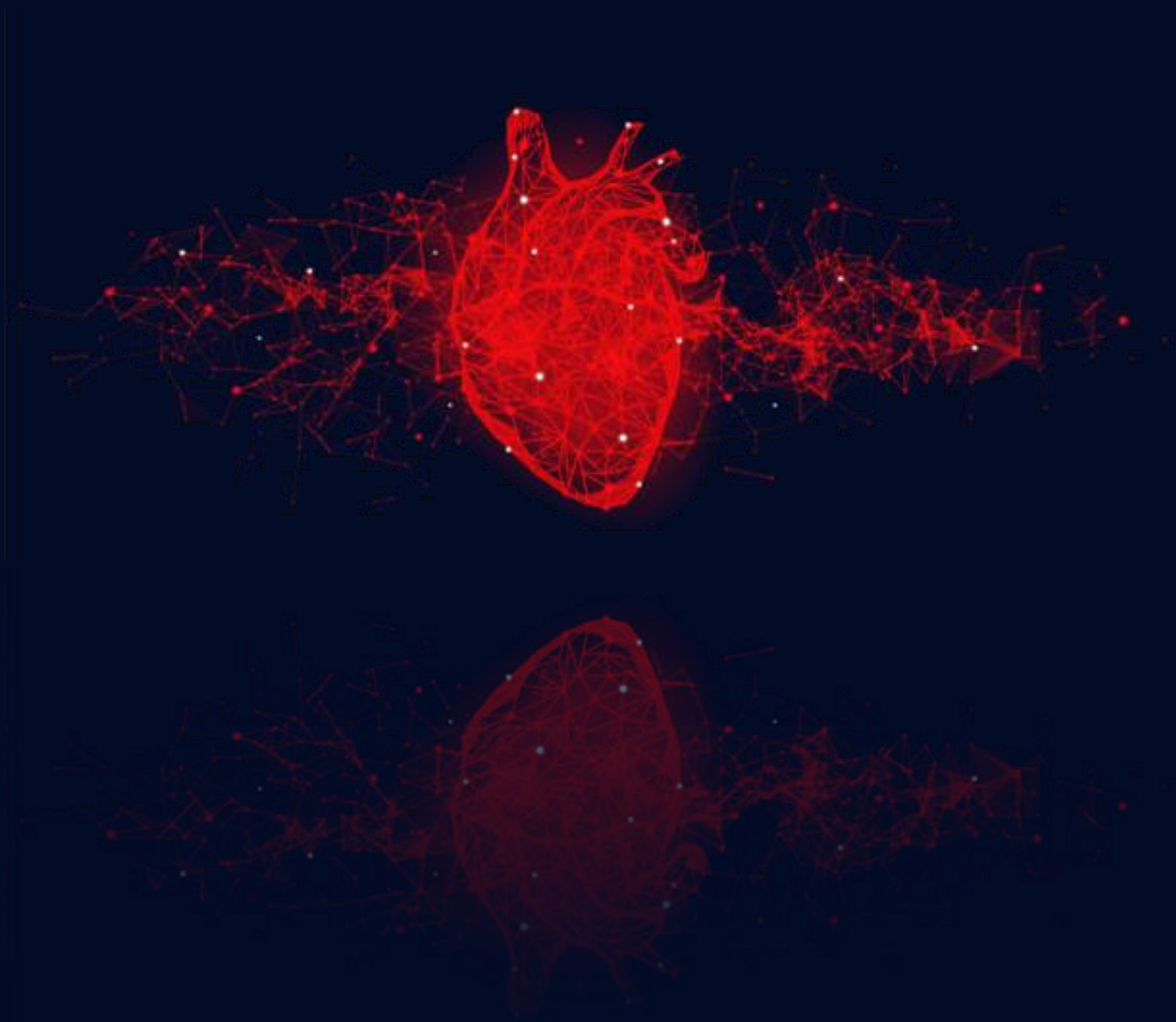
Chapter 1

In-silico identification of Clock associated miRNAs



Introduction

Evolving under the selective pressure of 24h world, circadian time keeping is present apparently in all the living organisms. Circadian rhythms are indicative of organismal coordination with the external environment. Physiological processes including metabolism, digestion, circulation, respiration etc. exhibit circadian rhythmicity. This circadian entrainment is maintained and orchestrated by a set of genes referred to as 'core clock genes', comprising of CLOCK, BMAL-1, PER (Period genes) and CRY (Cryptochrome genes) genes (Yan *et al.*, 2008). These genes function in an autoregulatory feedback loops with several transcription – translation processes at their disposition. A myriad of downstream genes is subsequently controlled under circadian fashion and referred to as clock-controlled genes. Cumulative functioning of these set of genes, successfully operates and entrain the tissue level circadian clock (peripheral circadian clock) with the central clock regulator SCN. Along with maintaining the peripheral circadian rhythms these genes hold ascendency over several physiological functions in the body. Circadian rhythms have been studied since decades. However, advancements in current research unveils a more complex operation and regulation of circadian rhythms. This involves post-transcriptional, translational and post-translational regulations as well as novel regulatory candidates that holds fine control over circadian rhythm at physiological levels (Kojima *et al.*, 2011). There have been several reports of miRNA mediated regulation of circadian clock. A new paradigm of gene expression regulation has emerged with the discovery of miRNAs. Majority of miRNAs are thought to control gene expression, mostly by base pairing with miRNA-recognition elements or seed sequences found in their mRNA targets. Although a large number of human miRNAs have been reported, many of their mRNA targets remain unknown.

