Genomics Proteomics Bioinformatics 12 (2014) 57–63

Genomics Proteomics Bioinformatics

www.elsevier.com/locate/gpb
www.sciencedirect.com

ESSAY

Ribogenomics: the Science and Knowledge of RNA

Jiayan Wu, Jingfa Xiao, Zhang Zhang, Xumin Wang, Songnian Hu, Jun Yu *

CAS Key Laboratory of Genome Sciences and Information, Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing 100101, China

Received 4 April 2014; accepted 14 April 2014
Available online 24 April 2014

KEYWORDS

RNA;
Ribogenomics;
Operational;
Informational;
Transcriptome

Abstract Ribonucleic acid (RNA) deserves not only a dedicated field of biological research — a discipline or branch of knowledge — but also explicit definitions of its roles in cellular processes and molecular mechanisms. Ribogenomics is to study the biology of cellular RNAs, including their origin, biogenesis, structure and function. On the informational track, messenger RNAs (mRNAs) are the major component of ribogenomes, which encode proteins and serve as one of the four major components of the translation machinery and whose expression is regulated at multiple levels by other operational RNAs. On the operational track, there are several diverse types of RNAs — their length distribution is perhaps the most simplistic stratification — involving in major cellular activities, such as chromosomal structure and organization, DNA replication and repair, transcriptional/post-transcriptional regulation, RNA processing and routing, translation and cellular energy/metabolism regulation. An all-out effort exceeding the magnitude of the Human Genome Project is of essence to construct just mammalian transcriptomes in multiple contexts including embryonic development, circadian and seasonal rhythms, defined life-span stages, pathological conditions and anatomy-driven tissue/organ/cell types.

Introduction

Ribogenomics is the science and knowledge about ribonucleic acid (RNA). As one of the four major macromolecules (percentage weight in mammalian cell: DNA, ~7 pg, 0.3%; RNA, ~20 pg, 1%; protein, ~500 pg, 20%; and polysaccharide, ~2 μg, 78.7% [1,2]) of cellular life forms, RNA deserves not only a dedicated research field but also definitions of its roles in cellular processes and molecular mechanisms. Therefore, ribogenomics, at least in a sense of cellular mass, in terms of research focus and priority, may not be more imperative than proteomics but certainly has no reason to draw less attention than genomics.

RNA molecules can be divided into two essential functional categories: operational (including what have been defined as catalytic) and informational. At the center of the informational RNAs (other types of informational RNAs including those guiding processes that change mRNA sequences) is messenger RNA (mRNA). In a typical mammalian cell, mRNA takes ~4% of the total RNA mass and aside from 80% ribosomal RNA (rRNA), other operational RNAs make up the rest. If we take the constitutive RNAs — transfer RNAs (tRNAs) and rRNAs — out of the total, the ratio of the operational RNA *vs.* the informational RNA [3,4] is about four.

* Corresponding author.
E-mail: junyu@big.ac.cn (Yu J).

Peer review under responsibility of Beijing Institute of Genomics, Chinese Academy of Sciences and Genetics Society of China.

ELSEVIER Production and hosting by Elsevier

http://dx.doi.org/10.1016/j.gpb.2014.04.002

What are the operational RNA types in the dynamic portion of the total RNA? First, all RNA macromolecules are operational. Only the protein-coding portion of all mRNAs is relatively informational, together with certain sequence-specific guiding RNAs (including small sequence-matching RNAs) that may actually aid RNA editing and splicing, while the chemical entity of them remains operational. Second, all non-coding RNAs (ncRNAs) are exclusively operational, including mRNA-like transient transcripts that are often generated from gene duplications — genome-wide, segmental or individual — but may not be translated into functional proteins [5–7]. Transcripts of such kind have been confusing as some of them neither are conserved across closely-related species nor contain normal reading frames albeit often polyadenylated [8,9]. Third, all small RNAs (sRNAs), such as microRNAs (miRNAs), small nuclear RNAs (snRNAs), tRNA-derived sRNAs (tsRNAs), small nucleolar RNAs (snoRNAs) and small interfering RNAs (siRNAs), are operational although they may be processed to become functional in different ways [10–12]. Fourth, long ncRNAs (lncRNAs), intron-encoded or intergenic sequence-encoded, are also all operational, which may act on different aspects of cellular activities and mechanisms [12]. In this article, we first divide ribogenomes into informational and operational tracks, pointing out the obvious differences and intricate relationships between the two tracks, and then provide insights on the research scopes and fundamental scientific questions of ribogenomics under such a scheme.

