Decoding the Functional Roles of Intronic microRNA miR-2355 and Its Host Gene *KLF7* in Immunobiology of Cervical Cancer

By

Nure Sharaf Nower Samia
17136032

A thesis submitted to the Department of Mathematics and Natural Sciences in partial fulfillment of the requirements for the degree of
Bachelor of Science in Biotechnology

Department of Mathematics and Natural Sciences
Brac University
June 2021

Declaration

It is hereby declared that

1. The thesis submitted is my/our own original work while completing degree at Brac University.
2. The thesis does not contain material previously published or written by a third party, except where this is appropriately cited through full and accurate referencing.
3. The thesis does not contain material which has been accepted, or submitted, for any other degree or diploma at a university or other institution.
4. I have acknowledged all main sources of help.

Student's Full Name & Signature:

Nure Sharaf Nower Samia

Nure Sharaf Nower Samia
Student ID: 17136032

Approval

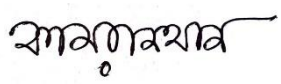
The thesis/project titled "**Decoding the Functional Roles of Intronic microRNA miR-2355 and Its Host Gene *KLF7* in Immunobiology of Cervical Cancer**" submitted by

1. Nure Sharaf Nower Samia (Student ID: 17136032)

of Spring, 2017 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of Bachelor of Science in Biotechnology on June 2021.

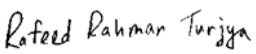
Examining Committee:

Supervisor:
(Member)



Md. Abdullah-Al-Kamran Khan
Lecturer, Department of Mathematics and Natural Sciences
Brac University

Co-Supervisor:
(Member)



Rafeed Rahman Turjya
Lecturer, Department of Mathematics and Natural Sciences
Brac University

Program Coordinator:
(Member)

Iftekhar Bin Naser, PhD
Assistant Professor, Department of Mathematics and Natural Sciences
Brac University

Departmental Head:
(Chair)

A F M Yusuf Haider, PhD
Professor and Chairperson, Department of Mathematics and Natural Sciences
Brac University

Dedication

I would like to dedicate my parents, my teachers and my friends for their undying love and support without whom I wouldn’t have been able to make it this far.

Acknowledgement

I would like to commence after expressing my earnest gratitude to the Almighty for endowing me with the opportunity of this research course and then providing in me the boldness needed throughout this journey to fulfil it successfully.

I acknowledge my gratitude to Professor **A F M Yusuf Haider**, Chairperson of the Mathematics and Natural Sciences Department of BRAC University for allowing me to complete my undergraduate thesis.

I express my sincere gratitude and respect to my supervisors, **Mr. Md. Abdullah-Al-Kamran Khan** and **Mr. Rafeed Rahman Turjya** who have constantly walked me through each and every step of this study and inspired in me to bring out my best in the most enjoyable way possible.

I also convey my utmost gratefulness to **Dr. Iftekhar Bin Naser** (Assistant Professor and coordinator of Biotechnology program, Department of Mathematics and Natural Sciences) for his kind words and continuous guidance, patience, motivation, enthusiasm, scholarly inputs to keep me in the right track along the way.

My deepest regards and gratitude to my friend, **Shazneen Hossain** for being with me throughout the study and for her endless help and also **Sumya Nishat, Sheikh Shafin Ahmed, Mehnaz Tabassum** for always encouraging me in all means. Without them it would be really difficult to survive the current situation.

Lastly and most importantly, I would like to declare my gratitude to my family for their endless support and encouragement in every aspect of my life.

Nure Sharaf Nower Samia

ID: 17136032

Abstract

Intronic miRNAs are human miRNA genes that sit within the intronic regions of protein-coding host genes and influence their hosts' function either antagonistically or synergistically. However, previous research has shown that the co-expression of intronic miRNAs and their related host genes is required to keep the biological system resilient and to elucidate their functional relationship. In this study, through processing the raw gene expression data of the whole transcriptome sequencing in human cervical adenocarcinoma and normal cervical tissues from the GEO database (GSE145372), we identified the deregulated miRs and deregulated genes in cervical cancer patients. Among them we found 22 DE intronic miRNAs along with their associated DE host genes. We selected the DE intronic miRNA *hsa-mir-2355* and its DE transcription factor-encoding host gene *KLF7* as they are associated with prognosis in cervical cancer patients. Besides, the selected gene pair play roles in the several immune signaling pathways such as IGF1 pathway, IFN-gamma pathway, TRAIL signalling pathway, and S1P1 pathway etc. Furthermore, we identified a total of 227 common targets of *KLF7* and hsa-mir-2355, and among them 25 genes were significantly associated with cervical cancer patient survival. Then to shed insights onto the oncoimmunology, we further selected four genes namely *AGT12, KRAS, PRKAR1* and *REL* that can play synergetic or antagonistic role in the immunology. While searching for the correlation between the expression of these genes with immune infiltration in cervical cancer, we observed that expression of *REL* has significant association of recruiting CD4+ T-cells in the tumor microenvironment which evidently supports the prevalence of observed CD4+ T-cells amongst the tumor cells in the analysed expression data of the cervical cancer patients. All these results imply that intronic microRNAs have a regulatory influence on the transcriptor factor encoding

host genes, thereby promoting cancer spreading or assisting in tumor destruction in the tumor microenvironment.The result identified through these approaches might be used as potential candidates for cancer diagnosis and therapy.

Keywords: MiRNAs, Intronic miRNAs, Transcription factor, Immune infiltration, Cervical cancer.

Table of Contents

Table of figures

List of Acronyms

CESC	Cervical Squamous Cell Carcinoma
DE	Differentially Expressed
GO	Gene Ontology
GOBP	Gene Ontology Biological Process
CHG	Cancer Hallmark Gene
GSEA	Gene Set Enrichment Analysis
KEGG	Kyoto Encyclopedia of Gene and Genomes

Chapter 1

Introduction

In recent years, cancer has been a significant cause of death for humans. Cervical cancer is the fourth most common form of cancer in women. It mostly affects the cells of the cervix, which is the lower portion of the uterus that leads to the vagina. The majority of cases arise in less developed countries where adequate screening programs are not available [1]. Cervical cancer screening, based on indirect data, could minimize the incidence and mortality of invasive cervical cancer by approximately 90% [2]. In recent years, there has been a surge of interest among researchers in the function of microRNA (miRNA) in both normal and disease processes. Due to the difficulty of identifying genuinely expressed miRNAs from random hairpin secondary structure sequences in genomic DNA, it is advantageous to predict them using expressed sequences such as expressed sequence tags (ESTs) and intronic sequences [3]. Upon analysis, it is discovered that 25% of known human miRNAs reside in intronic regions of known protein-coding genes, with approximately 50% of these intronic miRNAs located in introns longer than 5,000 base pairs. It is possible that these intronic miRNAs have their own independently regulated transcription units that are regulated by either RNA polymerase II (Pol II) or RNA polymerase III (Pol III) (Pol III) [4].

MicroRNAs (miRNAs) are tiny non-coding RNAs that play a significant role in the formation and progression of a variety of cancer forms by controlling the expression of their targets. They are small non-coding regions of 20–22 nucleotides in RNAs that are involved in any biological pathway in multicellular organisms, including humans[5]. Numerous studies suggest that more than half of miRNA genes are found in regions

of the genome associated with cancer or in vulnerable sites[6]. While the vast majority of published research focuses on a single mRNA target, the majority of miRNAs exert their effects by targeting several mRNAs, some of which may be located in the same cellular pathway [5].

On the other hand intronic miRNAs are those which are capable of interfering with the expression of other genes. About 82 percent of miRNA genes are located in intronic areas. But for the 6% of mixed miRNAs, the percentages of intronic and exonic miRNAs are 87% and 13%, respectively [7]. Intronic miRNAs are thought to be processed from their host transcription units' introns, and thus share regulatory mechanisms and expression patterns with their host gene [8]. However, intronic miRNAs and their host genes may regulate independently, the intronic miRNA may still be able to down-regulate its own host protein-coding gene by targeting the host gene's untranslated region (UTR) [4]. It is possible that these intronic miRNAs are transcribed from distinct transcription units with separate promoters from their host genes [8]. Additionally, intronic miRNAs can play a critical role as negative feedback regulators [4].

MicroRNAs have the ability to control a wide range of biological and pathological mechanisms such as cell cycle, differentiation, proliferation, apoptosis, stress tolerance, energy metabolism, and immune response, including the formation and development of cancer [9]. It has been reported that miRNAs regulate about 30% of human genes and half of these genes are tumor related or found in fragile locations where some act as tumor genes and others as tumor suppressor genes interestingly some may act as both [9].Additionally, epigenetic therapy stimulates miRNAs not only directly from their own promoters, but also intronic miRNAs in conjunction with their host genes, indicating epigenetic therapy's anticancer impact [10]. We intended to

determine whether intronic RNAs have a role in oncogenesis in this research work using RNA sequencing data from GSE145372 and TCGA CESC. Essentially, we want to underline that because microRNAs are intronic in nature, they have an influence on their target genes, which results in oncogenesis. The role of miRNA in cervical cancers are observed where miR-200a and miR-9 are seen to be the vital regulators of cervical cancer control as they alter the metastatic potential of cervical cancer cells by co-suppressing numerous genes involved in cell motility [11].

Dendritic cells are crucial for eliciting innate and adaptive cellular immune responses against HPV-associated cervical cancers, which can be avoided or treated by inducing virus-specific immune responses in patients [12]. Cytokines produced by immune system cells, as well as tumor cells, appear to act by influencing the host immune system, resulting in an immunosuppressive response rather than an efficient protective immunological response, hence contributing to the tumor's development and development [13]. The purpose of this study was to determine whether intronic microRNAs have a role in the tumorigenic pathway by altering their host gene targets. Additionally, whether those intronic microRNAs are abundantly expressed in practically all samples and are controlled in such a way that they are connected with the overall survival of CESC patients or not. Subsequently, to delineate the relationship between the common targets of intronic miRNA and transcription factor-encoded host genes and immune infiltrating cells in cervical cancers in order to gain a better understanding of the extent of immune infiltration and whether these genes are guiding cancer metastasis or assisting in tumor destruction in the tumor microenvironment.

1.1 miRNAs

The miRNA or microRNA is a short non-coding RNA molecule with approximately 22 nucleotides. The primary function of this enzyme is to silence RNA and to regulate gene expression post-transcriptionally [14], which they accomplish by base-pairing with complementary sequences contained within messenger RNA molecules [15]. As a result, these mRNA molecules are silenced via strand cleavage; instability of the mRNA via shortening its poly(A) tail; and less efficient translation of the mRNA into proteins by ribosomes occur [15, 16]. They are structurally similar to small interfering RNAs (siRNAs) in the RNA interference (RNAi) pathway, except that miRNAs originate from portions of RNA transcripts that fold back on themselves to form short hairpins, whereas siRNAs originate from longer double-stranded RNA sections [17]. The miRNA strand's directionality defines the identification of the mature miRNA form. The 5p strand derives from the pre-miRNA hairpin's 5′ end, while the 3p strand comes from the 3′ end [18].

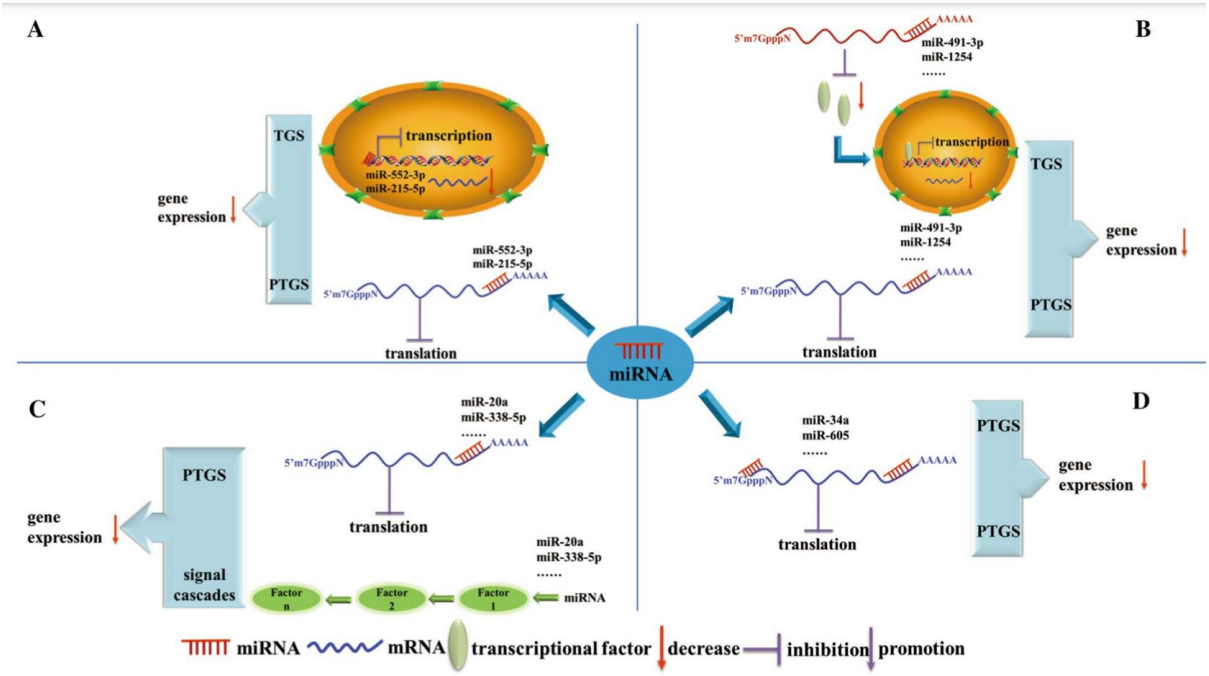


***Figure 1**: Different forms of dual regulation by miRNA on a target gene. Adapted from [19].*

Over 1917 annotated hairpin precursors and 2654 mature sequences may be found in the human genome [20]. They are abundant as extracellular circulating miRNAs in a variety of mammalian cell types [21] which are released into bodily fluids such as blood and cerebrospinal fluid and may serve as biomarkers for a variety of disorders. Numerous miRNAs are evolutionary conserved, implying that they perform crucial biological activities [22]. Enhancing the precision of biogenesis, for example, to maximize targeting specificity [14]. Oncogenic miRNAs and oncosuppressor miRNAs are two classes of dysregulated miRNAs. Both categories of miRNAs have been shown to correlate with a variety of biological processes in CC, including invasion and metastasis, implying that miRNAs may serve as a set of potential biomarkers for the diagnosis and treatment of metastatic CC using molecule-targeted therapy [23].