miRNAs are small 20-22 nucleotide non-coding RNA molecules that majorly function by binding to 3' UTR of mRNA and induce its degradation or repress its translational process. miRNAs are processed by RNA pol II and further synthesized via canonical or non-canonical pathways as mentioned earlier. A fully mature miRNA identified as -5p miRNA or -3p miRNA is a functionally active form. Recent studies have shown that several miRNAs also exhibit circadian rhythmicity. Yang *et al* (2008) performed a microarray to find that about 78 miRNAs showed circadian entrainment in drosophila head (Yang *et al.*, 2008). Xu *et al* (2007) showed that 12 miRNAs, solely expressed in retina, exhibited diurnal expression pattern using murine model (Xu *et al.*, 2007). In another study Na *et al* (2009) identifies 85 murine hepatic miRNAs showing circadian patterns (Na *et al.*, 2009). Further the authors defined circadian miRNA–mRNA target pair wherein both elements showed rhythmic expression and for which the sequence-based target relationship was established computationally. Based on these criteria, 33 significant miRNA–mRNA target pairs were identified between a total of 24 miRNAs and 10 core clock genes (Na *et al.*, 2009). Apart from these, there are several studies showing miRNA-based regulation of circadian clock genes and their physiological implications. Chen *et al* (2013) had shown that the period of the circadian clock shortens in Dicer mutant cells as well as mice, pertaining to PER1 and PER2 cytoplasmic accumulation under influence of three different miRNAs (Chen *et al.*, 2013).

Computational methods play important roles in the identification of new miRNAs. Conventionally, some major characteristics such as the hairpin-shaped stem loop structure, high minimal folding free energy, and high evolutionary conservation are important features used in the computational identification of miRNAs. The current state-of-the-art methods focus mainly on three major categories: sequence or structure

conservation-based approaches, machine learning-based approaches, and experimental data-driven approaches.

Recently, these computational methods are artificially classified into two different generations. The first-generation methods like miRanda (Enright *et al.*, 2003), DIANA-microT (Kiriakidou *et al.*, 2004), RNAhybrid (Krüger and Rehmsmeier, 2006), MicroInspector (Rusinov *et al.*, 2005), and TargetScan and TargetScans (Lewis *et al.*, 2003; Lewis *et al.*, 2005) and miRDB (Wong and Wang, 2015) are based mainly on three characteristic properties:

(1) The 5' seed sequence of the miRNA (nucleotide positions 2–8 of the miRNA) is complementary to the 3' UTR of the target mRNA.

(2) The RNA–RNA duplex has a higher negative folding free energy.

(3) Mature miRNAs, binding sites of miRNA to mRNA, and miRNA:mRNA duplex all are highly conserved from species to species, predominantly within the same kingdom.

