorial regulation during animal development. *Sci STKE*, 355:pe38, 2006.

5. R. R. Breaker, G. M. Emilsson, D. Lazarev, S. Nakamura, I. J. Puskarz, A. Roth, and N. Sudarsan. A common speed limit for RNA-cleaving ribozymes and deoxyribozymes. *RNA*, 9:949–957, 2003.

6. G. A. Calin, C. G. Liu, M. Ferracin, T. Hyslop, R. Spizzo, C. Sevignani, M. Fabbri, A. Cimmino, E. J. Lee, S. E. Wojcik, M. Shimizu, E. Tili, S. Rossi, C. Taccioli, F. Pichiorri, X. Liu, S. Zupo, V. Herlea, L. Gramantieri, G. Lanza, H. Alder, L. Rassenti, S. Volinia, T. D. Schmittgen, T. J. Kipps, M. Negrini, and C. M. Croce. Ultraconserved regions encoding ncRNAs are altered in human leukemias and carcinomas. *Cancer Cell*, 12:215–229, 2007.

7. P. Carninci. Constructing the landscape of the mammalian transcriptome. *J. Exp. Biol.*, 210:1497–1506, 2007.

8. P. Carninci, T. Kasukawa, S. Katayama, J. Gough, M. Frith, N. Maeda, R. Oyama, T. Ravasi, B. Lenhard, C. Wells, and *et al.; FANTOM Consortium; RIKEN Genome Exploration Research Group and Genome Science Group (Genome Network Project Core Group).* The transcriptional landscape of the mammalian genome. *Science*, 309:1559–1563, 2005.

9. L. A. Cassiday and L. J. Maher. Having it both ways: transcription factors that bind dna and rna. *Nucleic Acids Res*, 30(19):4118–4126, Oct 2002.

10. J. L. Childs, M. D. Disney, and D. H. Turner. Oligonucleotide directed misfolding of RNA inhibits *Candida albicans* group I intron splicing. *Proc. Natl. Acad. Sci. USA*, 99:11091–11096, 2002.

11. H. A. Coller, J. J. Forman, and A. Legesse-Miller. "Myc'ed messages": *myc* induces transcription of E2F1 while inhibiting its translation via a microRNA polycistron. *PLoS Genet.*, 3:e146, 2007.

12. F. F. Costa. Non-coding RNAs: new players in eukaryotic biology. *Gene*, 357(2):83–94, 2005.

13. R. Dahm, M. Kiebler, and P. Macchi. RNA localisation in the nervous system. *Semin Cell Dev Biol*, 18(2):216–223, Apr 2007.

14. L. David, W. Huber, M. Granovskaia, J. Toedling, C. J. Palm, L. Bofkin, T. Jones, R. W. Davis, and L. M. Steinmetz. A high-resolution map of transcription in the yeast genome. *Proc. Natl. Acad. Sci. USA*, 103:5320–5325, 2006.

15. E. Davidson. *The Regulatory Genome.* Elsevier, Amsterdam, 2006.

16. W. Driever and C. Nüsslein-Volhard. A gradient of bicoid protein in drosophila embryos. *Cell*, 54(1):83–93, Jul 1988.

17. J. Dworkin and R. Losick. Differential gene expression governed by chromosomal spatial asymmetry. *Cell*, 107(3):339–346, Nov 2001.

18. J. Feng, C. Bi, B. S. Clark, R. Mady, P. Shah, and J. D. Kohtz. The Evf-2 noncoding RNA is transcribed from the Dlx-5/6 ultra conserved region and functions as a Dlx-2 transcriptional coactivator. *Genes Dev*, 20:1470–1484, 2006.

19. J. L. Fink, R. N. Aturaliya, M. J. Davis, F. Zhang, K. Hanson, M. S. Teasdale, C. Kai, J. Kawai, P. Carninci, Y. Hayashizaki, and R. D. Teasdale. Locate: a mouse protein subcellular localization database. *Nucleic Acids Res*, 34(Database issue):213–217, 2006.

20. M. A. Fuentes and D. C. Krakauer. The evolution of developmental pat-

16

terning under genetic duplication constraints. *J R Soc Interface*, Jun 2007. 10.1098/rsif.2007.1074.