1.2 Intronic miRNAs

DNA sequences are characterized as exons or introns according to their ability to translate genes. After transcription, the introns are spliced out of the mRNA because they do not contribute to protein synthesis. However, introns include up to 40% of miRNA genes [24]. To date, bioinformatic techniques have found over 90 intronic miRNAs, however the function of the great majority of these molecules has yet to be defined [25]. Typically, they are present in a sense orientation and are controlled in association with their host genes [7, 26, 27].

Intergenic miRNAs are known to be transcribed independently of their host transcription units, but intronic miRNAs are expected to be processed from their host transcription units' introns and hence share common regulatory mechanisms and expression patterns with their host gene [8]. However, an intronic miRNA can mute

genes that are functionally antagonistic to the host gene product, so establishing a positive feedback loop that benefits the host gene's physiological function [28].

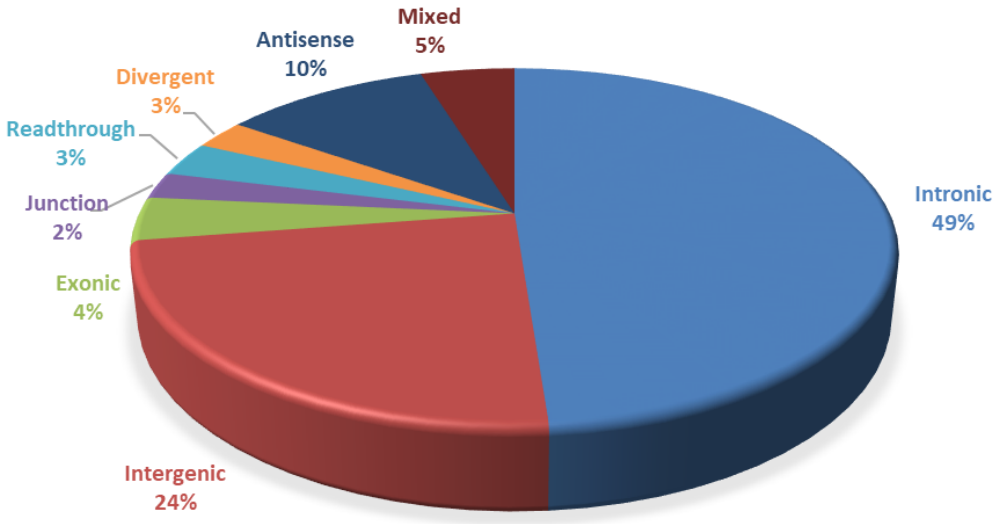


***Figure 2**: MiRNA categorization.*

1.3 Biogenesis of intronic miRNAs

Intronic miRNA are generated from an mRNA sequence's intron regions. Two distinct ways for processing intronic miRNAs from the host gene primary transcript have been reported [29]. In one case, Drosha-mediated processing happens following integration of the transcript into the spliceosome complex but prior to intron excision. Colocalization of primary miRNAs (pri-miRNAs) and Drosha at transcription sites, as well as pri-miRNA concentration in the nuclear chromatin fraction, enables simultaneous transcription and processing of miRNA sequences [7, 30, 31]. Following Drosha cleavage, Exportin-5 exports precursor miRNA (pre-miRNA) sequences to the cytoplasm [23, 24], where they are processed by Dicer into mature miRNAs and integrated into the RNA-induced silencing complex (RISC) [32]. According to other research, intronic miRNAs located in short introns (Mirtrons), are exported straight to the cytoplasm following splicing and Dicer cleavage [24]. Thus, Mirtrons circumvent

the first Drosha/DGCR8 processing [33]. Those mirtrons, are highly co-expressed with their host genes [29]. According to bioinformatics research, over 20% of intronic miRNAs are projected to target their host genes [34], such as miR-26b, which regulates neuronal development by suppressing its host transcript, *ctdsp2* [35].

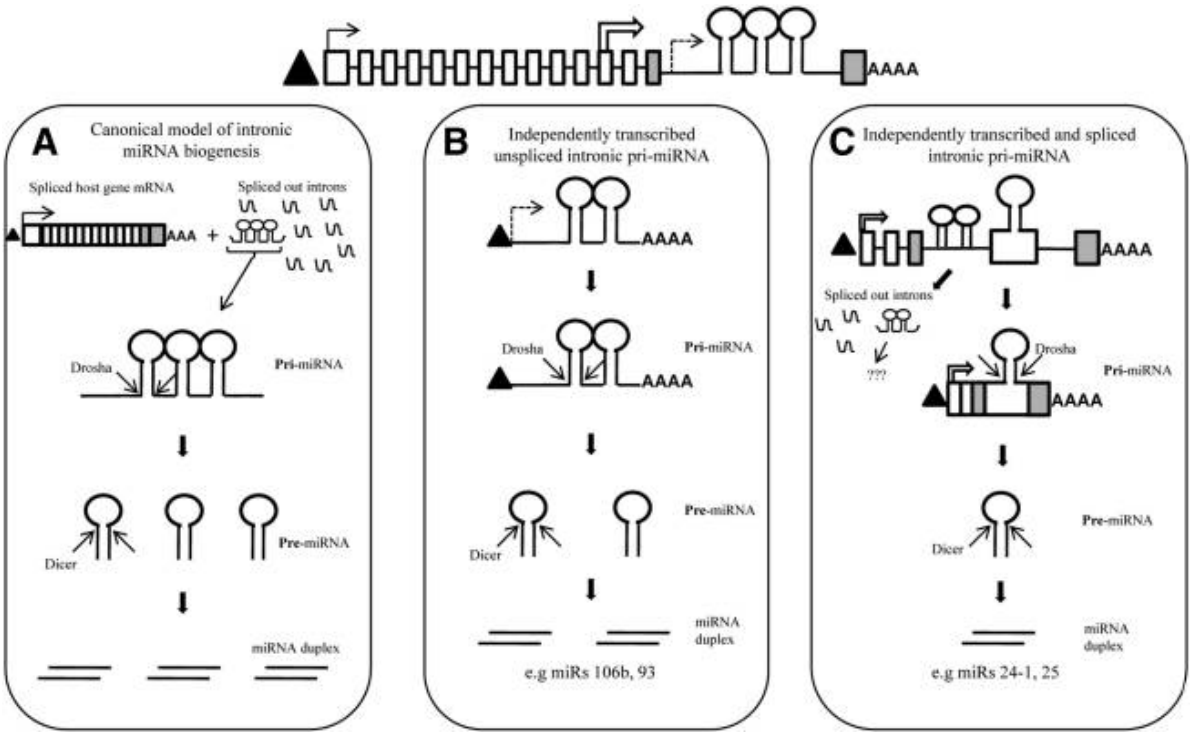


***Figure 3**: Intronic miRNA biogenesis model. Adapted from [8]*

According to the canonical pathway of intronic miRNA production, it is expected that the expression of intronic miRNAs will coincide with the expression of their respective host genes [8]. Drosha, the primary nuclear RNase III-like enzyme in humans, works in conjunction with its cofactor DGCR8 (DiGeorge syndrome critical region gene 8) to form a microprocessor complex that creates a 70 nt pre-miRNA precursor molecule. Following Drosha breakage, the receptor protein Exportin 5 exports the pre-miRNA precursor to the cytoplasm. In the cytoplasm, another RNase III enzyme, Dicer, transforms the pre-miRNA to a mature miRNA duplex that subsequently targets certain messenger RNAs for translational inactivation and/or degradation [36].However,Drosha

can control the level of the microprocessor complex by cleaving hairpins in the DGCR8 mRNA's 5-UTR and coding region, thereby destabilizing the mature transcript and resulting in a decrease in DGCR8 protein [37].

1.4 Molecular mechanisms of intronic miRNAs mediated gene expression regulation

Between 50% and 80% of miRNAs are encoded within the introns of host mRNA genes, implying that the miRNA biogenesis and pre-mRNA splicing processes are co-regulated [38]. If co-transcriptional intron breakage occurs between an alternatively spliced exon and its intronic regulatory components, it can affect alternative splicing. These findings suggest a requirement for RNA continuity between exons and their regulatory elements, as well as the critical role of such elements being tethered to the elongating Pol II transcription complex during splicing [39].

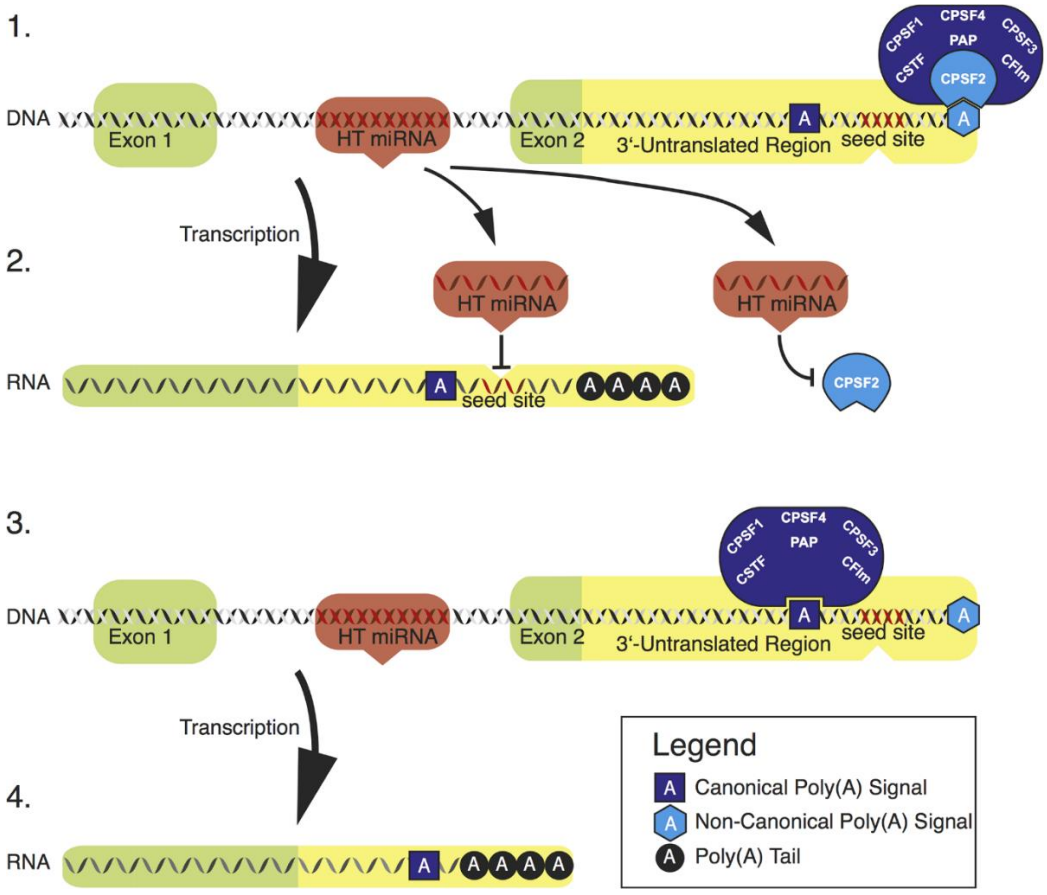


***Figure 4**: Intronic mirna negative feedback regulation model.Adapted from [40]*

As for example, it is well established that intronic transcript cleavage events contribute to the production of mature mRNA in Saccharomyces cerevisiae. Co-transcriptional cleavage of intronic dsRNA stem–loop structures by RNAse III activity can result in the degradation of unspliced pre-mRNA and lariat introns, hence regulating the amount of mRNA produced from such transcripts [41]. After miRNA and host gene coexpression, the miRNA influences its host gene as well as CPSF2. The polyadenylation-complex is biased towards identification of canonical sites once CPSF2 is removed [42]. The canonical site that comes before the miRNA binding site is used in the following transcription cycle. As a result, the intronic miRNA that regulates the host gene is inhibited [40]. Theoretically, intronic miRNAs attach to cytoplasmic messenger RNA, depriving it of its ability to express genes. As a result, the gene becomes silent.

1.5 Roles of intronic miRNAs in diseases

MiRNAs play a role in regulating and directing a variety of immunological and cancer cell interactions, as well as critical immune response mechanisms [43]. Several immune cell pathways are regulated by miRNAs. By targeting suppressor of cytokine signaling 1 (SOCS1), which limits proinflammatory responses, and IL-13RA1, which promotes alternative M2 macrophage polarization, miR-155 can boost classical macrophage activation [44]. Additionally, MiR-155 modulates the effector actions of natural killer cells (NK cells). MiR-155 promotes the manufacture of IFN- by activated NK cells by targeting hematopoietic cell specific 5′ inositol phosphatase (SHIP-1), a significant negative regulator of NK cell effector actions [45]. MiR-155 has the ability to influence the polarization of CD4+ T cells (Th1 vs Th2). Here in this case, by boosting the levels of its target c-Maf, a powerful transactivator of the IL-4 promoter, miR-155 promotes the differentiation of CD4+ T cells into Th2 phenotypes, resulting

in increased production of Th2 cytokines IL-4, IL-5, and IL-10 [46]. Thus we can see mir-155 play an important role also in the regulation of immune cells similar examples are also seen in cases of some other miRNAs like miR-17–92, miR-18a, miR-19a, miR-20a etc [47].

Despite the fact that intronic miRNAs and their host genes may be regulated independently, the intronic miRNA may still be able to down-regulate its own host protein-coding gene by targeting the host gene's untranslated region (UTR). Intronic miRNAs could also serve as negative feedback regulators, which has biological implications [4].