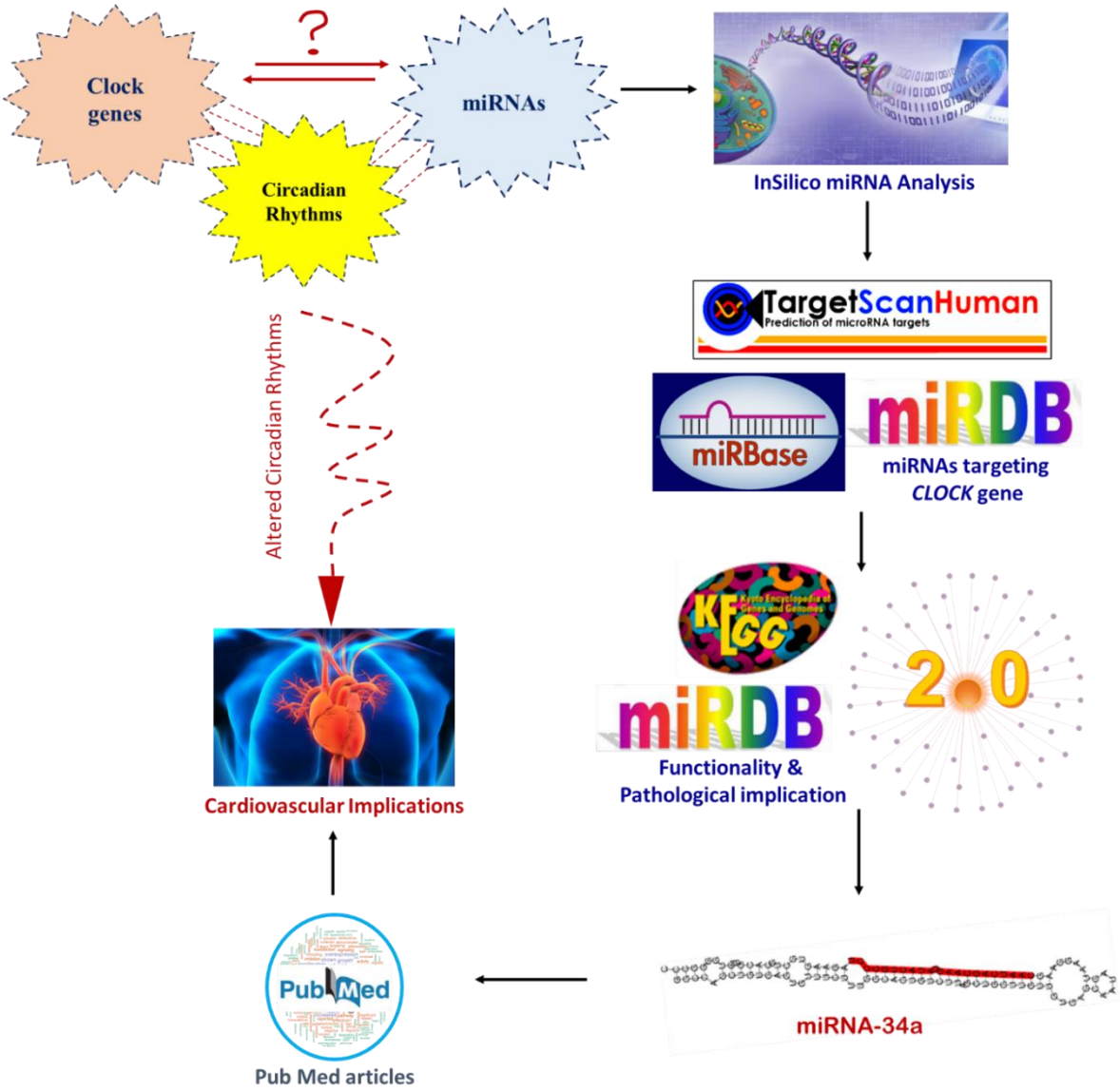
Whereas the new generation methods operate upon machine learning-based approaches viz. PicTar (Krek *et al.*, 2005). In this format the algorithm scans alignments of 3' UTRs for those displaying seed matches to miRNAs, then filters the alignments according to their thermodynamic stability. Each predicted target is scored by using HMM maximum-likelihood fit approach. wherein, synergistic effects of multiple binding sites of one miRNA or several miRNAs acting together are accounted for along with appropriate scoring of overlapping site and background for binding. The probabilities are assigned according to experimental and computational results.

This study exploits such computational databases, definite algorithmic structures and focuses on ploughing through miRNAs associated to Clock gene and their plausible

implications in cardiovascular disorders, generating prima face evidence for potential link between chronodisruption induced cardiovascular disorder.

Methodology

In-silico Experimentations



Results

Algorithmic exploration of miRNAs exhibiting complementarity to the 3'UTR seed sequence of human and murine *Clock* gene.

Potential mRNA-miRNA interactions for Clock gene were assessed using computational databases miRDB V 6.0 and TargetScan V 7.0 that uses miRTarget and The Basic Seed-based algorithms respectively. Characteristic traits incorporated while developing mRNA-miRNA pairing included target site, distance to UTR, terminal A-T/G-C content, GC count upstream/downstream of the seed sequence etc. Herein, we identified a total of 628 and 397 miRNAs showing potential complementarity to 3'UTR of *Clock* gene in human and mice respectively (Fig 1.1a & b). Further, all the miRNAs that showed target score above 50 were included for the analysis. An overlapping screening of common miRNA targets shared between human and mice was done for further evaluation (Fig 1.2) The resultant 86 miRNAs were subjected to FuncMir test, an experimental evaluation-based database developed in miRDB software, wherein miRNAs are reported based on their pathological/physiological implications. Several enlisted miRNAs viz. miR30a-5p, miR106a, miR33, miR34a, miR15b, miR20a-5p etc. were identified to have a role in cardiac arrest and other cardiovascular disorders. Based on the miRNA conservation across species, gene targets and their cardiovascular implications, miR34a was carried further for detailed investigations in regard to the study.