21. M. Fujioka, Y. Emi-Sarker, G. L. Yusibova, T. Goto, and J. B. Jaynes. Analysis of an even-skipped rescue transgene reveals both composite and discrete neuronal and early blastoderm enhancers, and multi-stripe positioning by gap gene repressor gradients. *Development*, 126(11):2527–2538, Jun 1999.

22. M. B. Gerstein, C. Bruce, J. S. Rozowsky, D. Zheng, J. Du, J. O. Korbel, O. Emanuelsson, Z. D. Zhang, S. Weissman, and M. Snyder. What is a gene, post-ENCODE? history and updated definition. *Genome Res.*, 17:669–681, 2007.

23. N. Gilbert and B. Ramsahoye. The relationship between chromatin structure and transcriptional activity in mammalian genomes. *Brief Funct Genomic Proteomic*, 4(2):129–142, Jul 2005.

24. T. R. Gingeras. Origin of phenotypes: Genes and transcripts. *Genome Res.*, 17:682–690, 2007.

25. J. Hackermüller, N.-C. Meisner, M. Auer, M. Jaritz, and P. F. Stadler. The effect of RNA secondary structures on RNA-ligand binding and the modifier RNA mechanism: A quantitative model. *Gene*, 345:3–12, 2005.

26. J. Han, D. Kim, and K. V. Morris. Promoter-associated RNA is required for RNA-directed transcriptional gene silencing in human cells. *Proc. Natl. Acad. Sci. USA*, 104:12422–12427, 2007.

27. M. Havilio, E. Y. Levanon, G. Lerman, M. Kupiec, and E. Eisenberg. Evidence for abundant transcription of non-coding regions in the sac charomyces cerevisiae genome. *BMC Genomics*, 6:93, 2005.

28. H. He, J. Wang, T. Liu, X. S. Liu, T. Li, Y. Wang, Z. Qian, H. Zheng, X. Zhu, T. Wu, B. Shi, W. Deng, W. Zhou, G. Skogerbø, and R. Chen. Mapping the *C. elegans* noncoding transcriptome with a whole-genome tiling microarray. *Genome Res.*, 2007. DOI: 10.1101/gr.6611807.

29. O. Hobert. Common logic of transcription factor and microRNA action. *Trends Biochem Sci*, 29(9):462–468, Sep 2004.

30. J. Ihmels, S. Bergmann, M. Gerami-Nejad, I. Yanai, M. McClellan, J. Berman, and N. Barkai. Rewiring of the yeast transcriptional network through the evolution of motif usage. *Science*, 309:938–940, 2005.

31. Y. Ikeda, R. S. Daughters, and L. P. Ranum. Bidirectional expression of the SCA8 expansion mutation: One mutation, two genes. *Cerebellum*, 2007. doi: 10.1080/14734220701413781.

32. O. Isken and L. E. Maquat. Quality control of eukaryotic mRNA: safeguarding cells from abnormal mRNA function. *Genes Dev.*, 21:1833–1856, 2007.

33. L. J. Johnson and J. F. Brookfield. Evolution of spatial expression pattern. *Evol Dev*, 5(6):593–599, Nov-Dec 2003.

34. D. Kampa, J. Cheng, P. Kapranov, M. Yamanaka, S. Brubaker, S. Cawley, J. Drenkow, A. Piccolboni, S. Bekiranov, G. Helt, H. Tammana, and T. R. Gingeras. Novel RNAs identified from an in-depth analysis of the transcriptome of human chromosomes 21 and 22. *Genome Res.*, 14:331–342, 2004.

35. P. Kapranov, J. Cheng, S. Dike, D. Nix, R. Duttagupta, A. T. Willingham, P. F. Stadler, J. Hertel, J. Hackermüller, I. L. Hofacker, I. Bell, E. Cheung,

J. Drenkow, E. Dumais, S. Patel, G. Helt, G. Madhavan, A. Piccolboni, V. Sementchenko, H. Tammana, and T. R. Gingeras. RNA maps reveal new RNA classes and a possible function for pervasive transcription. *Science*, 316:1484–1488, 2007.