Mir-128-2 is an example of this. Mi et al. (2007) show that the intronic miRNA miR-128-2 is significantly overexpressed in patients with acute lymphoblastic leukemia (ALL), but not in individuals with acute myeloid leukemia [48]. The underlying mutation in 99 percent of people with fragile X syndrome is an increase of the intronic CGG repeat (rCGG) in the 5′-UTR of the FMR1 gene[49].FMR1 encodes FMRP, an RNA-binding protein that is involved in polyribosome assembly in an RNP-dependent manner and can limit translation via an RNAi-like route [50].

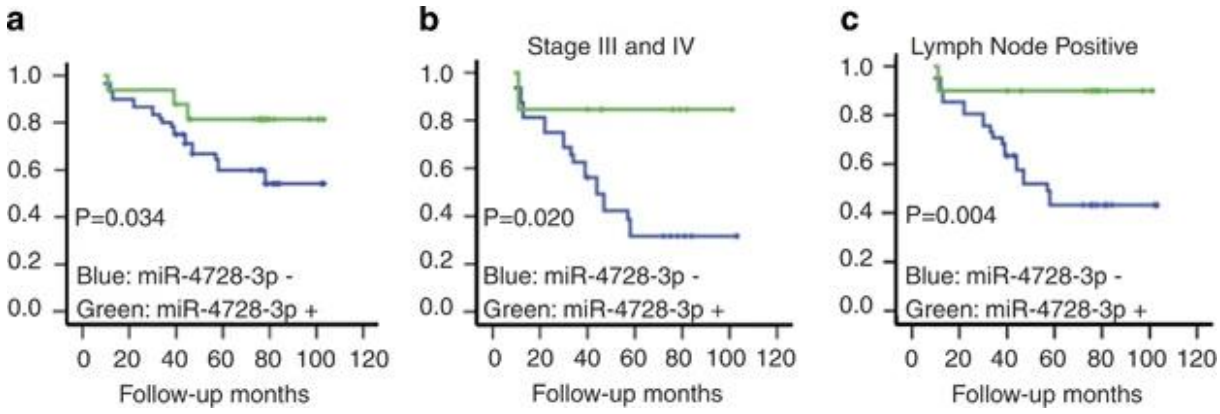


Figure 5*: Loss of miR-4728-3p expression occurs in breast cancer and correlates with worse patient overall survival which shows antagonistic function.Adapted from [51].*

Some of the intron-derived miRNA fragments can form mature miRNAs and efficiently quiet the target genes through the RNAi mechanism after Pol-II RNA processing and splicing excision, while the exons of pre-mRNA are ligated together to produce a mature mRNA for protein production [52]. Pol-II-mediated miRNA synthesis can avoid the harmful effects of dsRNA and siRNA in vitro and in vivo since miRNAs are single-stranded molecules insensitive to PKR- and 2–5A-induced interferon systems [25]. These suggests that introns in intracellular miRNA synthesis and gene silencing can be employed as tools for analyzing gene activities and developing gene-specific therapies for cancer and viral infections [25]. Furthermore, an intronic miRNA can antagonize the canonical signaling of its host gene indirectly [51]. Thus, based on the evidence, we may infer that miRNAs play a significant role in tumor growth; miRNAs are variably expressed according on the molecular subtypes of tumors; and specific miRNA expression profiles categorize tumor malignancies [53].

1.6 Rationale

Pathologic and circumstantial dysfunction typically comes from changes in several intronic microRNAs, and this eventually leads to various disease phenotypes. Recently the association of the intronic miRNA with various types of tumorigenesis and tumor suppressor genes. Whether or if the mechanisms used by transcription factors to regulate gene expression are just as significant as the mechanisms used by introns in gene regulation are yet to be seen. It is possible that introns contained in intronic miRNA are able to activate cell transitions when they sense external stimuli without the necessity of repetitive protein production. Conversely, it has been shown that non-coding small RNAs, or miRNAs, have an effect on oncogenes and tumor suppressor genes. MiRNAs have been implicated and shown to have a role in a variety of malignancies, including breast, colon, gastric, lung, prostate, and thyroid.

The fourth most frequent form of cancer in women is cervical cancer. It mostly impacts the cells of the lower section of the uterus, the cervix, which is situated within the vagina. The screening of the CESC patients in the less developed countries is not available whereas this cancer is most prevalent there. Since microRNA are seen to be mostly associated with various cancers its role in CESC is also prevalent. They play very important regulatory role in cervical cancer control and affect the metastatic potential of cervical cancer cells by coordinate suppression of multiple genes controlling cell motility.

As such, in this study we sought to explore the role of intronic miRNA in association with their host genes in exerting regulatory control on cervical cancer carcinogenesis. Additionally, to explicate the correlation between the common targets of intronic miRNA and transcription factor-encoded host genes with immune infiltrating cells in cervical cancers in order to gain a better understanding of the level of immune infiltration and whether these genes are guiding cancer metastasis or assisting in tumor destruction in the tumor microenvironment.

1.7 Hypothesis:

In this study we have hypothesized that:-

1. Intronic microRNAs have a regulatory effect on their host genes, which are ultimately responsible for oncogenesis of cervical cancers.
2. If those intronic miRNAs, as well as their targets, expressed in all samples and are modulated in such a way that they are linked with CESC patient survival.
3. Intronic miRNA and its associated host gene are associated with immune infiltration in the tumor microenvironment.
4. Those specific intronic miRNAs can be identified as a possible biomarker.

1.8 Objectives of the study:

The main objectives of this study are:

I. Determining the differentially regulated miRNAs and genes from the GEO dataset.
II. Functional enrichment analysis of the genes to see their association with different biological pathways and cancer hallmarks.
III. Determining the miRNAs which are intronic in nature.
IV. Finding out the targets of the intronic miRNA and their host genes
V. The transcription factors of the DE intronic miRs and host genes are identified.
VI. Predicting the correlation of the intronic miRs and host genes with patient survival in CESC.
VII. Searching for the association of the intronic miRs and their host genes in immune infiltration in CESC.

Chapter 2

Methodology

i. Finding out the DE miRs and DE genes:

RNA-seq raw read-count data of the whole transcriptome sequencing in human cervical adenocarcinoma and normal cervical tissues were obtained from the GEO database (GSE145372) [54]. For the differential expression (DE) analysis, we used the Bioconductor package DESeq2 [55] with R version 4.0.3 [56] with a model based on the negative binomial distribution. To avoid false positives, we considered only those log2FC values significant which have a p-value < 0.05.

Then using the ggplot2 [57] and deplyr [58] functions we have visualized the DE miRs and DE genes in volcano plot where we categorized log2FC value of 0.8 or above as "**Upregulated**" and log2FC of -0.8 or below as "**Downregulated**" and the others as "Non DEGs" which has a p value of 0.05 or below. We marked the **upregulated**, **downregulated** and **non-significant** deregulated genes as **red**, **blue** and **grey** color respectively.

ii. Identification of the intronic miRNAs and associated host genes

From the identified differentially expressed (DE) genes and miRNAs, we intended to find out the DE intronic miRNAs and their associated DE host genes. Firstly, to identify the DE intronic miRNAs, we utilized the intragenic_microRNA_v12 dataset of miRIAD [59] database. We used venn analysis to compare between our DE miRNAs and the intronic miRNAs from the dataset by Venny [60]. From the identified DE intronic miRNAs, we then compared the host genes with our DE genes following the same

approach. At the end of this step, we selected once those intronic miRNAs, which are themselves deregulated and have deregulated host genes.

iii. Detection of intronic miRNAs which have transcription factor (TF) encoding host genes

We again filtered our list of DE intronic miRNAs by taking only those which have DE host genes that encode transcription factors. To achieve this, we utilized the human transcription factor database obtained from "The Human Transcription Factors" database developed by Lamber et al. (http://humantfs.ccbr.utoronto.ca/cite.php). We then used Venny tool [60] to compare between this dataset and our targeted DE host genes to find out the DE host genes which are also transcription factors. We selected only those DE intronic miRNAs which have DE transcription factor-encoding host genes. Finally, we selected only those pair of DE intronic miRNAs and DE TF-coding host genes in which both have associated with significant prognosis in cervical cancers (CESC) which we elucidated from the survival analysis from GEPIA2 [61] and UALCAN [62] tools.

iv. Finding out the common targets of the intronic miRNAs and the associated host genes

To find out the common targets of the intronic miRNA and its associated host gene, we first identified the targets of the intronic miRNA and host genes using several different databases to reduce the false positive targets. In order to get the targets of the intronic miRNA, we utilized three different databases of miRNA targets, namely- TarBase v8 [63] , TargetScan v7.1 (http://www.targetscan.org), and miRDB [64]. We took only those targets of the intronic miRNA which have been predicted by at least two of these three databases. For identifying the targets of host genes, we utilized the

hTFtarget [65] database and Cistrome data browser [66]. From Cistrome DB, we downloaded the putative targets of the host genes of *Homo sapiens* and taking those predictions only where TF binding peaks have been observed around 5kbp up- and downstream of the gene transcription start site (TSS). Now from our generated datasets of targets of the intronic miRNA and the host gene, we used Venny tool [60] to find out the commonly targeted targets of both intronic miRNA and its host gene. We obtained the PPI data of these genes from STRING [67] database and then created a network of the target genes utilizing Cytoscape software [68].

v. Functional enrichment analysis

We performed functional enrichment analysis to illustrate the functional significances of the DE host genes, DE miRNAs and common targets of DE intronic miRNA-DE host gene. We carried out the enrichment analysis using gitools v1.8.4 [69] and FunRich tool [70] utilizing the KEGG [71] pathway module. We also utilized the cancer hallmarks module obtained from GSEA [72] and CHG [73] databases. We used the Benjamini-Hochberg False Discovery Rate (FDR) [74] method of multiple test correction model to reduce the false positive enriched terms/pathways. We considered only those enriched terms/pathways significant which have a corrected p-value < 0.05.

vi. Searching for the target genes having prognostic values in CESC

From our generated list of the common target genes of the selected intronic miRNA and its host gene, we then filtered and selected only those targets which have important prognostic significance in CESC. To achieve this, we used GEPIA2 [61] tool for performing the survival analysis using these target genes and then selected only those target genes which have significant logrank p-values (<0.05). Then from the selected target genes, we intended to find out if any of these genes are associated

with immune functions. To find out this, we gathered the immune related genes from InnateDB [75] database and compared it with our filtered list of targets.

vii. Association of the target genes with immune infiltrations in CESC

Next, we sought to find out whether the immune function associated target genes have roles in the tumor immune cells infiltrations. Firstly, we utilized the estimation function of TIMER2.0 tool [76] which uses TIMER [77], CIBERSORT [78], quanTIseq [79], xCell [80] , MCP-counter [81] and EPIC [82] immune deconvolution algorithms to find out the abundances of immune cells for our analysed RNA-seq data. Then we used the "Survival" module of TIMER [77] tool to find out the important immune cells associated with patient survival in CESC. Finally, to find out the correlation between the expression of the immune related target genes with immune infiltration of specific immune cells, we used the "Gene" module of TIMER [77] tool. We considered only those correlation values significant which have a p-value < 0.05.

An overview of the whole workflow has been briefly illustrated in figure 6.

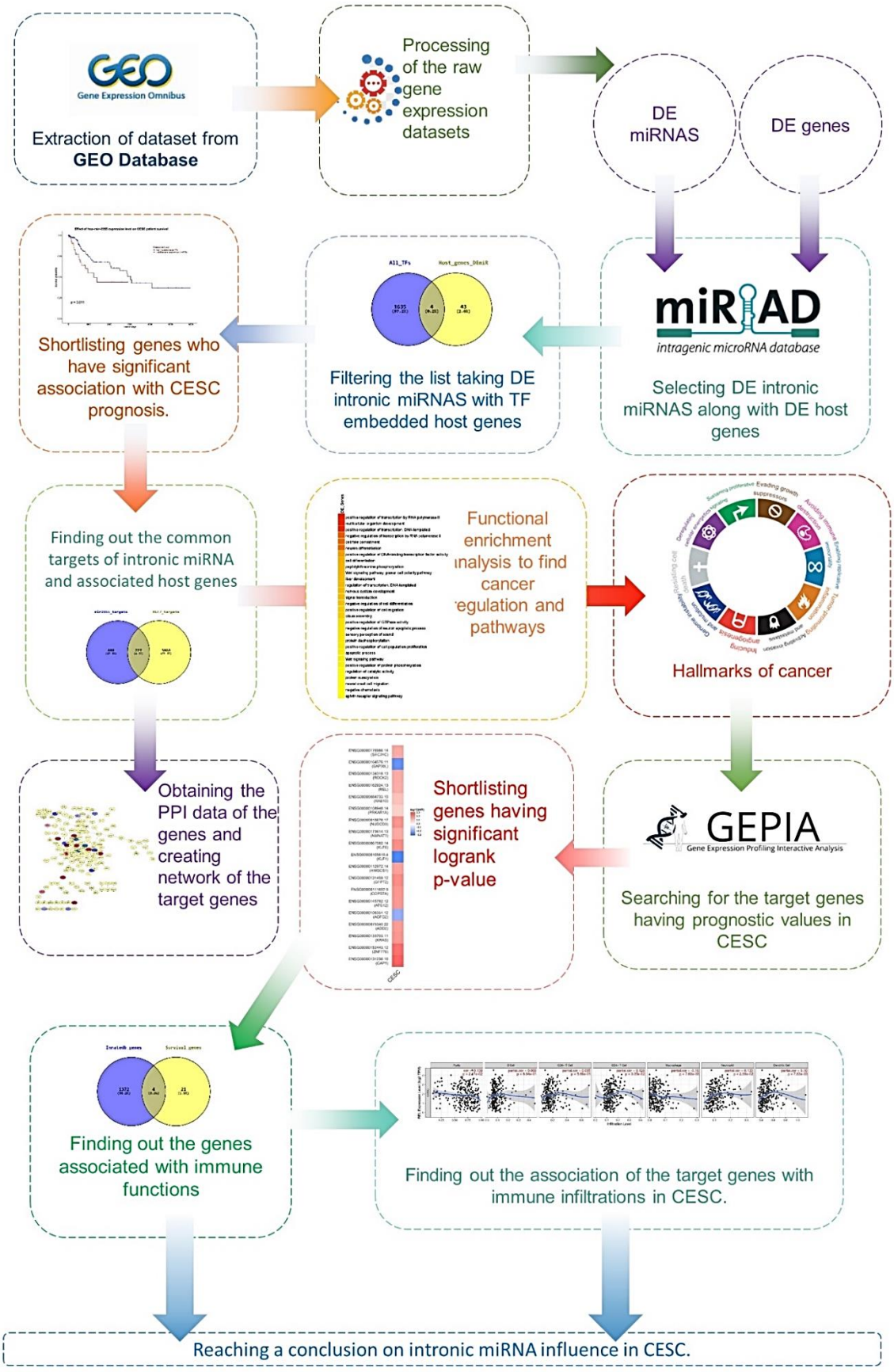


***Figure 6**: Overview of entire workflow.*

Chapter 3

Results

3.1 Identification of differentially expressed (DE) genes in cervical cancer

The dataset of the present study contained 8 samples; 4 normal tissue samples and 4 CESC tumor samples. Differentially expressed genes were screened with P<0.05 and fold control (Log|FC|)>1 criteria was used to obtain DEGs from the dataset from where a total of 2,907 genes were identified. Of these, 1,692 were upregulated and 1,214 were downregulated. Additionally in case of DE miRNA a total of 116 miRs were degulated of which 68 were upregulated and 47 were downregulated.Then using the ggplot2 function of the R software the volcano plot of the DEGs distribution was obtained (Figure 7).