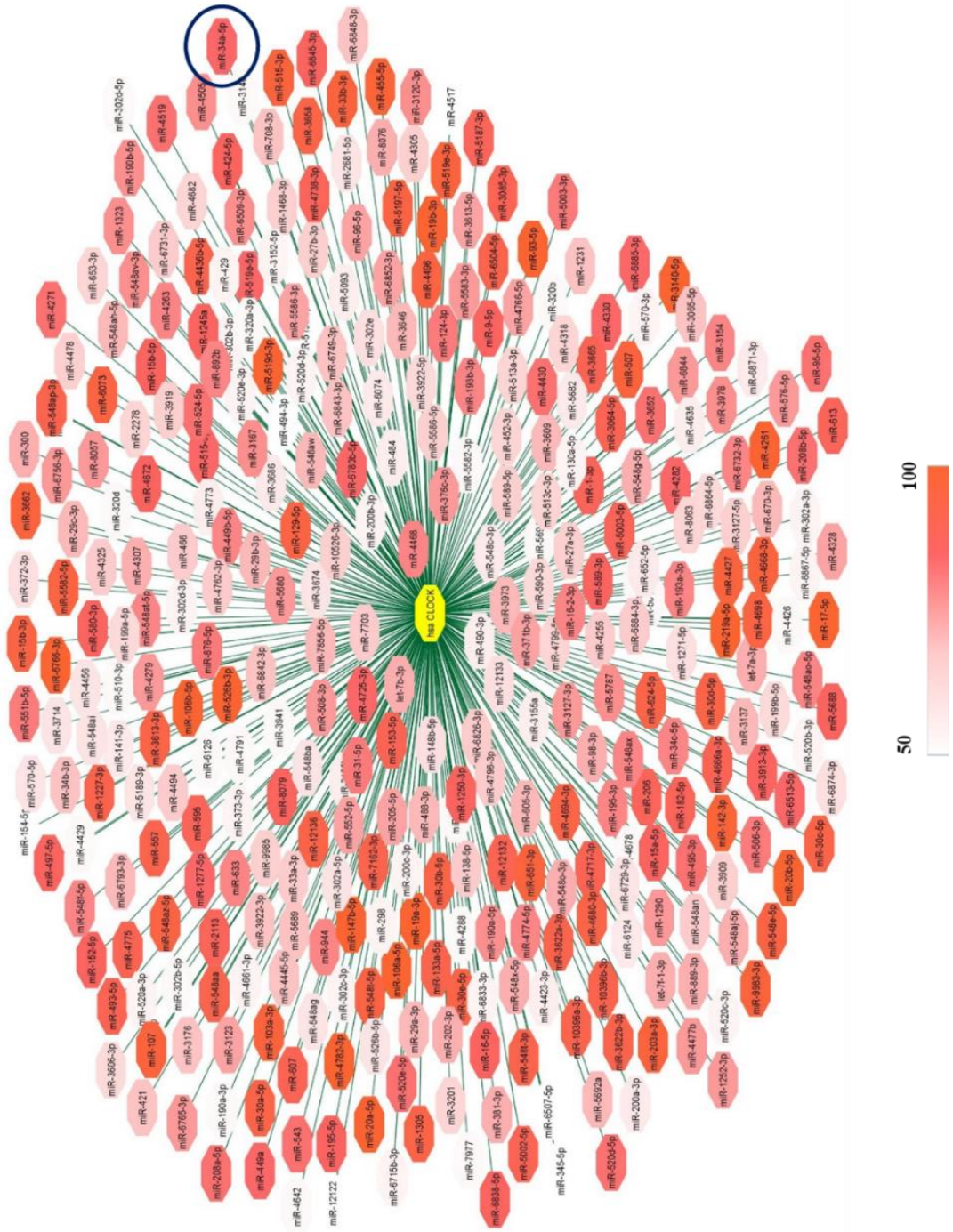
Evaluating gene targets and pathological implications of miR34a

miR34a matures into miR34a-5p and miR34a-3p as a functional miRNA. Both the forms were seeded as the miRNA query in the Target search engine for identification of the gene targets with complementary seed sequence. miR-34a-3p displayed a total

of 371 gene targets in miRDB database whereas a total of 899 gene targets were obtained with query of miR34a-5p. miR34a-3p did not show complementary seed sequence for CLOCK gene and hence was not carried forward in investigation. miR34a-5p targets showed several key genes vital for physiological functions as well as active in various pathological conditions. These targets included genes like *CLOCK, PER2, RORA, KLF4, SIRT1, BCL2, FOXO1, CDK6, IL6R, ADIPOR2, AGO4, XPO5, NOTCH2, NOTCH1, COL26A1, ACE, UCP3* etc., vouching for its vital circadian function and plausible atherogenic implications (Fig. 1.3a).

With that miR 34a-5p was assessed for its stability across the evolution. The conserved sequence across the species for miR34a-5p was deduced from NCBI repository database, wherein it showed conservation throughout the vertebrates (Fig. 1.3b). miR34a-5p showed two seed sequences in the 3'UTR of *CLOCK* gene of human, whereas single seed sequence was found in mice and rat (Table 1.1). That was further subjected to *Function Enrichment* operation using Hypogeometric Test and KEGG database wherein miR34a showed implications in about 44 diseases established experimentally and 159 pathological conditions recorded from the Human microRNA Disease Database (Fig. 1.4a). miR34a-5p target genes were further evaluated for its established functional implications in CVDs using PubMed Literature survey wherein key words "*Gene name* + cardiovascular disorder/atherosclerosis" were utilized. Herein we found that about 39.9% of the target genes exhibited functional role in Cardiovascular disorders (Fig. 1.4b). Contemplating over the attained results, miR34a was sought as a way forward with a detailed scrutiny done on its mature form miR34a-5p.

