

Chromatin and transcriptome-based integrative approaches to profile functional long noncoding RNAs – A computational approach

Akademisk avhandling

Som för avläggande av medicine doktorsexamen vid Sahlgrenska akademien, Göteborgs universitet kommer att offentligens försvaras i hörsal **Carl Kylberg (2320)**, **Medicinaregatan 9**, den **07th december 2020**, klockan **13:00**

av **Santhilal Subhash**

Fakultetsopponent:

Professor Johan Jakobsson
Lund University, Lund

Avhandlingen baseras på följande delarbeten

- I. Ali MM*, Akhade VS*, Kosalai ST*, **Subhash S***, Statello L, Meryet-Figuieri M, Abrahamsson J, Mondal T, Kanduri C. PAN-cancer analysis of S-phase enriched lncRNAs identifies oncogenic drivers and biomarkers. *Nature communications*. 2018;9(1):1-20. doi: 10.1038/s41467-018-03265-1. (**Co-first authors*) [**Paper I**]
- II. **Subhash S***, Mishra K*, Akhade VS, Kanduri M, Mondal T, Kanduri C. H3K4me2 and WDR5 enriched chromatin interacting long non-coding RNAs maintain transcriptionally competent chromatin at divergent transcriptional units. *Nucleic acids research*. 2018;46(18):9384-400. doi: 10.1093/nar/gky635. (**Co-first authors*) [**Paper II**]
- III. **Subhash S**, Kanduri M, Kanduri C. Sperm Originated Chromatin Imprints and LincRNAs in Organismal Development and Cancer. *iScience*. 2020;23(6):101165. doi: 10.1016/j.isci.2020.101165. [**Paper III**]

SAHLGRENKA AKADEMIN
INSTITUTIONEN FÖR BIOMEDICINE



Chromatin and transcriptome-based integrative approaches to profile functional long noncoding RNAs – A computational approach

Santhilal Subhash

Department of biochemistry and cell biology, Institute of biomedicine, Sahlgrenska akademien, Göteborgs universitet, Sverige, 20.

Abstract

One of the major hallmarks of cancer is aberrant or uncontrollable cell division, which occurs due to a defective cell cycle process. During the synthesis phase (S-phase) of the cell cycle, before cell division or mitosis phase, the DNA in the cell makes a new copy to pass on genetic information to the daughter cells. Therefore, S-phase is one of the crucial steps for a successful cell division to occur. The DNA in the nucleus is wrapped around a set of proteins called histones, forming nucleosomes, and multiple nucleosomes give rise to the higher-order chromatin structure. This well-established chromatin structure determines which portion of DNA or gene gets activated or suppressed by switching between open or closed chromatin states. Tri- or di-methylation of lysine 4 from histone 3 (H3K4me2/3) leads to open chromatin, which in turn promotes active gene transcription. The product of gene transcription is either protein-coding mRNA that translates into protein for its function or noncoding RNA, which do not code for any protein and function as RNA. However, the human genome project has identified that protein-coding genes only constitute 2% of the genome, and the vast majority of it is noncoding. Unlike protein-coding genes, the significance of RNAs transcribed from the noncoding genome is not well-established. Apart from housekeeping noncoding RNAs (rRNA, tRNA, snRNA, and snoRNA) and microRNAs (miRNAs), most functional noncoding RNAs fall into the long noncoding RNA (lncRNA) category. In this thesis, we implemented comprehensive computational approaches to identify functionally relevant lncRNAs by analyzing chromatin and transcriptome-based sequencing datasets. In the first study (**paper I**), using a transcriptome approach, we profiled lncRNAs associated with the S-phase stage of the cell cycle. We demonstrated the oncogenic properties of various S-phase associated lncRNAs in multiple cancers. Earlier, studies proposed that chromatin-associated RNAs, with the help of chromatin-modifying enzymes, determines the active/open or close chromatin status to promote or suppress gene transcription. Hence, in the second study (**paper II**), we used chromatin-based approaches to propose a possible mechanism through which the active chromatin-associated lncRNAs may function. We show that active chromatin-associated lncRNAs regulate their partner genes in-cis by recruiting the WDR5 chromatin modifier to establish an open chromatin structure at the partner protein-coding gene promoters. In our third study (**paper III**), we integrated both transcriptome and chromatin-based approaches to find early development-associated lncRNAs. Here, we focused on tracing the molecular footprints of sperm lncRNAs throughout the stages of organismal development. For this purpose, we integrated datasets from gametes, preimplantation and post-implantation stages of an embryo. Interestingly, we observed distinct chromatin structures in the sperm. Also, sperm lncRNAs were active during the onset of zygotic genome activation in the preimplantation stages and in cancers. In summary, this study reveals a unique set of sperm-specific lncRNAs that are temporally activated during preimplantation stages and also aberrantly expressed in multiple cancers. Overall, the present thesis provides an extensive catalogue of functionally relevant lncRNAs that can take part in cell cycle regulation, cancer, chromatin modulation, and organism development. Our studies can serve as a comprehensive resource for future investigations on lncRNAs.

Keywords: Computational biology, Long noncoding RNAs, Chromatin, Histone, Cell Cycle, S-phase, Cancer, Epigenetics, ChIP-seq, RNA-seq, ChRIP-seq, WDR5, H3K4me2, DNA methylation, Histone modifications.

ISBN: 978-91-8009-062-9 (TRYCK)

<http://hdl.handle.net/2077/66195>

ISBN: 978-91-8009-063-6 (PDF)