36. S. Katayama, Y. Tomaru, T. Kasukawa, K. Waki, M. Nakanishi, M. Nakamura, H. Nishida, C. C. Yap, M. Suzuki, J. Kawai, H. Suzuki, P. Carninci, Y. Hayashizaki, C. Wells, M. Frith, T. Ravasi, K. C. Pang, J. Hallinan, J. Mattick, D. A. Hume, L. Lipovich, S. Batalov, P. G. Engström, Y. Mizuno, M. A. Faghihi, A. Sandelin, A. M. Chalk, S. Mottagui-Tabar, Z. Liang, B. Lenhard, C. Wahlestedt, RIKEN Genome Exploration Research Group, Genome Science Group (Genome Network Project Core Group), and FANTOM Consortium. Antisense transcription in the mammalian transcriptome. *Science*, 309:1564–1556, 2005.

37. D. Kosman, C. M. Mizutani, D. Lemons, W. G. Cox, W. McGinnis, and E. Bier. Multiplex detection of RNA expression in drosophila embryos. *Science*, 305(5685):846–846, 2004.

38. S. Kumar, K. Jayaraman, S. Panchanathan, R. Gurunathan, A. Marti-Subirana, and S. J. Newfeld. Best: a novel computational approach for comparing gene expression patterns from early stages of drosophila melanogaster development. *Genetics*, 162(4):2037–2047, Dec 2002.

39. M. Lapidot and Y. Pilpel. Genome-wide natural anti-sense transcription: coupling its regulation to its different regulatory mechanisms. *EMBO Rep.*, 7:1216–1222, 2006.

40. N. C. Lau, A. G. Seto, J. Kim, S. Kuramochi-Miyagawa, T. Nakano, D. P. Bartel, and R. E. Kingston. Characterization of the piRNA complex from rat testes. *Science*, 313:363–367, 2006.

41. D. Lebrecht, M. Foehr, E. Smith, F. J. Lopes, C. E. Vanario-Alonso, J. Reinitz, D. S. Burz, and S. D. Hanes. Bicoid cooperative dna binding is critical for embryonic patterning in drosophila. *Proc Natl Acad Sci U S A*, 102(37):13176–13181, Sep 2005.

42. E. Leygue. Steroid receptor RNA activator (SRA1): unusual bifaceted gene products with suspected relevance to breast cancer. *Nucl. Recep. Signaling*, 5:e006, 2007.

43. L. C. Li, S. T. Okino, H. Zhao, D. Pookot, R. F. Place, S. Urakami, H. Enokida, and R. Dahiya. Small dsRNAs induce transcriptional activation in human cells. *Proc. Natl. Acad. Sci. USA*, 103:17337–17342, 2006.

44. S. J. Lolle, J. L. Victor, J. M. Young, and R. E. Pruitt. Genome-wide non-mendelian inheritance of extra-genomic information in arabidopsis. *Nature*, 434(7032):505–509, Mar 2005.

45. N. Maeda, T. Kasukawa, R. Oyama, J. Gough, M. Frith, P. G. Engström, B. Lenhard, R. N. Aturaliya, S. Batalov, K. W. Beisel, C. J. Bult, C. F. Fletcher, A. R. Forrest, M. Furuno, D. Hill, M. Itoh, M. Kanamori-Katayama, S. Katayama, M. Katoh, T. Kawashima, J. Quackenbush, T. Ravasi, B. Z. Ring, K. Shibata, K. Sugiura, Y. Takenaka, R. D. Teasdale, C. A. Wells, Y. Zhu, C. Kai, J. Kawai, D. A. Hume, P. Carninci, and Y. Hayashizaki. Transcript annotation in FANTOM3: Mouse gene catalog based on physical

18

cdnas. *PLoS Genetics*, 2:e62, 2006. doi:10.1371/journal.pgen.0020062.

46. G. Mainguy, J. Koster, J. Woltering, H. Jansen, and D. A. Extensive poly-cistronism and antisense transcription in the mammalian Hox clusters. *PLoS ONE*, 2:e356, 2007.

47. J. R. Manak, S. Dike, V. Sementchenko, P. Kapranov, F. Biemar, J. Long, J. Cheng, I. Bell, S. Ghosh, A. Piccolboni, and T. R. Gingeras. Biological function of unannotated transcription during the early development of Drosophila melanogaster. *Nat Genet*, 38:1151–1158, 2006.