(a)



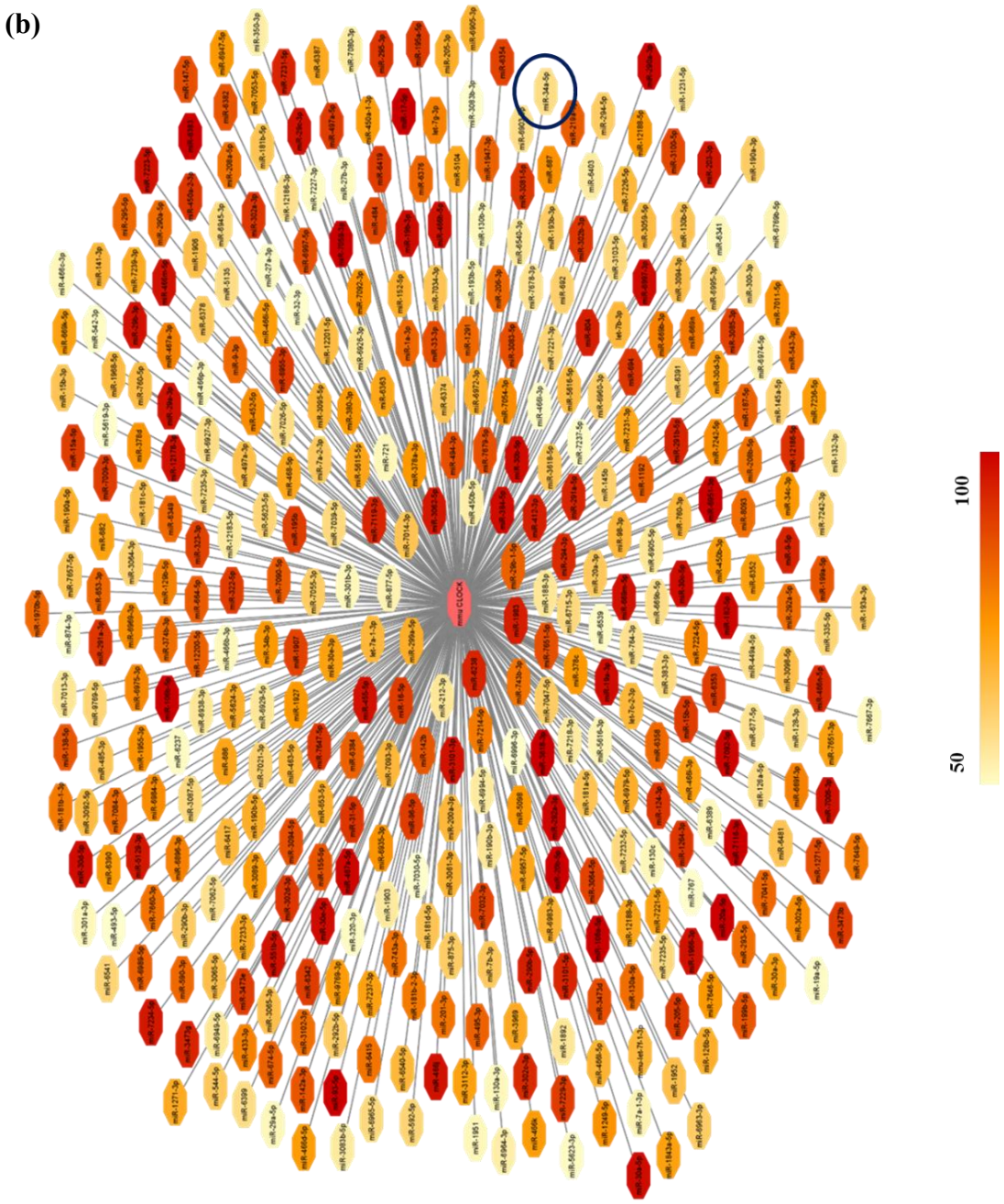


Fig. 1.1: miRNA complementary to 3'UTR seed sequence of Clock gene in human and mice. Computational algorithms miRDB and TargetScan were used to plough the miRNAs complementing the seed sequence in the 3'UTR of Clock gene in (a) human and (b) mice, respectively. Resultant miRNAs have been plotted using Cytoscape with color gradation in accordance with the dock score generated by the system, represented as scale below. Encircled blue tab shows miR34a-5p.