Ribogenomics on the operational track

Life had started with RNAs [13–15]. Molecular mechanisms and cellular processes of the operational ribogenomic track have to be created earlier than those of the informational track until the genetic code was created [16–20]. We have argued before that early RNA-built life forms may have begun as eukaryote-like organisms since simple life forms might not be able to utilize DNA at all initially and bacteria might be too greedy to keep all complicated molecular mechanisms going, such as RNA splicing and polyadenylation [16]. Fundamentally, RNA macromolecules and their intermediates, as well as building blocks, must have performed all essential cellular functions but some may have lost to proteins over evolutionary time scales. Therefore, active searches for the function of various RNA macromolecules should focus on systematic discovery at all levels and for all facets of molecule mechanisms and cellular processes rather than taking the attitude of "guarding the stump for dumb hare to hit on".

Operational RNAs are diverse in function (**Figure 1**) as well as in sequence length and genomic origin [12,21]. In function-seeking studies, any effort should include both size classes,

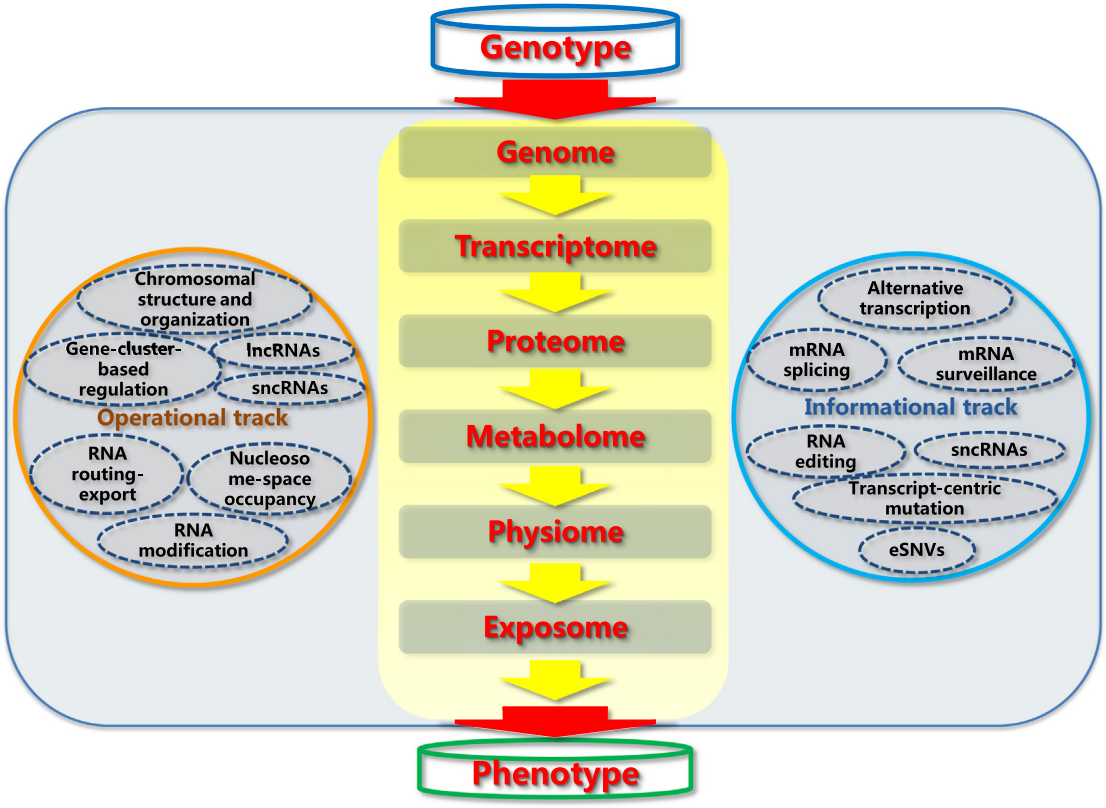

Figure 1 Schematic view from genotype to phenotype with informational and operational tracks

Ribogenomics in the context of a genotype-to-genotype view. Genotype becomes one of the deterministic factors that include ribogenomic, epigenomic, homoeostatic, compartmental and plastic tracks. For the sake of discussion, here we simply classify non-coding RNAs (ncRNAs) into small ncRNAs (sncRNAs) and long ncRNAs (lncRNAs). To emphasize the influence of transcript-centric mutations, we identify expression-related simple nucleotide variations (eSNVs) as an important class of sequence variations that are not yet considered in the context of traditional population genetics and evolution.