48. F. Markowetz and R. Spang. Inferring cellular networks — a review. *BMC Bioinformatics*, 8(Suppl 6):S5, 2007.

49. K. C. Martin and R. S. Zukin. RNA trafficking and local protein synthesis in dendrites: an overview. *J Neurosci*, 26(27):7131–7134, 2006.

50. S. C. Materna and E. H. Davidson. Logic of gene regulatory networks. *Curr Opin Biotechnol.*, 18:351–354, 2007.

51. J. S. Mattick. A new paradigm for developmental biology. *J. Exp. Biol.*, 210:1526–1547, 2007.

52. J. S. Mattick and I. V. Makunin. Non-coding RNA. *Hum Mol Genet.*, 15:R17–29, 2006.

53. A. Mazo, J. W. Hodgson, S. Petruk, Y. Sedkov, and H. W. Brock. Transcriptional interference: an unexpected layer of complexity in gene regulation. *J Cell Sci.*, 120:2755–2761, 2007.

54. N.-C. Meisner, J. Hackermüller, V. Uhl, A. Aszódi, M. Jaritz, and M. Auer. mRNA openers and closers: A methodology to modulate AU-rich element controlled mRNA stability by a molecular switch in mRNA conformation. *Chembiochem.*, 5:1432–1447, 2004.

55. L. M. Mendes Soares and J. Valcárcel. The expanding transcriptome: the genome as the 'Book of Sand'. *EMBO J.*, 25:923–931, 2006.

56. B. D. Metscher, K. Takahashi, K. Crow, C. Amemiya, D. F. Nonaka, and G. P. Wagner. Expression of hoxa-11 and hoxa-13 in the pectoral fin of a basal ray-finned fish, polyodon spathula: implications for the origin of tetrapod limbs. *Evol Dev*, 7(3):186–195, May-Jun 2005.

57. F. Miura, N. Kawaguchi, J. Sese, A. Toyoda, M. Hattori, S. Morishita, and T. Ito. A large-scale full-length cDNA analysis to explore the budding yeast transcriptome. *Proc. Natl. Acad. Sci. USA*, 103:17846–17851, 2006.

58. K. Mochizuki and M. A. Gorovsky. Small RNAs in genome rearrangement in tetrahymena. *Curr Opin Genet Dev.*, 14:181–187, 2004.

59. K. Ng, D. Pullirsch, M. Leeb, and A. Wutz. *Xist* and the order of silencing. *EMBO Rep.*, 8:34–39, 2007.

60. Y. Okazaki, M. Furuno, T. Kasukawa, J. Adachi, H. Bono, S. Kondo, I. Nikaido, N. Osato, R. Saito, H. Suzuki, and *et al.* Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs. *Nature*, 420:563–573, 2002.

61. I. M. Palacios and D. St Johnston. Getting the message across: the intracellular localization of mRNAs in higher eukaryotes. *Annu Rev Cell Dev Biol*, 17:569–614, 2001.

62. M. Paulus, M. Haslbeck, and M. Watzele. RNA stem-loop enhanced expres-

sion of previously non-expressible genes. *Nucl. Acids Res.*, 32:9/e78, 2004. doi 10.1093/nar/gnh076.

63. J. S. Pedersen, G. Bejerano, A. Siepel, K. Rosenbloom, K. Lindblad-Toh, E. S. Lander, J. Kent, W. Miller, and D. Haussler. Classification of conserved RNA secondary structures in the human genome. *PLoS Comput. Biol.*, 2:e33, 2006.

64. K. V. Prasanth and D. L. Spector. Eukaryotic regulatory RNAs: an answer to the 'genome complexity' conundrum. *Genes Dev.*, 21:11–42, 2007.

65. T. Ravasi, H. Suzuki, K. C. Pang, S. Katayama, M. Furuno, R. Okunishi, S. Fukuda, K. Ru, M. C. Frith, M. M. Gongora, S. M. Grimmond, D. A. Hume, Y. Hayashizaki, and J. S. Mattick. Experimental validation of the regulated expression of large numbers of non-coding RNAs from the mouse genome. *Genome Res.*, 16:11–19, 2006.