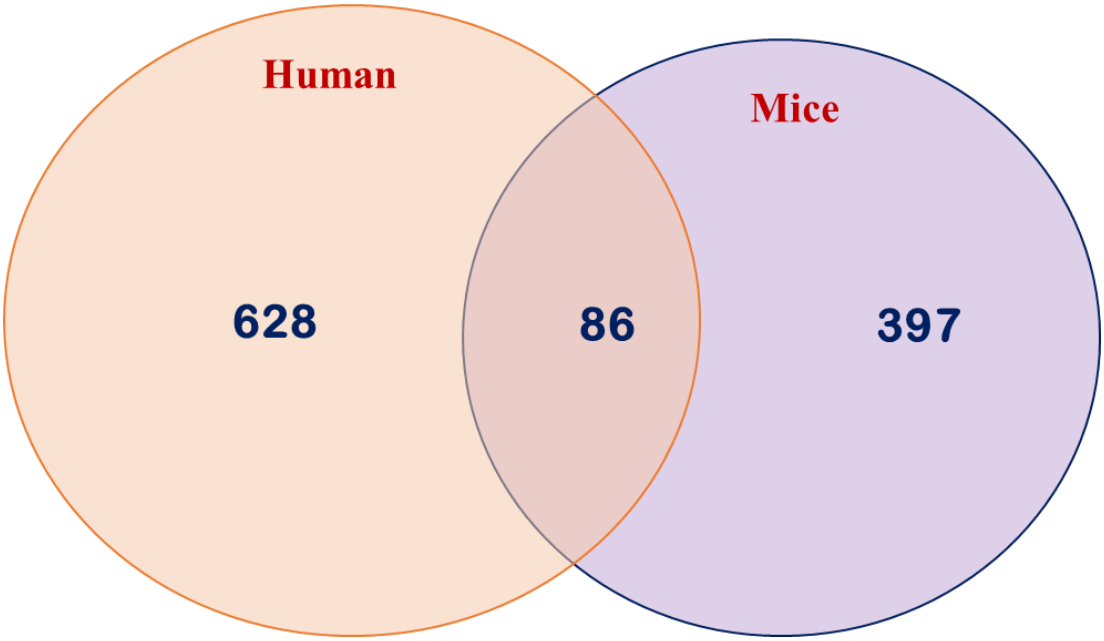
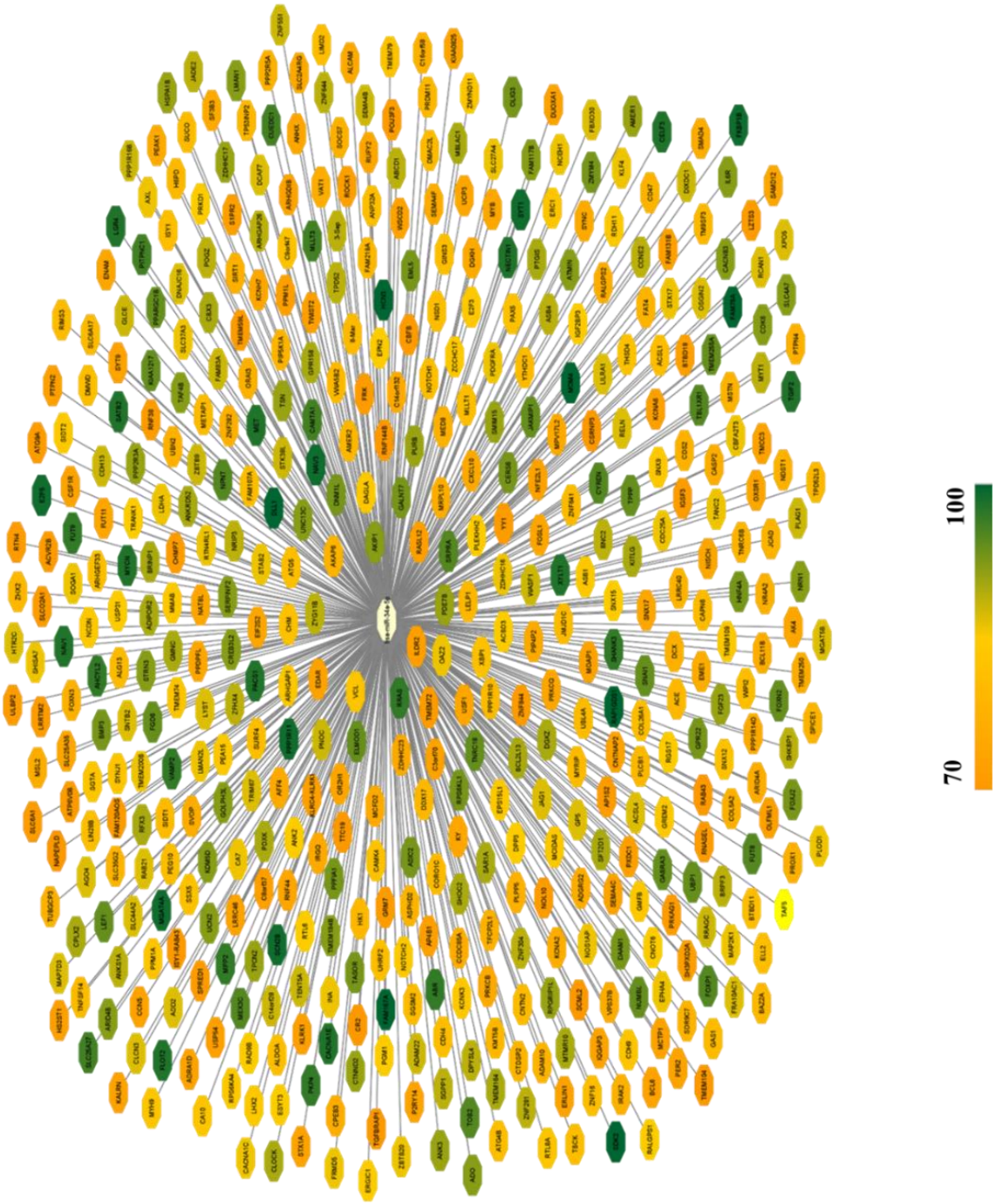


Fig. 1.2: Venn diagram representing number of unique and common Clock associated miRNAs shared between human and mice.