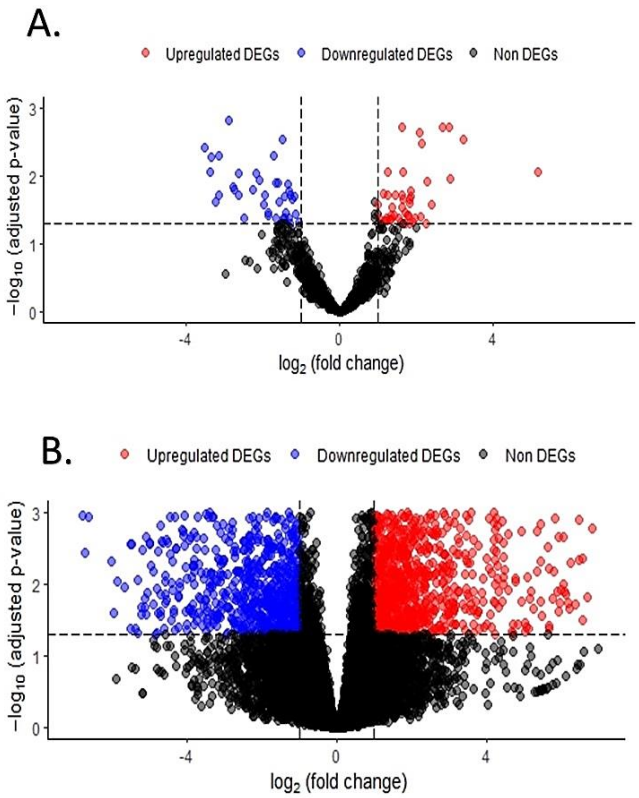


***Figure 7**: Volcano plot of the distribution of all differentially expressed, (A) miRNA genes, and (B) all other genes. Here, red colored bubbles suggest upregulated while bule colored bubbles mean downregulated and non-DEGs as non-significant are represented in grey.*

3.2 Functional enrichment analysis of the deregulated genes and miRNAs in cervical cancer

For the DE miRNAs and genes which we obtained the enrichment analysis is done to find their correlation with different pathways. This analysis allowed us to know the association of the DE miRNAs and DE genes with different pathways. The significance level was calculated on the basis of the corrected p value. The importance of enrichment in terms of modified p-value (0.05) is depicted in all heatmaps using a color-coded P-value scale.

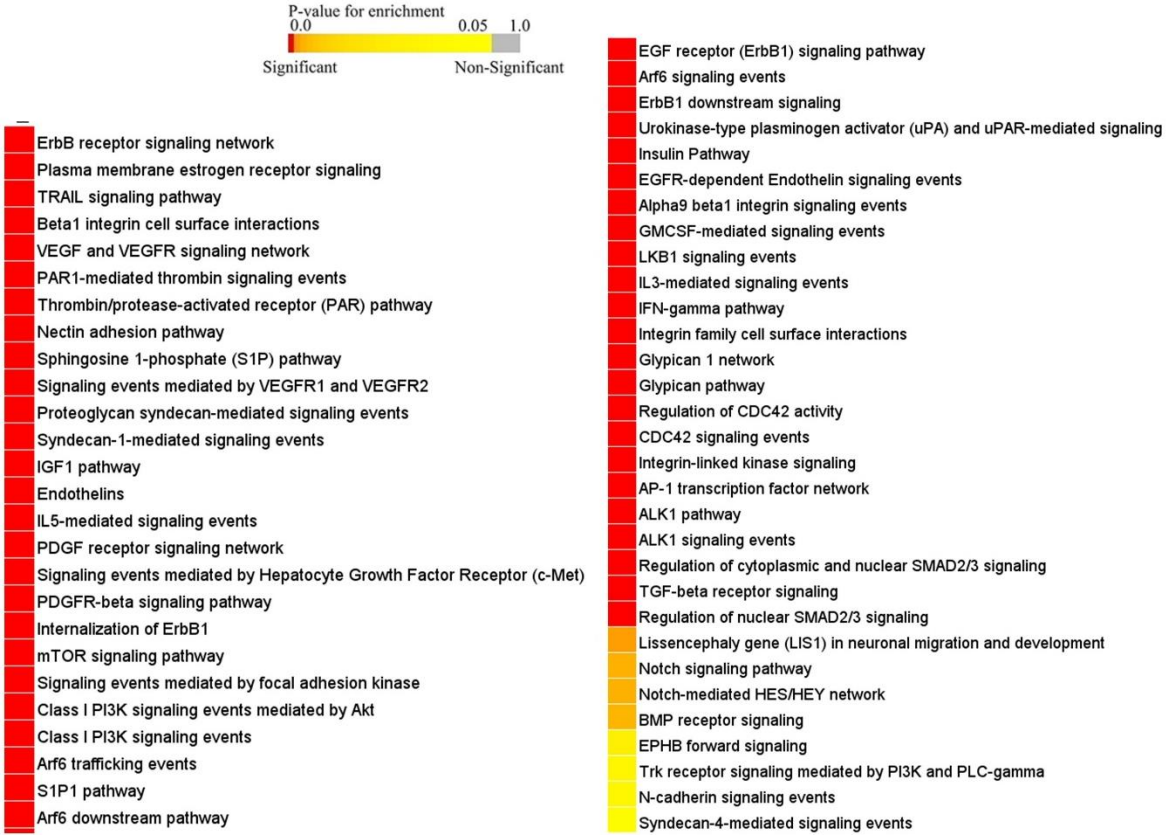


***Figure 8**: Enrichment analysis of DE miRNA using KEGG module. Colours closer to red denote greater significance, yellow denotes less significance, and grey denotes non-significant.*

The differentially expressed miRNAs are observed to be highly associated with different signalling pathway which includes ErBb receptor signalling network, IGF1 pathway, Nectin adhesion pathway, Gkypcian pathway, IFN-gamma pathway, Notch

signalling pathway, BMP receptor signalling, and N-cadherin signalling events (Figure 8). So the differentially regulated miRNA genes are targeting various pathways to regulate cell proliferation, migration, differentiation, apoptosis, and cell motility.

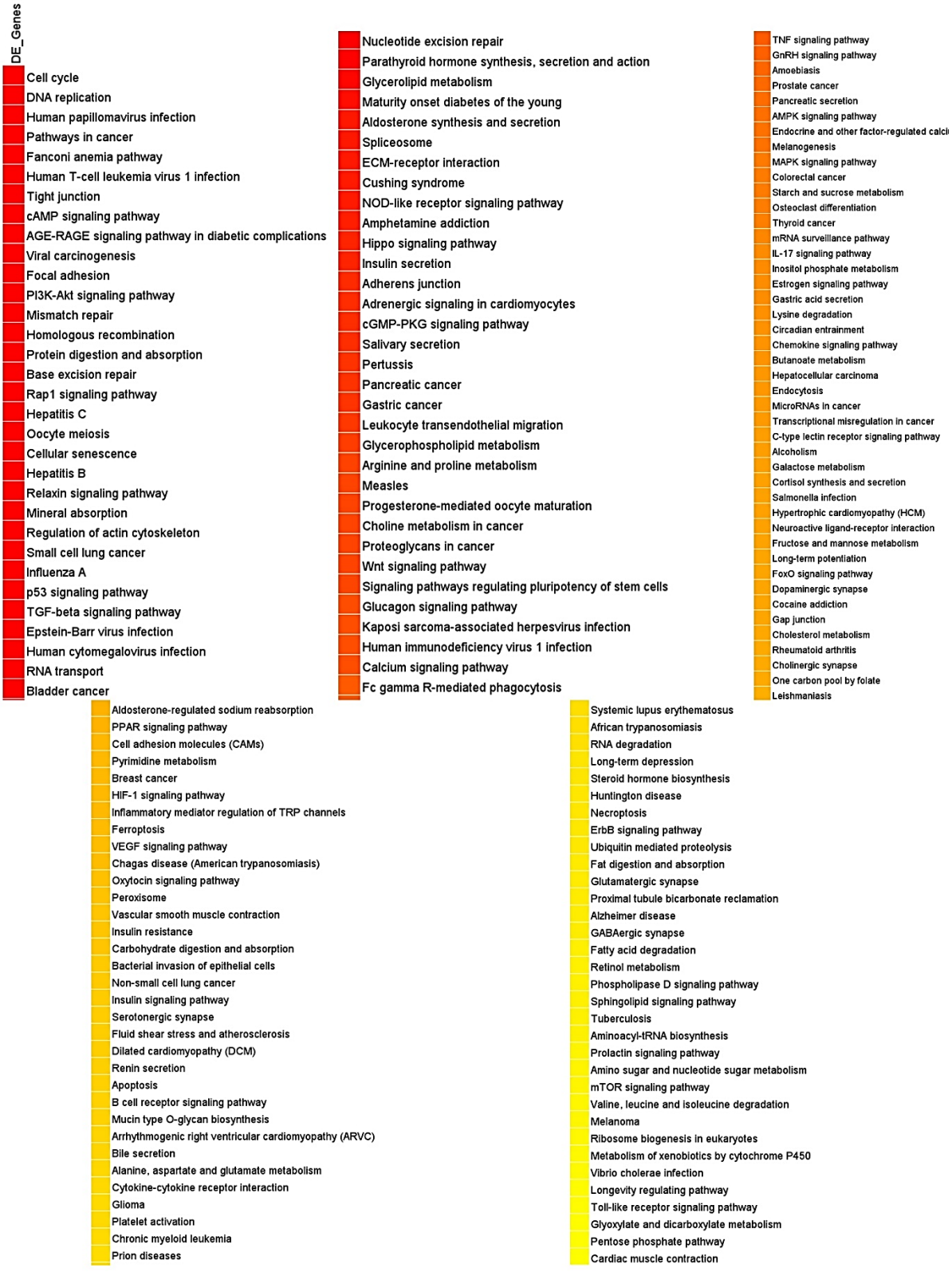


***Figure 9**: Enrichment analysis using KEGG pathway of DE genes.*

The DE genes through enrichment analysis in the context of KEGG module are also seen to be significantly associated with some signalling pathways. These association indicates most of our DE genes are associated significantly with different immune signalling pathways (Figure 9).

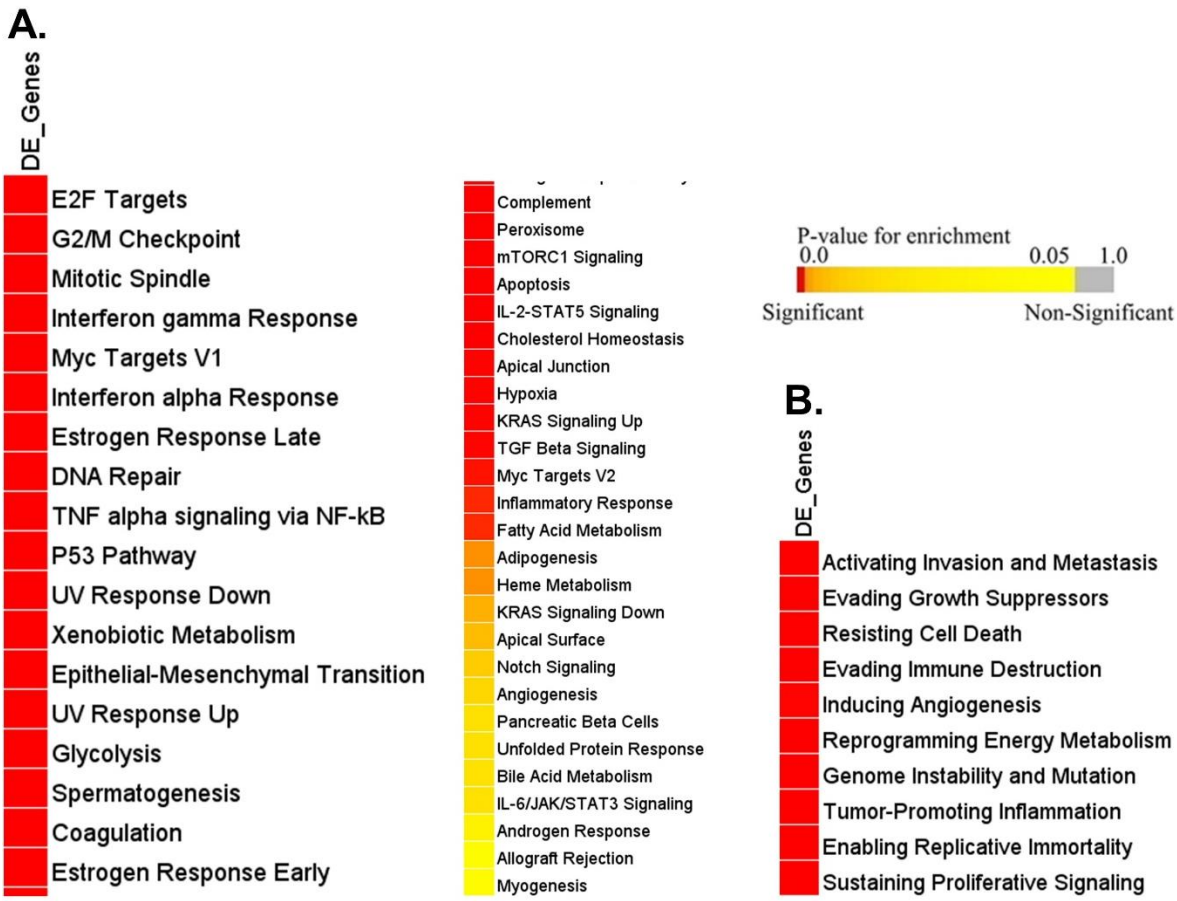


Figure 10 *: Cancer hallmarks for DE genes (A) GSEA hallmarks (B) CHG hallmarks.*

The DE genes are highly significant with most of the hallmarks of cancers in GSEA and CHG which includes p53 pathway, activating invasion and metastasis, inducing angiogenesis, UV response down, and E2F targets (Figure 10). These association indicates that the sustained progression of cancer can be observed as a result of the influence of DE genes. These genes encompass the majority of the characteristics required to denote the stage of cancer.