66. T. Ravasi, H. Suzuki, K. C. Pang, S. Katayama, M. Furuno, R. Okunishi, S. Fukuda, K. Ru, M. C. Frith, M. M. Gongora, S. M. Grimmond, D. A. Hume, Y. Hayashizaki, and J. S. Mattick. Experimental validation of the regulated expression of large numbers of non-coding RNAs from the mouse genome. *Genome Res.*, 16:11–19, 2006.

67. S. V. Razin, A. Rynditch, V. Borunova, E. Ioudinkova, V. Smalko, and K. Scherrer. The 33 kb transcript of the chicken alpha-globin gene domain is part of the nuclear matrix. *J Cell Biochem*, 92:445–457, 2004.

68. W. Reik. Stability and flexibility of epigenetic gene regulation in mammalian development. *Nature*, 447(7143):425–432, May 2007.

69. G. Riddihough and D. Ish-Horowicz. Individual stripe regulatory elements in the drosophila hairy promoter respond to maternal, gap, and pair-rule genes. *Genes Dev*, 5(5):840–854, 1991.

70. J. L. Rinn, M. Kertesz, J. K. Wang, S. L. Squazzo, X. Xu, S. A. Brugmann, L. H. Goodnough, J. A. Helms, P. J. Farnham, E. Segal, and H. Y. Chang. Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell*, 29:1311–1323, 2007.

71. D. R. Rose, J. Hackermüller, S. Washietl, S. Findeiß, K. Reiche, J. Hertel, P. F. Stadler, and S. J. Prohaska. Computational RNomics of drosophilids. *BMC Genomics*, 2007. accepted.

72. J. G. Ruby, C. Jan, C. Player, M. J. Axtell, W. Lee, C. Nusbaum, H. Ge, and D. P. Bartel. Large-scale sequencing reveals 21U-RNAs and additional microRNAs and endogenous siRNAs in *C. elegans*. *Cell*, 127:1193–1207, 2006.

73. K. Salehi-Ashtiani, A. Lupták, A. Litovchick, and J. W. Szostak. A genomewide search for ribozymes reveals an HDV-like sequence in the human CPEB3 gene. *Science*, 313:1788–1792, 2006.

74. T. Sanchez-Elsner, D. Gou, E. Kremmer, and F. Sauer. Noncoding RNAs of trithorax response elements recruit Drosophila Ash1 to Ultrabithorax. *Science*, 311:1118–1123, 2006.

75. F. Schacherer, C. Choi, U. Götze, M. Krull, S. Pistor, and E. Wingender. The TRANSPATH signal transduction database: a knowledge base on signal transduction networks. *Bioinformatics*, 17:1053–1057, 2001.

76. T. Schlitt and A. Brazma. Current approaches to gene regulatory network

20

modelling. *BMC Bioinformatics*, 8 (Suppl 6):S9, 2007.

77. S. Schmitt and R. Paro. RNA at the steering wheel. *Genome Biol*, 7(5):218–218, 2006.

78. W. Schubert, B. Bonnekoh, A. J. Pommer, L. Philipsen, R. Böckelmann, Y. Malykh, H. Gollnick, M. Friedenberger, M. Bode, and A. W. Dress. Analyzing proteome topology and function by automated multidimensional fluorescence microscopy. *Nat Biotechnol*, 24(10):1270–1278, Oct 2006.

79. K. E. Shearwin, B. P. Callen, and J. B. Egan. Transcriptional interference—a crash course. *Trends Genet.*, 21:339–345, 2005.

80. Y. Shimoni, G. Friedlander, G. Hetzroni, G. Niv, S. Altuvia, O. Biham, and H. Margalit. Regulation of gene expression by small non-coding RNAs: a quantitative view. *Mol. Syst. Biol.*, 3:138, 2007.

81. S. Small, A. Blair, and M. Levine. Regulation of even-skipped stripe 2 in the drosophila embryo. *EMBO J*, 11(11):4047–4057, Nov 1992.

82. S. Small, A. Blair, and M. Levine. Regulation of two pair-rule stripes by a single enhancer in the drosophila embryo. *Dev Biol*, 175(2):314–324, May 1996.

83. S. Soneji, S. Huang, M. Loose, I. J. Donaldson, R. Patient, B. Gottgens, T. Enver, and G. May. Inference, validation, and dynamic modeling of transcription networks in multipotent hematopoietic cells. *Ann N Y Acad Sci.*, 1106:30–40, 2007.