(a)



(b)

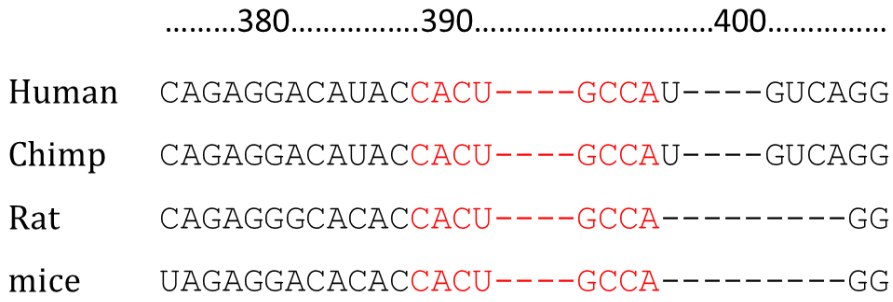


Fig. 1.3: potential gene targets and evolutionary conservation of miR34a-5p. Computational algorithms miRDB and TargetScan were used to plough the gene targets of miR34a-5p and the same have been plotted using Cytoscape with color gradation in accordance with the dock score generated by the system, represented as scale below. miRNA conserved sequence across the species was identified using NCBI database and represented here schematically.

Sr. No.	Species	Gene	Total no. of miRNAs docking to clock	miRNA	Dock Score	Seed location	Length of promoter region
1.	*Homo sapiens*	CLOCK	628	hsa-miR 34a-5p	90	389, 5954	7514 nt
2.	*Mus musculus*	CLOCK	397	mmu-miR 34a-5p	58	369	6845 nt
3.	*Rattus norvegicus*	CLOCK	163	rno-miR 34a-5p	64	363	6702 nt

Table 1.1: Information on miR34a-5p – CLOCK docking was studied using computational algorithms in human, mice, and rats. The dock score based on miRNA mRNA interactions, seed locations in the 3'UTR and the length of 3'UTR has been documented in the above table.