3.3 Identification of the differentially expressed intronic miRNAs

Using the intragenic_microRNA_v12 dataset of miRIAD we found out a total of 22 intronic miRNAs along with DE host genes. Among them 13 intronic miRNAs along with their associated host genes were upregulated and the rest 9 were downregulated (Table 1).

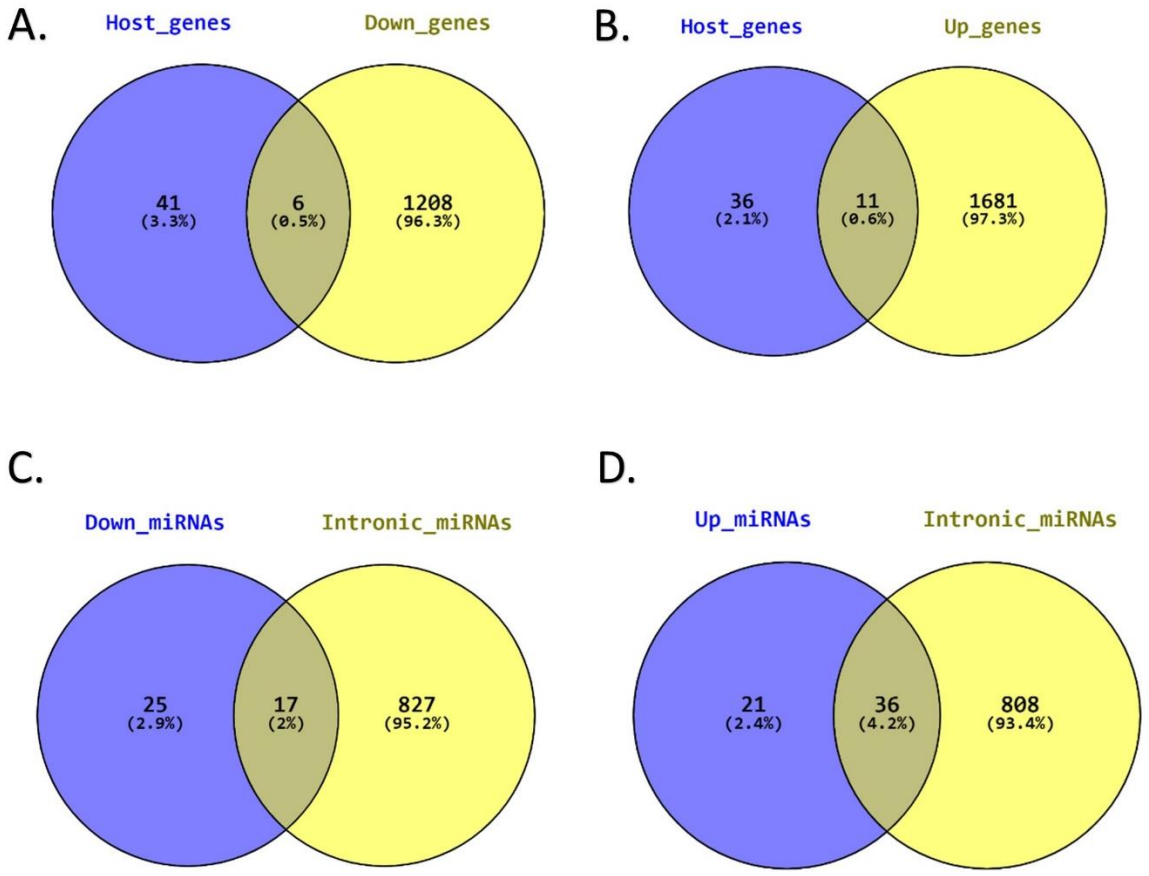


Figure 11: *Venn diagram showing differentially expressed genes among (A) Downregulated genes among host genes. (B) Upregulated genes among host genes. (C) Downregulated miRNAs among intronic miRNAs. (D) Upregulated miRNAS among intronic miRNAs.*

We found out that amongst the 47 host genes, 6 are downregulated and 11 are upregulated. Conversely, in case of the intronic miRNAs, out of the 844 intronic miRNAs, 17 are downregulated and 36 are upregulated (Figure 11).

Table 1: *The following is a list of DE intronic miRNAs and their DE host genes.*

DE Intronic miRNAs	DE status intronic miRNAs	Host genes	DE status of host genes
hsa-mir-1224	Downregulated	*VWA5B2*	Downregulated
hsa-mir-1264	Downregulated	*HTR2C*	Downregulated
hsa-mir-1298	Downregulated	*HTR2C*	Downregulated
hsa-mir-1911	Downregulated	*HTR2C*	Downregulated
hsa-mir-3616	Downregulated	*EYA2*	Downregulated
hsa-mir-448	Downregulated	*HTR2C*	Downregulated
hsa-mir-4700	Downregulated	*UNC119B*	Downregulated
hsa-mir-628	Downregulated	*CCPG1*	Downregulated
hsa-mir-6502	Downregulated	*IRAK3*	Downregulated
hsa-mir-190a	Upregulated	*TLN2*	Upregulated
hsa-mir-2355	Upregulated	*KLF7*	Upregulated
hsa-mir-28	Upregulated	*LPP*	Upregulated
hsa-mir-3065	Upregulated	*AATK*	Upregulated
hsa-mir-335	Upregulated	*MEST*	Upregulated
hsa-mir-4284	Upregulated	*STX1A*	Upregulated
hsa-mir-500a	Upregulated	*CLCN5*	Upregulated
hsa-mir-501	Upregulated	*CLCN5*	Upregulated
hsa-mir-532	Upregulated	*CLCN5*	Upregulated
hsa-mir-548u	Upregulated	*PRIM2*	Upregulated
hsa-mir-615	Upregulated	*HOXC5*	Upregulated
hsa-mir-625	Upregulated	*FUT8*	Upregulated
hsa-mir-652	Upregulated	*TMEM164*	Upregulated

3.4 Enrichment analysis of intronic miRNA

The intronic miRNAs which were differentially expressed are associated with **KEGG** pathway. It was observed those DE intronic miRNAs were associated with various pathways namely ErbB receptor signalling pathway [83], TRAIL signalling pathway [84],Insulin pathway [85], and VEGF signalling pathway [86] (Figure 12). They might also try to reduce some host-induced inflammatory responses by targeting those pathways and function to downregulate these to invoke a proper immune response.

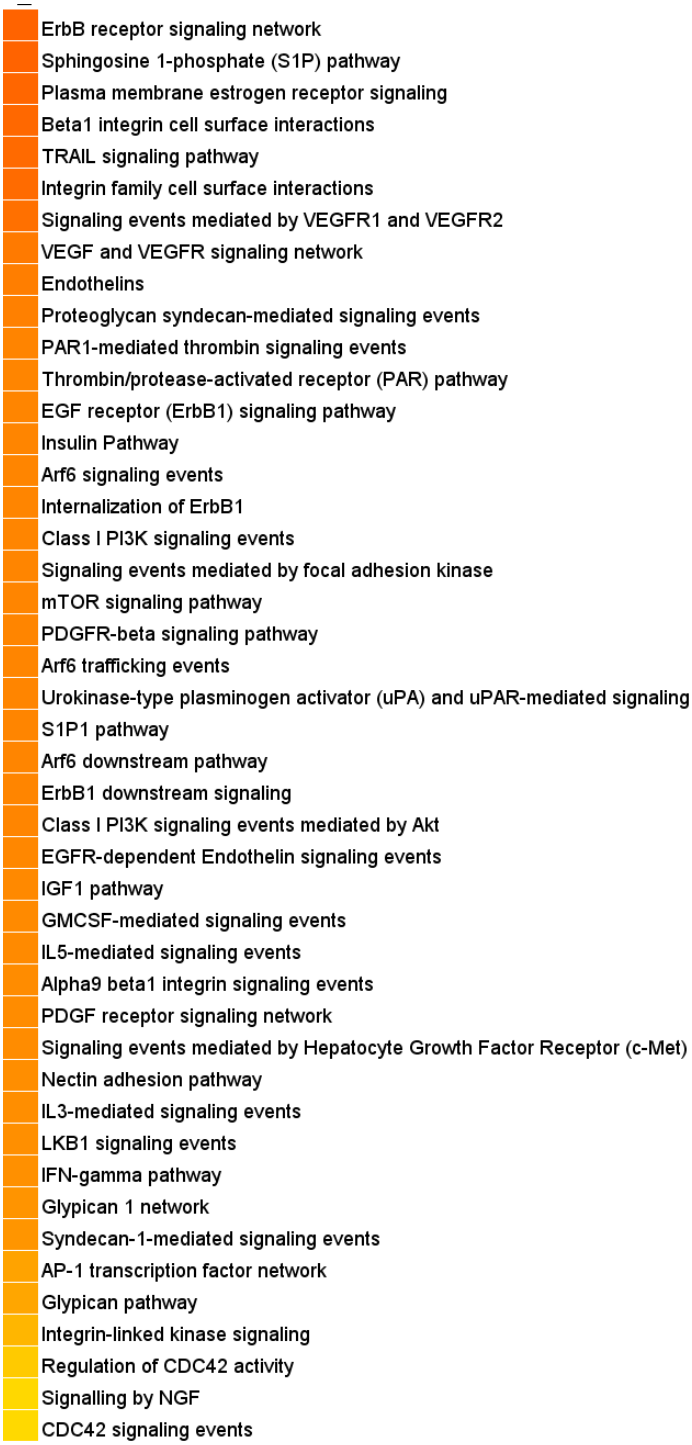


***Figure 12**: Enrichment analysis of DE Intronic miRNA gene pairs using KEGG module.*

3.5 Retrieval of DE intronic miRNAs with DE transcription factor (TF) encoding host genes

The host genes which are differentially regulated and along with them the associated intronic miRNAs are also differentially regulated. The host genes also encode some

transcription factor. Out of seventeen DE host genes, two pairs of genes were selected which DE intronic miRNAs and have DE transcription factor-encoding host genes. (Figure 13)

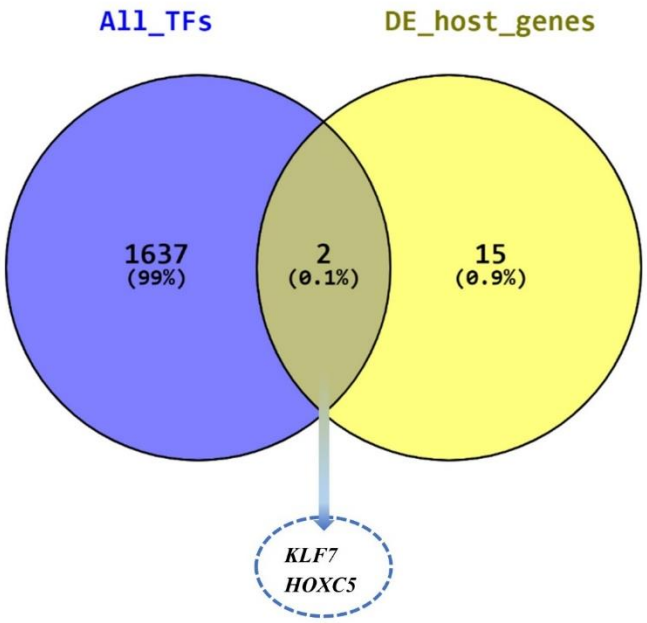


***Figure 13**: Venn diagram showing overlapping of DE host genes and all transcription factors.*

The survival and expression of the two selected genes namely *KLF7* and *HOXC5* are then seen in CESC. The Kaplan–Meier (KM) plot for **Cervical Cance**r or CESC shows 2 groups for *KLF7* and *HOXC5* (Figure 14) where one group has a high expression of gene and the other has a low expression each. For both of them, the highly gene expressed group has lower rate of survival in Cervical Cancer patients. Again, the box plot is showing lower expression of tumor cells than opposed to normal cells for *KLF7* gene (Figure 14).

Next, among the pair of host genes and miRs, we chose *KLF7* and *hsa-mir-2355* because both paired genes were associated with substantial survival in CESC, but the other pair, hsa-mir-615, was not associated with significant survival in CESC, and so was omitted.