84. A. Stark, J. Brennecke, N. Bushati, R. B. Russell, and S. M. Cohen. Animal microRNAs confer robustness to gene expression and have a significant impact on 3'UTR evolution. *Cell*, 123(6):1133–1146, Dec 2005.

85. The Athanasius F. Bompfünewerer RNA Consortium:, R. Backofen, C. Flamm, C. Fried, G. Fritzsch, J. Hackermüller, J. Hertel, I. L. Hofacker, K. Missal, S. J. Mosig, Axel Prohaska, D. Rose, P. F. Stadler, A. Tanzer, S. Washietl, and W. Sebastian. RNAs everywhere: Genome-wide annotation of structured RNAs. *J. Exp. Zool. B: Mol. Dev. Evol.*, 308B:1–25, 2007.

86. The ENCODE Project Consortium. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature*, 447:799–816, 2007.

87. The Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. *Nature*, 431:931–945, 2004.

88. P. Tomancak, A. Beaton, R. Weiszmann, E. Kwan, S. Shu, S. E. Lewis, S. Richards, M. Ashburner, V. Hartenstein, S. E. Celniker, and G. M. Rubin. Systematic determination of patterns of gene expression during drosophila embryogenesis. *Genome Biol*, 3(12), 2002.

89. E. Torarinsson, M. Sawera, J. Havgaard, M. Fredholm, and J. Gorodkin. Thousands of corresponding human an mouse genomic regions unalignable in primary sequece contain common RNA structure. *Genome Res.*, 16:885–889, 2006.

90. S. Washietl, I. L. Hofacker, M. Lukasser, A. Hüttenhofer, and P. F. Stadler. Mapping of conserved RNA secondary structures predicts thousands of functional non-coding RNAs in the human genome. *Nature Biotech.*, 23:1383–1390, 2005.

21

91. S. Washietl, I. L. Hofacker, and P. F. Stadler. Fast and reliable prediction of noncoding RNAs. *Proc. Natl. Acad. Sci. USA*, 102:2454–2459, 2005.

92. S. Washietl, J. S. Pedersen, J. O. Korbel, A. Gruber, J. Hackermüller, J. Hertel, M. Lindemeyer, K. Reiche, C. Stocsits, A. Tanzer, C. Ucla, C. Wyss, S. E. Antonarakis, F. Denoeud, J. Lagarde, J. Drenkow, P. Kapranov, T. R. Gingeras, R. Guigó, M. Snyder, M. B. Gerstein, A. Reymond, I. L. Hofacker, and P. F. Stadler. Structured RNAs in the ENCODE selected regions of the human genome. *Gen. Res.*, 17:852–864, 2007.

93. C. Weile, P. P. Gardner, M. M. Hedegaard, and J. Vinther. Use of tiling array data and RNA secondary structure predictions to identify noncoding RNA genes. *BMC Genomics*, 8:244, 2007.

94. S. Will, K. Missal, I. L. Hofacker, P. F. Stadler, and R. Backofen. Inferring non-coding RNA families and classes by means of genome-scale structure-based clustering. *PLoS Comp. Biol.*, 3:e65, 2007.

95. A. T. Willingham and T. R. Gingeras. TUF love for "junk" DNA. *Cell*, 125:1215–1220, 2006.

96. X. Wu, R. Vakani, and S. Small. Two distinct mechanisms for differential positioning of gene expression borders involving the drosophila gap protein giant. *Development*, 125(19):3765–3774, 1998.

97. F. Wyers, M. Rougemaille, G. Badis, J. C. Rousselle, M. E. Dufour, J. Boulay, B. Régnault, F. Devaux, A. Namane, B. Séraphin, D. Libri, and J. A. Cryptic pol II transcripts are degraded by a nuclear quality control pathway involving a new poly(A) polymerase. *Cell*, 121:725–737, 2005.

98. O. Yazgan and J. E. Krebs. Noncoding but nonexpendable: transcriptional regulation by large noncoding RNA in eukaryotes. *Biochem Cell Biol.*, 85:484–496, 2007.

99. Z. Zhang, A. W. Pang, and M. Gerstein. Comparative analysis of genome tiling array data reveals many novel primate-specific functional RNAs in human. *BMC Evol Biol.*, 7 (Suppl 1):S14, 2007.