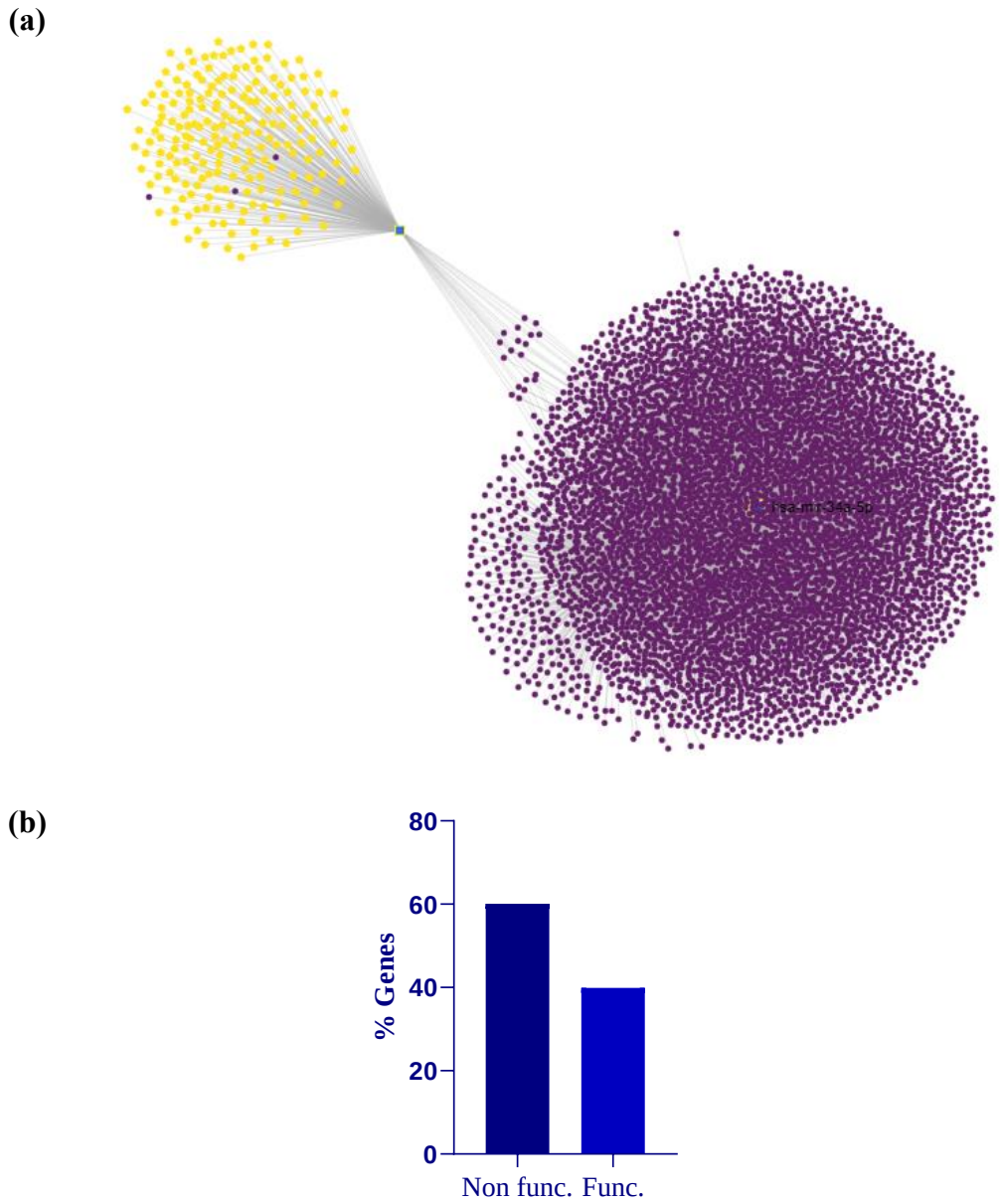


Fig. 1.4: Pathological association of miR34a-5p. (a) Topological representation of pathological association of miR34a-5p as assessed using KEGG database in miRNet; yellow represents the pathologies and purple represent the underlying gene ontology. (b) Graphical representation of miR34a-5p target-gene functional in cardiovascular disorders as identified in PubMed literature screening.

Discussion

Circadian misalignment has increasingly gained prominence as one of the key causative agents for onset and progression of several lifestyle disorders. Numerous genes, proteins and miRNAs have been documented to be responsive towards circadian rhythms of an organism. The present study primarily attempts to investigate circadian clock associated miRNAs with plausible implications in cardiovascular disorders. Exploring an unidentified target in cellular system is analogous to finding a needle in haystack. In recent years we have advanced in exploring and exploiting high throughput experimental techniques that broadly include NGS (next generation sequencing) or micro array analysis. Though these techniques are extensively used and are of great utility, they are cost expensive and preset us with humongous data that further needs to be sorted with dedicated software and compatibly developed algorithms. Along with this widely used practice, we employed another scheme of series of experiments that included target-fishing primarily at the computational level. Computational science has advanced extensively over the years wherein all the molecular docking including permutation combinations of miRNA, genes, mRNAs, proteins, DNA, RNA, lncRNA etc have been made possible with sophisticated programs and algorithms that not only take into consideration the physico-chemical properties but also physiochemical and biological properties of the target moiety.

Last decade has witnessed extensive work on miRNAs and that has obtained them due recognition pertaining to their biological role and capabilities. Several databases have been generated for miRNAs (Dweep *et al.*, 2014; Panwar *et al.*, 2017; Liu *et al.*, 2019) that have paved way to development of miRNA target predicting software which uses the biological properties of miRNA interactions to predict the targeted mRNA. These

algorithms have been worked upon lengthily to develop a sophisticated version that can identify miRNAs and/or mRNAs that are yet to be reported in biological systems.

Herein, the study commences miRNA identification with computational algorithms for miRNA – mRNA interactions, miRDB and TargetScan. These interactions deploy under defined physiological conditions including miRNA interaction to 3'UTR of mRNA exhibiting Watson & Crick base pair complementary seed sequence that is further followed with conditions like miRNA/mRNA concentrations, free energy of binding (-ΔG), cellular compartmentalization/location, cell type and competitive inhibitions.

Several miRNAs including miR30a-5p, miR106a, miR33, miR34a, miR15b, miR20a-5p were subjected to FuncMir test with primary objective of identifying their underlying cardiovascular implications. miR30a-5p showed several implications in cancer, Alzheimer disease, phenotype regulation and autophagy in VSMCs by myocardia regulation (Cai *et al.*, 2021; Shi *et al.*, 2022; Sun *et al.*, 2022; Zhang *et al.*, 2022). Apart from functional regulation in cancers, miR106a showed regulation in VSMCs by regulating miR-106a-5p/Foxo1/Vcam1 pathway and in patients with idiopathic neuropathy by regulating PTEN (Wang *et al.*, 2021; Wu *et al.*, 2021). miR15b was identified to regulate TGFβ1 induced proliferation in hepatic stellate cells, ox-LDL induced endothelial injury and growth arrest in relation to circCHFR downregulation (Ma *et al.*, 2021; Li and Wang, 2022). However, miR34a was identified to regulate macrophage cholesterol flux, suppressing apoptosis in endothelial cells by regulating BCL2, promoting VSMC proliferation and migration. miR34a was also functional in cancerous pathologies along with NASH, diabetic cardiomyopathy, and obesity. Moreover, miR34a was also reported to regulate PER2 and PER1 genes and hold circadian control. miR34a-5p was assessed further with KEGG database for its gene targets and consequential pathological implications to find its role in over hundreds of

pathogenic conditions including fatty liver, cardiovascular disorders, cancers, Parkinson's disease, Alzheimer's disease, vasculitis, obesity, heart failure, atherosclerosis, diabetes mellitus, rheumatoid arthritis, etc. Detailed scrutiny with PubMed literature unveiled functional role of about 39.9% of the target genes of miR34a-5p in cardiovascular disorders making it a potential candidate linking chronodisruption to onset and progression of cardiovascular disorders. Thereby, we decided to primarily validate the chronobiological role of miR34a-5p in invitro and in-vivo model systems followed by investigating in atherogenic implications in forthcoming chapters of this compile of literature.