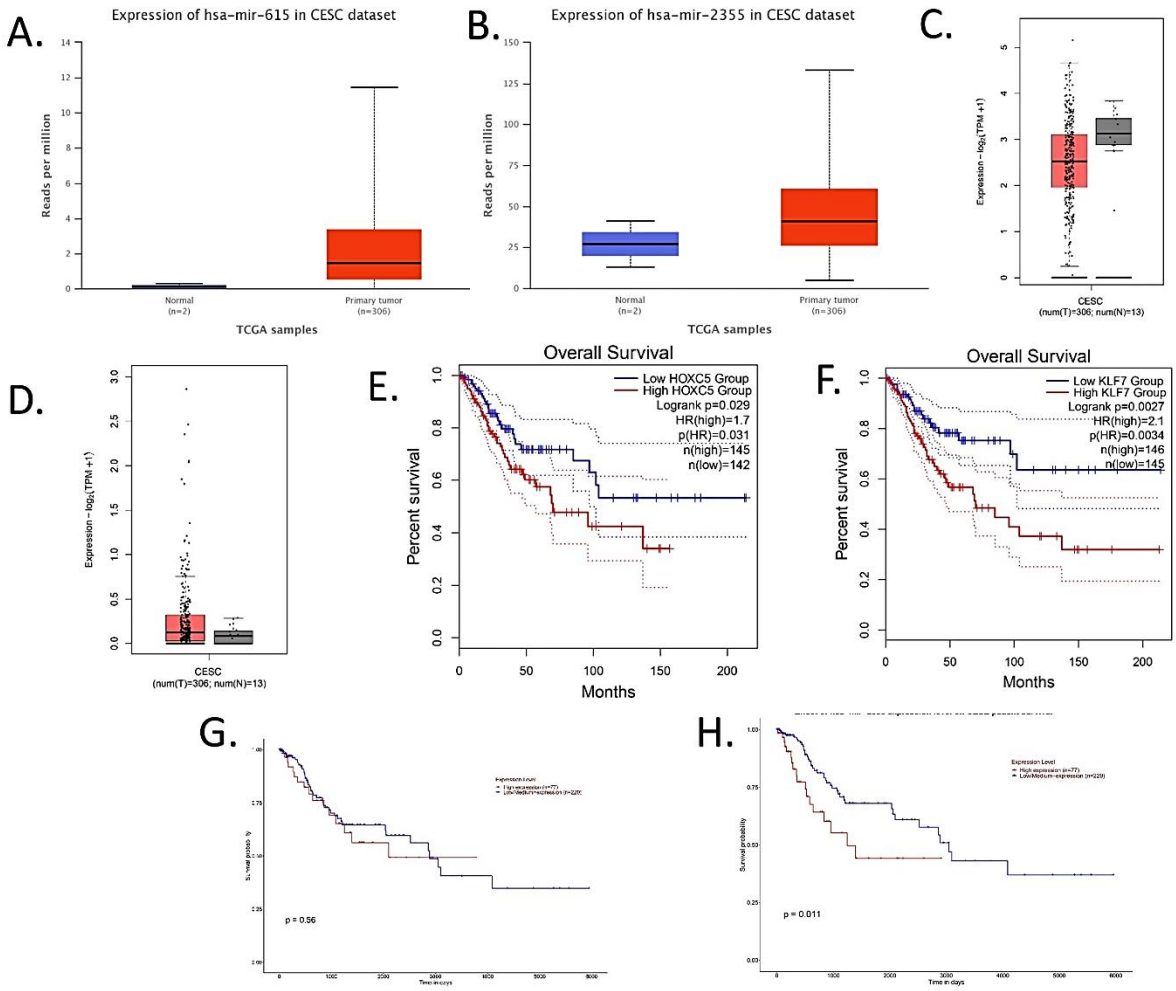


Figure 14*: For CESC, (**A**), (**B**) represents box plot of differential expression of hsa-mir-615 and hsa-mir-2355 respectively; (**C**), (**D**) represents box plot of differential expression data of KLF7 and HOXC5 respectively; (**E**), (**F**) represents Kaplan Mier plot for survival of HOXC5 and KLF7 respectively; (**G**), (**H**) shows effect of hsa-mir-615 and hsa-mir-2355 expression respectively.*

3.6 Two hundred twenty-seven common targets of *KLF7* and miR-2355 was retrieved

After that we sought to find out the common targets of *KLF7* and *hsa-mir-2355*. Using different databases, we got a total of 1115 and 4009 targets of *mir-2355* and *KLF7*, respectively. We obtained 227 overlapping targets of *KLF7* and *hsa-mir-2355*. Then we showed the comparative network of 227 common targets along with their expression in our dataset (Figure 15).

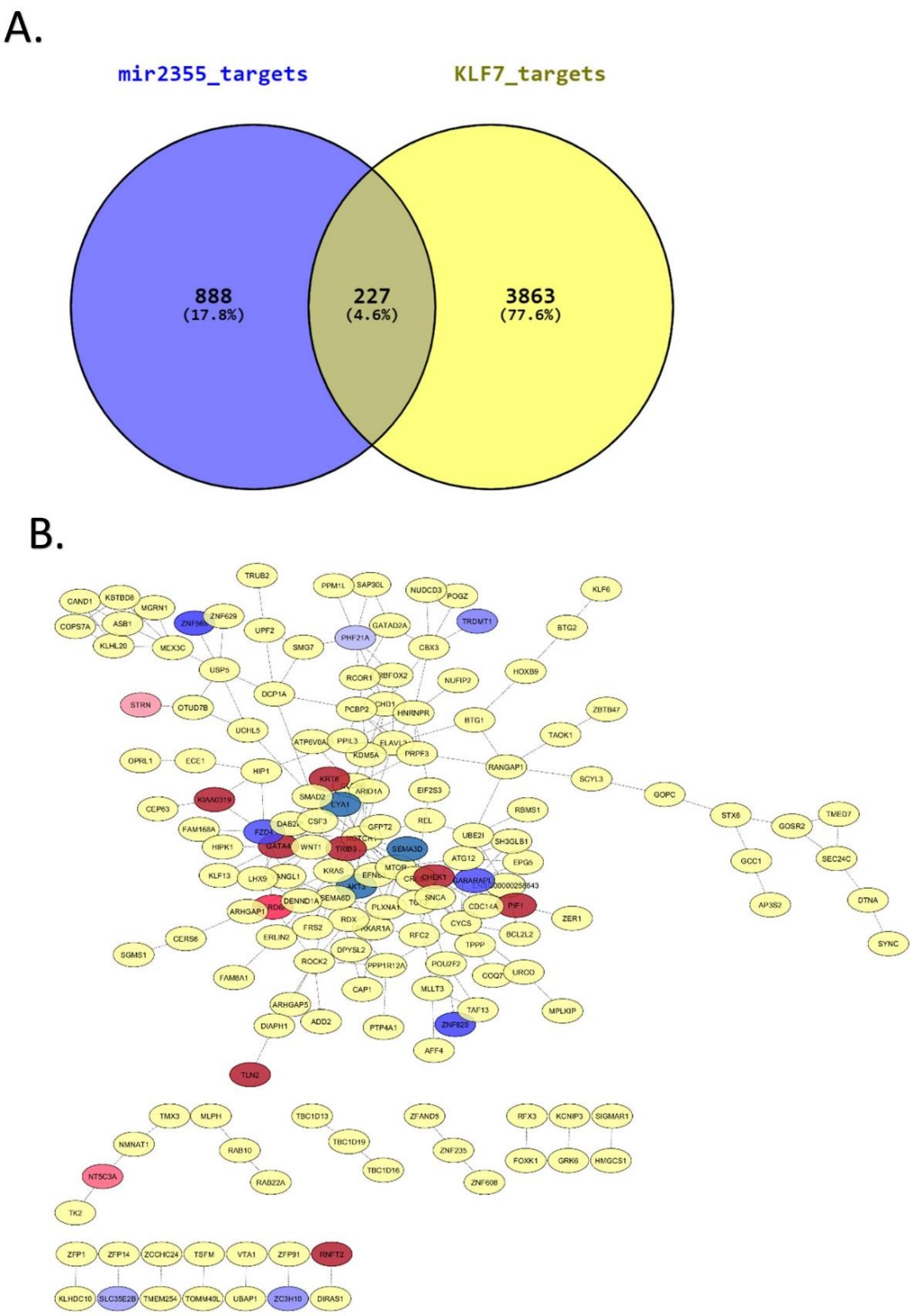


***Figure 15**: (**A**) Venn diagram showing the common targets between mir-2355 and KLF7. (**B**)Comparative network analysis of the common targets along with their expression in CESC dataset.*

3.8 Enrichment analysis of the common targets of *KLF7* and miR-2355

Then we do the enrichment analysis of the common targets to find out their association with different pathways and as well as cancer hallmarks.

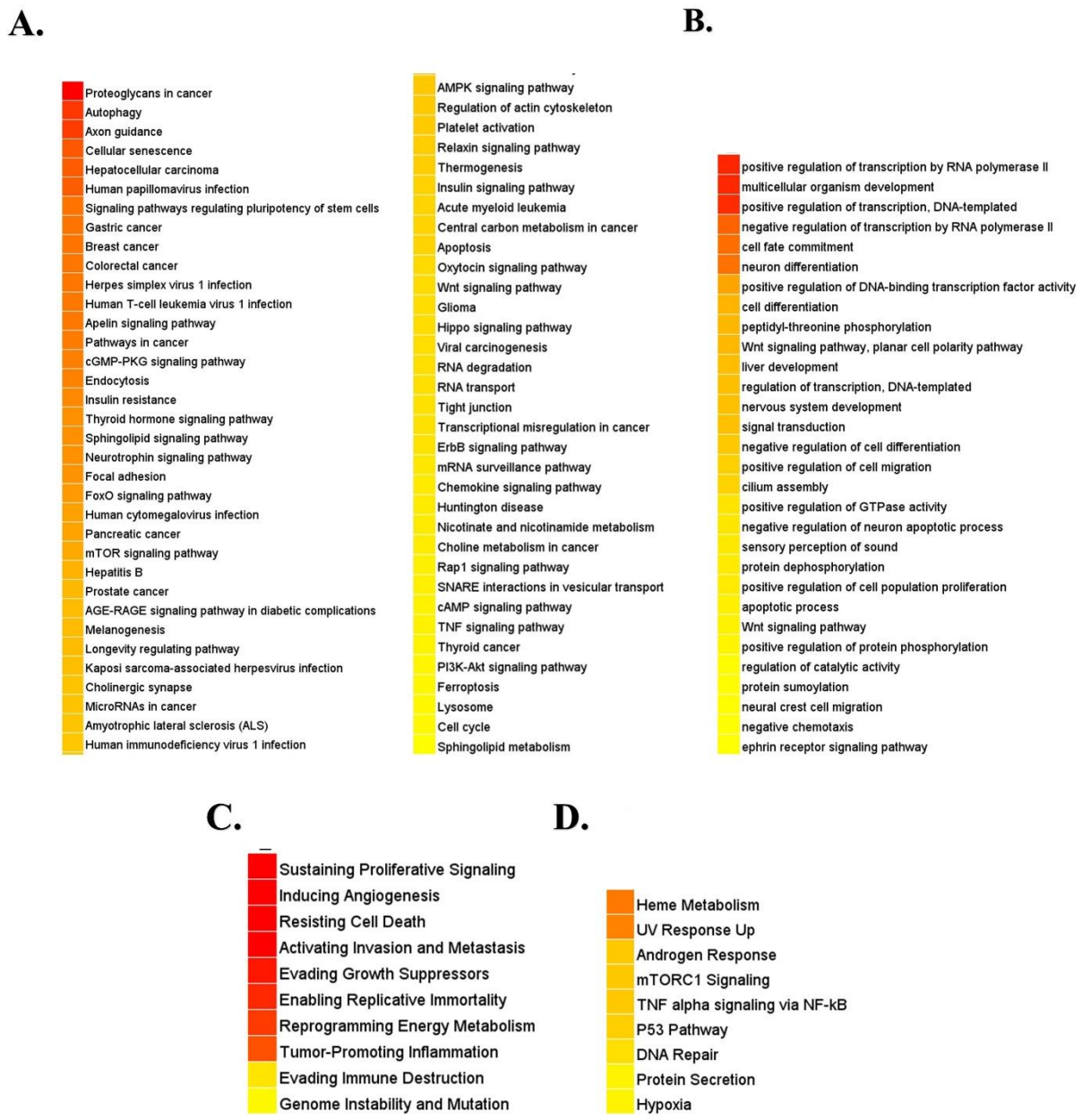


***Figure 16**: Enrichment Analysis of common target genes using (**A**) KEGG Module, (**B**) GOBP Module. (**C**), (**D**) Cancer hallmark association with the differentially expressed genes of common targets.*

Our common target genes are also searched for their association various reaction pathway which have been obtained by performing enrichment analysis in the context of **KEGG** and **GOBP** modules (Figure 16) from where we can see they have been involved in insulin resistance, autophagy, Human papillomavirus infection, endocytosis, cell differentiation, positive regulation of transcription by RNA polymerase II etc. When immune cells recognize infection or tissue damage, they activate several transcription factors. As a result, the majority of these common targeted genes actively follow the immunological pathway.

In addition to the leading to the immune pathways they are seen to be associated with some hallmarks of cancers. The hallmarks of cancer are related to tumor pathogenesis which is breached by a cell on the path towards cancer. The targeted genes are significantly associated with the hallmarks such as sustaining proliferating signalling, resisting cell death, inducing angiogenesis, evading growth suppressor, Hypoxia, p53 pathway, DNA repair etc which can be targeted for cancer diagnosis and therapy (Figure 16).

3.9 Immune Infiltration estimation in the analysed CESC samples

Next, we estimated a plot, using different prediction methods, where different infiltration rate of different types of cells are found in the four tumor samples of our dataset. We can see that NK cells are somewhat infiltrated in our sample according to Quantiseq prediction method whereas macrophage and neutrophil are more infiltrated. Myeloid dendritic cells are seen very much infiltrated in our sample according to the prediction of TIMER and infiltrated to some extent in all samples according to other prediction methods. According to all the prediction methods, B cells are seen to be involved significantly with this tissue in our sample. Overall we can see CD8+, CD4+,

macrophages, neutrophils, mDCs and B cell all are infiltrated in a good level according to the prediction methods whereas monocytes and NK cells are infiltrated less in our sample (Figure 17).

Figure 17*: Infiltration levels of different immune cells under different prediction methods for the four tumor samples.*

After that, we observe which immune cell types are more in our sample. Basically, some immune cells are more or less prevalent in all our tumor samples. Like in the figure 12, we can see CD4+ cell and B cell along with macrophages are prevalent in all of our tumor cells according to Cibersort method. In Quantiseq tool, we can see the proportion of neutrophil is much in all tumor samples. Lastly, in EPIC tool we can see the prevalence of CD4+ along with a slight prevalence of CD8+ is observed in all the tumor samples (Figure 18).

