Nucleic Acids Research Advance Access published October 13, 2016

Nucleic Acids Research, 2016 1
doi: 10.1093/nar/gkw909

Global and cell-type specific properties of lincRNAs with ribosome occupancy

Hongwei Wang[1,*,†], Yan Wang[1,†], Shangqian Xie[1], Yizhi Liu[1] and Zhi Xie[1,2,*]

[1]State Key Laboratory of Ophthalmology, Guangdong Provincial Key Lab of Ophthalmology and Visual Science, Zhongshan Ophthalmic Center, Sun Yat-sen University, Guangzhou, China and [2]Center for Precision Medicine, Collaborative Innovation Center for Cancer Medicine, Sun Yat-sen University, Guangzhou 510060, China

Received June 20, 2016; Revised September 24, 2016; Accepted October 10, 2016

ABSTRACT

Advances in transcriptomics have led to the discovery of a large number of long intergenic non-coding RNAs (lincRNAs), which are now recognized as important regulators of diverse cellular processes. Although originally thought to be non-coding, recent studies have revealed that many lincRNAs are bound by ribosomes, with a few lincRNAs even having ability to generate micropeptides. The question arises: how widespread the translation of lincRNAs may be and whether such translation is likely to be functional. To better understand biological relevance of lincRNA translation, we systematically characterized lincRNAs with ribosome occupancy by the expression, structural, sequence, evolutionary and functional features for eight human cell lines, revealed that lincRNAs with ribosome occupancy have remarkably distinctive properties compared with those without ribosome occupancy, indicating that translation has important biological implication in categorizing and annotating lincRNAs. Further analysis revealed lincRNAs exhibit remarkable cell-type specificity with differential translational repertoires and substantial discordance in functionality. Collectively, our analyses provide the first attempt to characterize global and cell-type specific properties of translation of lincRNAs in human cells, highlighting that translation of lincRNAs has clear molecular, evolutionary and functional implications. This study will facilitate better understanding of the diverse functions of lincRNAs.

INTRODUCTION

Long intergenic non-coding RNAs (lincRNAs) are an abundant class of endogenous RNA molecules that are transcribed from intergenic regions of the genome. Although originally defined as non-coding RNAs, accumulating evidence has revealed that lincRNAs play important roles in many cellular processes (1–3). The aberrant expression of lincRNAs has been associated with a wide variety of human diseases such as cancer, aging and ocular disorders (4–6), making them attractive candidates for biomarkers and therapeutic targets.

Notably, despite receiving remarkable attention in recent years, the biological roles of the majority of lincRNAs remain largely unknown. Due to the diverse functions and molecular mechanisms, lincRNAs are far more complex than initially thought. Previous studies have suggested they may act as signals, decoys, guides and scaffolds to regulate the expression of either neighbouring genes in cis or distant genes in trans (7). In recent years, advances in genomic technologies have made comprehensive understanding of lincRNA functions feasible (8). It is now possible, for example, to directly identify genomic localization of lincRNAs using chromatin isolation by RNA purification (ChIRP), to dissect biochemical partners using capture hybridization analysis of RNA targets (Chart) and to investigate biological functions using clustered regularly interspaced short palindromic repeat (CRISPR) (9–11).

Recently developed ribosome profiling allows us to globally monitor translation of transcripts by measuring RNAs associated with 80S ribosomes in cells (12,13). Many studies using ribosome profiling have shown apparent ribosome occupancy inside and outside of protein-coding regions, including lincRNA regions (14–17). Although the density of ribosomes in lincRNA regions is lower than that of protein-coding regions, several previous studies have suggested that many lincRNAs may undergo active translation and this translation closely resembles that observed at the 5′ leaders of protein-coding genes (14–15,17). Beyond these, more recently, emerging evidence has shown the existence of short peptides encoded by small open reading frames (sORFs) on lincRNAs (18–20), revealing that lincRNAs could be an important source of new peptides (16) and even orches-

*To whom correspondence should be addressed. Tel: +86 20 8733 5131; Fax: +86 20 87333271; Email: xiezhi@gmail.com
Correspondence may also be addressed to Hongwei Wang. Tel: +86 20 8733 5131; Fax: +86 20 87333271; Email: biocewhw@126.com
†These authors contributed equally to the paper as first authors.