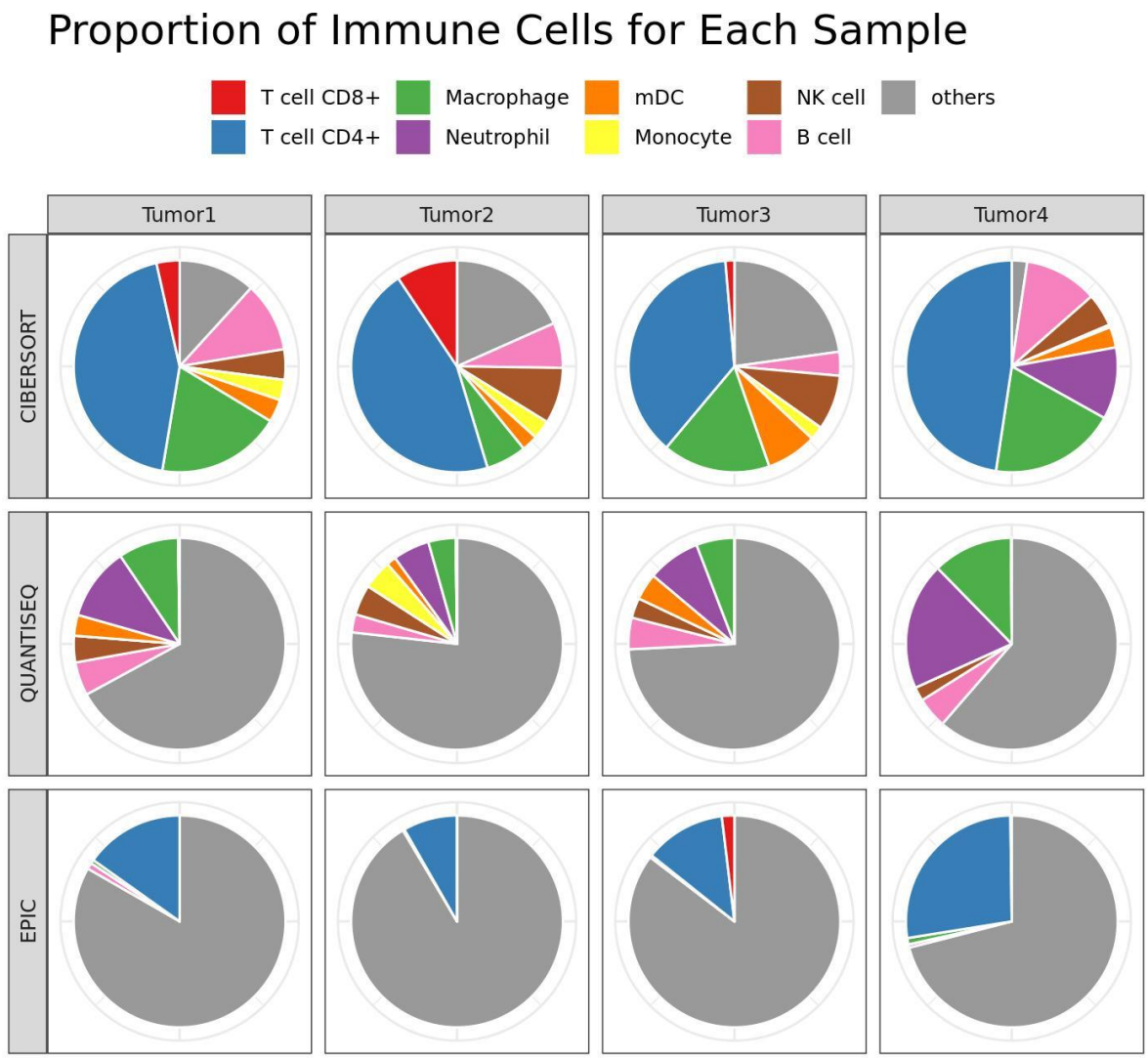


Figure 18: *Prevalence of each immune cell type in our tumor samples.*

3.10 CESC significant survival gene:

We previously obtained 227 target genes and examined their connection with cancer by monitoring their survival in CESC. The targeted genes are undergone through some analysis where only 25 genes among them were significantly associated with CESC patient survival. Thus, we identified 25 genes that are strongly linked with CESC patient survival which we used for further analysis.

The Kaplan–Meier (KM) plot for **CESC** also shows 2 groups for each of the genes with one group having a high gene expression and the other with low gene expression. As the time goes, the survival rate for both groups gets lowered in each of the high and low gene expressed group. Despite that in case of *AGFG2*, *KLF1* groups all of the lowly expressed group ultimately has the least rate of survival (Figure 19-21).

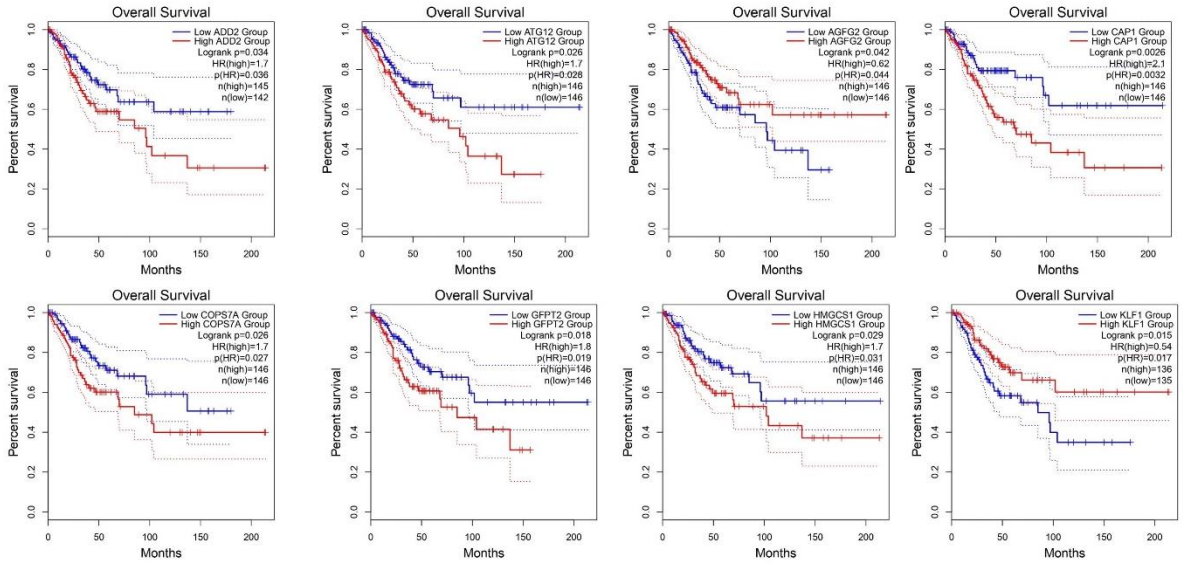


***Figure 19**: For CESC patient represent Kaplan Mier plot for survival of ADD2, ATG12, AGFG2, CAP1, COPS7A, GFPT2, HMGCS1, KLF1 (part 1)*

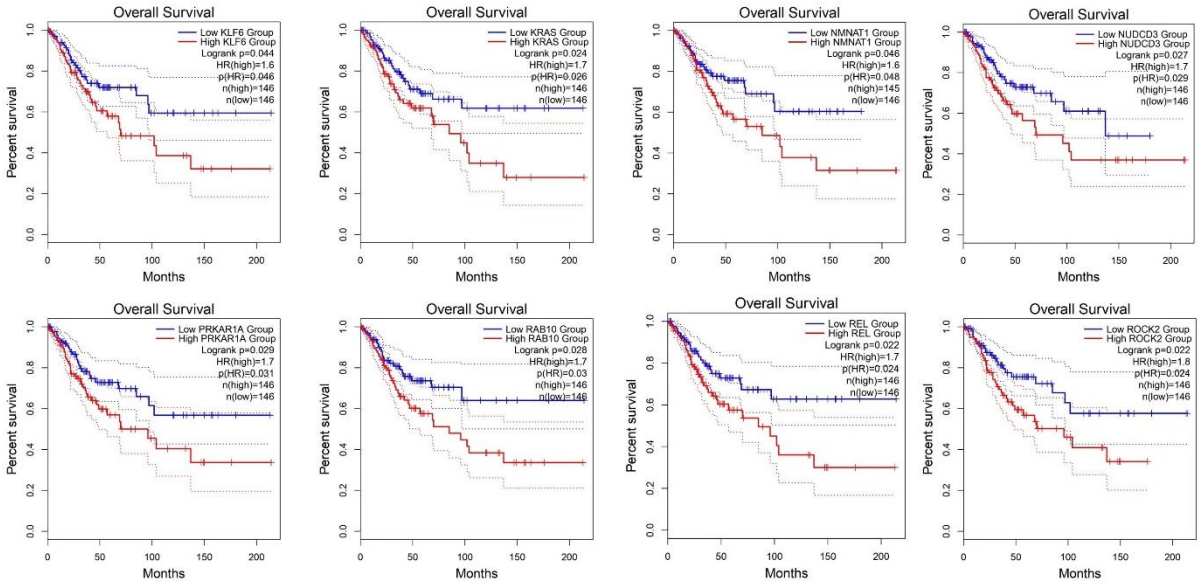


***Figure 20**: For CESC patient represent Kaplan Mier plot for survival of KLF6, KRAS, NMNAT1, NUDCD3, PRKAR1A, RAB10, REL, ROCK2.(part 2)*

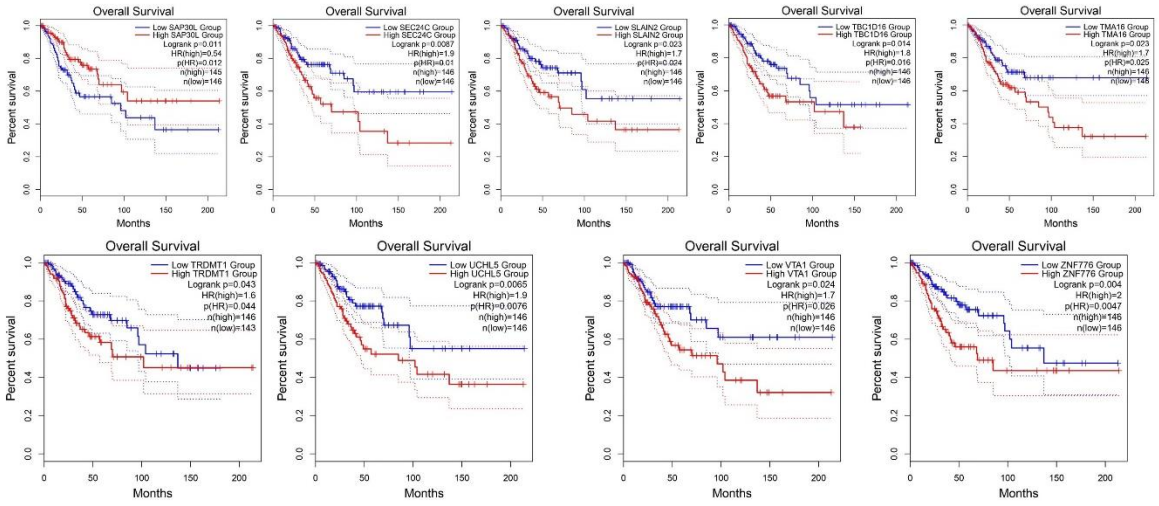


***Figure 21**: For CESC patient represent Kaplan Mier plot for survival of SAP30L, SEC24C, SLAIN2, TBC1D16, TMA16, TROMT1, UCHL5, VTA1, ZNF776. (part 3)*

Then, by analyzing the survival map, we determined these genes play some regulatory effect in patient survival. Here, the blue-colored genes indicate that if they are expressed at a low level, the patient will survive less, whereas the red-colored ones indicate if they are expressed at a high level, the patient will also survive less. This significance is associated is calculated in accordance with the hazard ratio which indicates the hazard in two groups (Figure 22).

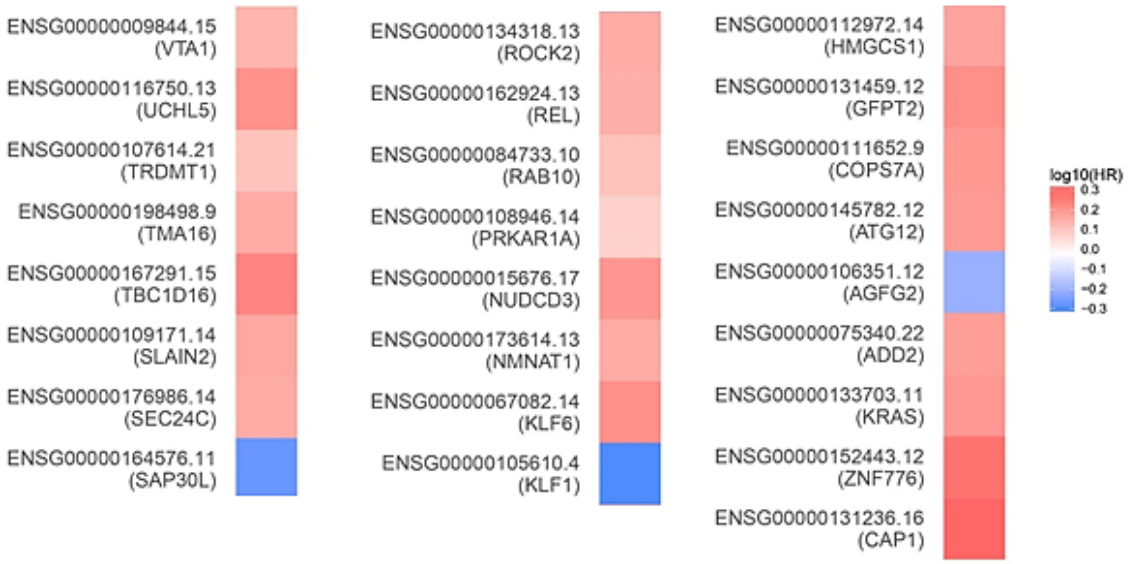


***Figure 22**: Survival map of the targeted genes in CESC.*

3.11 Immune related significant survival gene:

To enumerate immune-related survival genes, we used the InnateDB database and compared it to our survival genes. We discovered only four immune-related survival genes from all of these datasets (Figure 23).