Nucleic Acids Research Advance Access published October 18, 2016

Nucleic Acids Research, 2016 1
doi: 10.1093/nar/gkw950

MethSMRT: an integrative database for DNA N6-methyladenine and N4-methylcytosine generated by single-molecular real-time sequencing

Pohao Ye[1,†], Yizhao Luan[1,†], Kaining Chen[1], Yizhi Liu[1], Chuanle Xiao[1,*] and Zhi Xie[1,2,3,*]

[1]State Key Laboratory of Ophthalmology, Guangdong Provincial Key Lab of Ophthalmology and Visual Science, Zhongshan Ophthalmic Center, Sun Yat-sen University, Guangzhou 510060, China, [2]Center for Precision Medicine, Collaborative Innovation Center for Cancer Medicine, Sun Yat-sen University, Guangzhou 510000, China and [3]Collaborative Innovation Center of Genetics and Development, Fudan University, Shanghai 200438, China

Received August 11, 2016; Revised September 30, 2016; Editorial Decision October 07, 2016; Accepted October 10, 2016

ABSTRACT

DNA methylation is an important type of epigenetic modifications, where 5- methylcytosine (5mC), 6-methyadenine (6mA) and 4-methylcytosine (4mC) are the most common types. Previous efforts have been largely focused on 5mC, providing invaluable insights into epigenetic regulation through DNA methylation. Recently developed single-molecule real-time (SMRT) sequencing technology provides a unique opportunity to detect the less studied DNA 6mA and 4mC modifications at single-nucleotide resolution. With a rapidly increased amount of SMRT sequencing data generated, there is an emerging demand to systematically explore DNA 6mA and 4mC modifications from these data sets. MethSMRT is the first resource hosting DNA 6mA and 4mC methylomes. All the data sets were processed using the same analysis pipeline with the same quality control. The current version of the database provides a platform to store, browse, search and download epigenome-wide methylation profiles of 156 species, including seven eukaryotes such as *Arabidopsis*, *C. elegans*, *Drosophila*, mouse and yeast, as well as 149 prokaryotes. It also offers a genome browser to visualize the methylation sites and related information such as single nucleotide polymorphisms (SNP) and genomic annotation. Furthermore, the database provides a quick summary of statistics of methylome of 6mA and 4mC and predicted methylation motifs for each species. MethSMRT is publicly available at http://sysbio.sysu.edu.cn/methsmrt/ without use restriction.

INTRODUCTION

DNA methylation is an important type of epigenetic modifications, which greatly expands the information content of DNA. The most common types of DNA methylation are 5-methylcytosine (5mC), 6-methyadenine (6mA) and 4-methylcytosine (4mC) (1). In eukaryotes, 5mC is the dominant type, playing an important role in gene regulation, transposon suppression and genomic imprinting (2). Aberrant 5mC patterns have been associated with many diseases and cancers (3). Take retinoblastoma for example, DNA hypermethylation silenced gene expression of RAS-associated domain family 1A in tumor but not in normal tissue (4). In prokaryotes, 6mA and 4mC are the most prevalent DNA modifications that are primarily used for distinguishing host DNA from foreign pathogenic DNA (5). In contrast, 6mA and 4mC are suggested to be minimal and only detectable by highly sensitive technologies in eukaryotes (5). Until recently, several studies reported the epigenome-wide patterns of 6mA in eukaryotes, including *Chlamydomonas*, *C. elegans* and *Drosophila*, showing wide existence of 6mA in eukaryotes and its important functions in regulating gene regulation and development (6–8).

To date, many DNA methylation databases had been constructed, providing invaluable resources for the epigenetic community. MethDB is the first database that stores DNA methylation profiles and associated gene expression information (9). NGSMethDB hosts DNA methylation profiles generated from bisulfite sequencing technique (10). MethBank focuses on methylome changes during embryonic development (11) while MethyCancer and MENT focus on cancers (12,13). PubMeth is another cancer methylation database, based on text-mining of published literature (14). All these databases hosted DNA 5mC profiles and no database provided DNA 6mA or 4mC information so far.

*To whom correspondence should be addressed. Tel: +86 20 87335131; Email: xiezhi@gmail.com
Correspondence may also be addressed to Chuanle Xiao. Tel: +86 20 87335131; Email: xiaochuanle@126.com
†These authors contributed equally to this work as the first authors.

The MetaSUB International Consortium Microbiome (2016) 4:24
DOI 10.1186/s40168-016-0168-z

Microbiome

MEETING REPORT Open Access

The Metagenomics and Metadesign of the Subways and Urban Biomes (MetaSUB) International Consortium inaugural meeting report

The MetaSUB International Consortium

Abstract

The Metagenomics and Metadesign of the Subways and Urban Biomes (MetaSUB) International Consortium is a novel, interdisciplinary initiative comprised of experts across many fields, including genomics, data analysis, engineering, public health, and architecture. The ultimate goal of the MetaSUB Consortium is to improve city utilization and planning through the detection, measurement, and design of metagenomics within urban environments. Although continual measures occur for temperature, air pressure, weather, and human activity, including longitudinal, cross-kingdom ecosystem dynamics can alter and improve the design of cities. The MetaSUB Consortium is aiding these efforts by developing and testing metagenomic methods and standards, including optimized methods for sample collection, DNA/RNA isolation, taxa characterization, and data visualization. The data produced by the consortium can aid city planners, public health officials, and architectural designers. In addition, the study will continue to lead to the discovery of new species, global maps of antimicrobial resistance (AMR) markers, and novel biosynthetic gene clusters (BGCs). Finally, we note that engineered metagenomic ecosystems can help enable more responsive, safer, and quantified cities.

Keywords: Microbiome, Biosynthetic gene clusters, Built environment, Next-generation sequencing, Antimicrobial resistance markers

Introduction

In the past few years, novel work has characterized the microbiota and metagenome of urban environments and transit systems and demonstrated species-specificity to certain areas of a city, "molecular echoes" of environmental events, and even a forensic capacity for geospatial metagenomic data [1–8]. These data are especially helpful for understanding the sites of greatest points of contact between humans and the microbial world within cities, such as their subways or mass-transit systems [1–3, 7]. Indeed, how humans interact with (or acquire) new species of bacteria and other organisms depends on the environment they transit, the types of surfaces they touch, and the physical dynamics of their environment in their city. While a wide variety of methods, protocols, algorithms, and approaches for such large-scale studies are available for researchers, best practices, normalized methods, and ideal taxonomic approaches for global work are still being developed to ensure data quality and the promotion of robust data interpretation [9–12].

Since the majority of the world's population (54 %) currently resides in cities, the use of integrative functional genomic methods to elucidate the molecular dynamics (DNA, RNA, proteins, and small molecules) and ecosystems of cities has potentially large implications for the sustainability, security, safety, and future planning of cities [13]. This includes the concept of "smart cities," which could detect and respond to pathogens, improve water safety and treatment, and track the ever-changing metagenomic complexity of urban environments [14–17]. Indeed, by establishing a baseline genomic profile for a city, it is then possible to create differentials and density maps of organisms relevant for the built environment, such as

* Correspondence: chm2042@med.cornell.edu
Dept. of Physiology and Biophysics, Weill Cornell Medicine, New York, NY 10021, USA

文章连接请见学会网页“会员成果”栏。