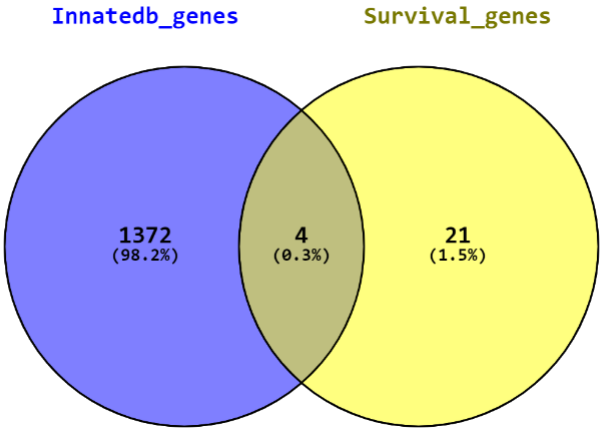


***Figure 23**: Venn diagram showing common genes overlapping the innante db genes and survival genes.*

Target genes related with immune function play a role in tumor immune cell infiltration. *ATG12*, *KRAS*, *PRKAR1A* and *REL* are the four target genes which are both immune signalling related as well as significant with patient survival.

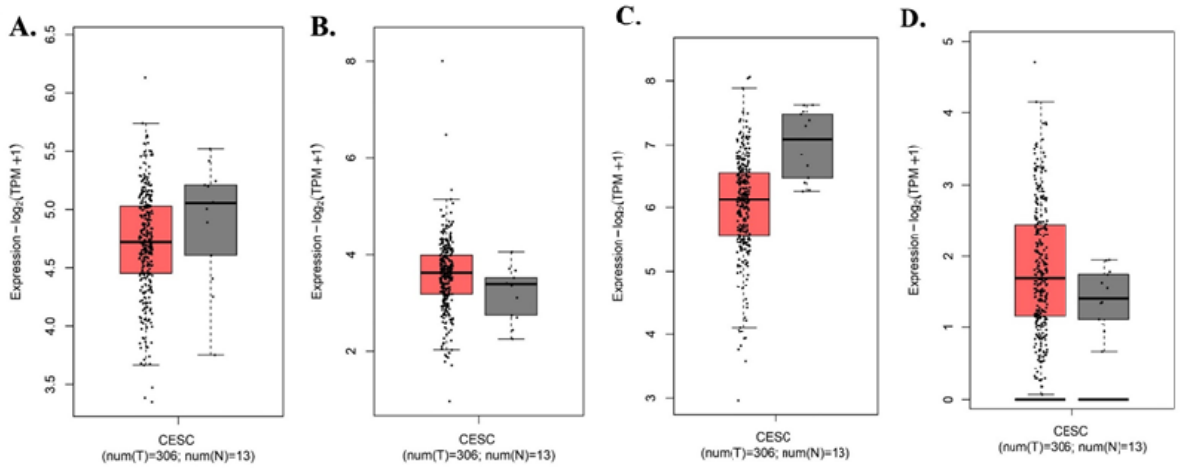


***Figure 24**: Box plot represents the expression of (**A**)**ATG12** (**B**)**KRAS** (**C**)**PRAKAR1A** and (**D**) **REL** in CESC.*

Significantly higher amount of tumor cell expression level than normal cells in CESC is observed for all the genes except for *KRAS* and *REL* genes (Figure 24).

The correlation between the expression of the immune related target genes with immune infiltration of specific immune cells were determined. In CESC dataset we checked which cells are significantly related with patient related and found out that CD4+ T immune cells is related with CESC patient survival. Then we compared it with our 4 target genes and found that they can recruit CD4+ T cells. Among the four genes, only *REL* is considered as it has correlation values significant to a p-value < 0.05 (Figure 25).

In CESC patient survival, the immune cells are seen to have some regulatory role. Low level of B cell, CD8+ T cells, CD4+ T cells and macrophages has an increase rate in overall patient survival. While the high level of neutrophil and dendritic cell is associated with an increase rate in patient survival. Since CD4+ has a significant p value we can say it is highly associated with CESC patient survival (Figure 25A).

ATG12 is seen to negatively correlated with B cell and macrophage while being weakly correlated with the other immune cells. Whereas in case of *KRAS* gene B cell, macrophage and CD4+ T cells are seen to be negatively correlated and the others are weakly correlated. Conversely, *PRKAR1* is seen to be weakly correlated with all the other immune cells. *REL*, on the other hand, is correlated with all the other immune cells except for the macrophages where it is negatively correlated (Figure 25B).

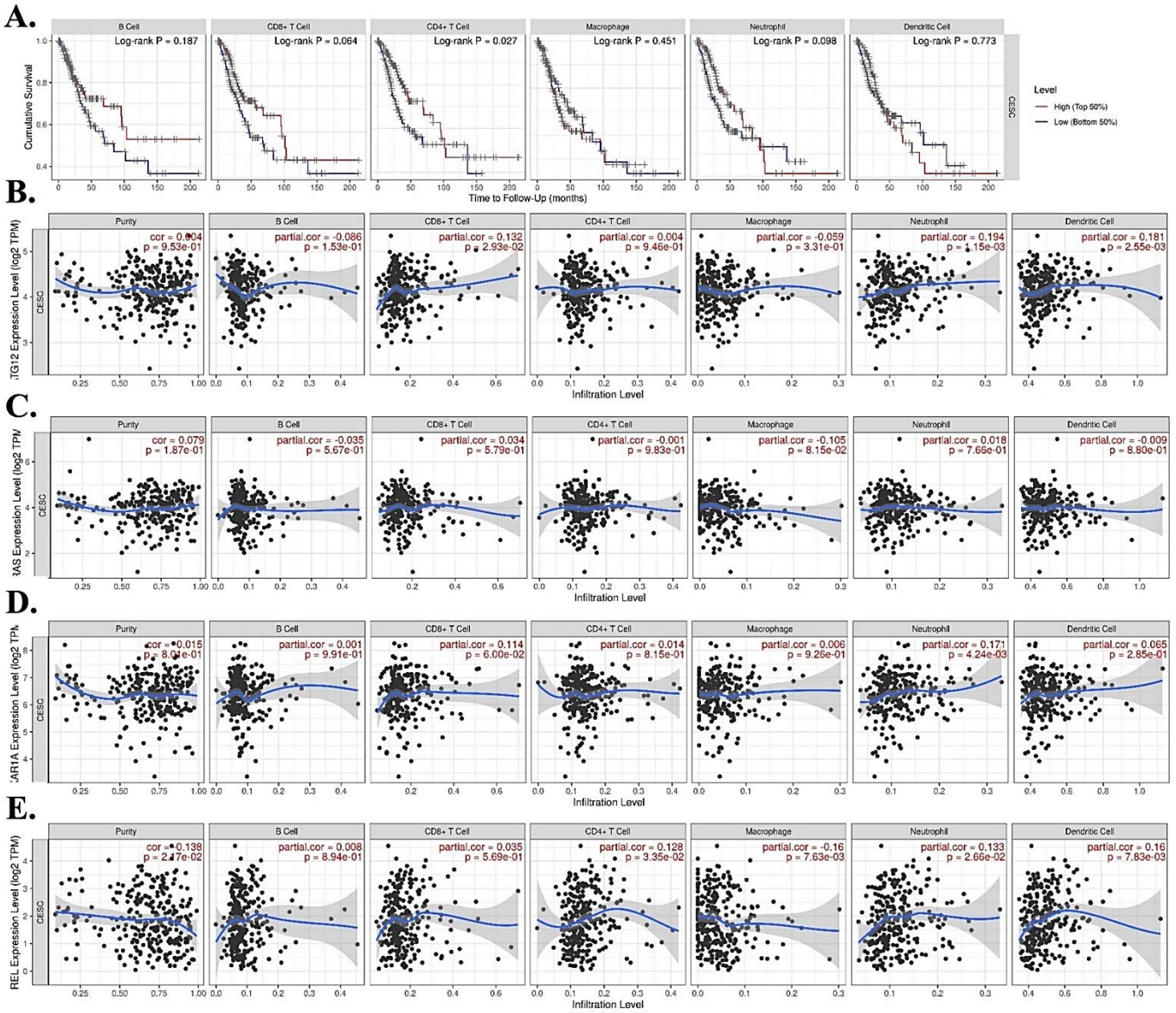


***Figure 25**: (**A**) shows association of immune cells with patient survival in CESC.(**B**), (**C**), (**D**) ,(**E**) shows correlation value of CESC in **ATG12** ,**KRAS** , **PRKAR1A** and **REL** genes with immune cells respectively.*

Due to increase in B cell, CD8+ T cell, CD4+ T cell and macrophage the patient survival goes in risk prone zone. Here, CD4+ T cells are observed to be significantly associated with the patient survival of CESC. Therefore, among the four genes only *REL* directly points out towards recruiting CD4+ T immune cell during cancer prognosis for CESC.

From the figure 18, we can see that in our four tumor samples CD4+ is prevalent to be in greater portion than the other immune cells. Therefore, we can conclude that *REL* can be used as a prognostic biomarker in case of CESC cancer therapy.

Chapter 4

Discussion

Recently, microRNA has gotten a lot of interest as a possible candidate for researching their involvement in gene expression regulation to delineate a variety of physiological processes connected to the progression and development of human diseases. Introns of annotated protein-coding genes encode about half of the experimentally found human miRNAs, which are referred to as intronic miRNA anchored within the host gene.

The fact that intronic miRNAs are co-expressed and have similar transcriptional regulatory mechanisms with their host genes in humans has elevated the importance of research in this subject. Until recently, very little is known about the regulatory cascade in humans that is mediated by intronic miRNA and host gene apart from its physiological importance.

MiRNAs modulate a number of cancer and immune cell pathways in the human body. The goal of this work was to focus due to the regulatory effect exerted by the intronic miRNA on its host genes it ultimately drives to the development of tumor. We also took a wider look at the intronic miRNA, and their host gene pairs who have a correlation in cancer prognosis with immune cells.

We extracted data of CESC tumor cells' whole genome transcriptomes from the GEO dataset. Then from their we grouped the deregulated genes in the following way genes which have a log2Fold value 0.8 and over were classified as upregulated genes and those having negative 0.8 and below were classified as downregulated genes while the others as non-significant having a value of p-value 0.05 or more. After that we

tried out to analyse the significant association of the DE genes and DE miRs with the KEGG pathway where most of the genes were seen to have highly significance with various biological pathway like TRAIL signalling pathway, IGF1 pathway, Endothelins etc. and cell cycle, DNA repair, pathways in cancer, RNA transport, Wnt signalling pathways etc respectively. Next we analyse the enrichment for seeing the statistical significance of DE genes with the hallmarks of cancer where it showed highly significance with all the cancer hallmarks. So, from here we can say that these groups of genes share common biological function, chromosomal location, or regulation with that of cancer. MicroRNAs which are intronic in nature and they have transcription factors embedded in their host genes are filtered. From here, we found the pair *KLF7* and *hsa-mir-2355* were both significantly expressed and survived in accordance with the CESC data. Concisely, intronic miRNAs were significantly associated with the KEGG module and were moderately related with some pathways namely ErbB receptor signalling network, IFN-gamma pathway, S1P1 pathways etc. This draws us to the fact that the DE intronic miRNAs plays a role in regulation of those biological processes.

The common targets of the pair *KLF7* and *hsa-mir-2355* were shortlisted and a general network analysis along with expression data among the common targets were observed. The association with the KEGG pathways and GOBP pathway showed they were mostly moderately significant with the pathways like signal transduction, negative regulation of cell differentiation, positive regulation of cell migration, apoptotic process etc. which was a little unexpected. In case of the hallmarks, they were highly associated with the cancer hallmarks of CHG than that of GSEA hallmarks namely sustaining proliferative signalling, inducing angiogenesis, resisting cell death,

activating invasion and metastasis etc. which indicates their regulating role in the cancer prognosis.

Next, upon estimation of a plot of the tumor infiltration rate in the four tumor samples of our dataset we found some unexpected results in all the samples.While other research has found that practically every immune cell type, including T cells, B cells, NK cells, NKT cells, basophils, neutrophils, and myeloid dendritic cells (mDCs), plays a role in cancer progression regulation [87]. Here the CD8+ cells, CD4+ cells, macrophages, neutrophil, myeloid dendritic cells and B cells were infiltrated almost in all cells to some extent in all the tumor samples according to various estimation software. But the infiltration of NK cells and monocytes were very less which was expected to be higher. Next, after observation the proportion of the immune cells in those tumor samples was found where the proportion of CD4+ and macrophages was in greater amount almost in all the samples and monocytes and myeloid dendritic cells were in lower amount. The presence of these specific immune cells indicates their regulatory role in cancer development.

Inflammation and immune evasion are thought to be hallmarks of cancer growth, implying that immune cells are directly involved [88]. In the next analysis part, we found out the survival associated genes which are related to CESC and got 25 genes which are highly associated with the patient’s survival. According to the hazard ratio, the expression of *SAP30L*, *KLF1* and *AGFG2* if expressed in highly then the patient will survive less whereas if they are expressed less then the patient will survive more.

Innate immune signalling is vital for inflammation, and disruption of this pathway's signalling components is becoming more widely recognized as a key mediator in cancer development and progression [89]. Taking this into account, we discovered

that among these 25 genes, four activate immune signaling genes and are associated with patient survival: *ATG12*, *KRAS*, *PRKAR1A*, and *REL*. Although certain innate signalling mediators are abundant in myeloid malignancies, it is unknown how the activation of innate immune pathways contributes to the promotion or maintenance of leukaemia [90]. In our study we saw the activation of innate immune cells in the promotion and prognosis in CESC patients. Following the mapping of patient survival in CESC, we showed that CD4+ is significantly important with the patient survival rate of CESC patients. Besides among the four genes which can induce immune cells only *REL* is positively correlated with CD4+ cells recruitment. The prevalence of CD4+ in our four samples was also substantially greater, indicating that it plays a key role in the regulation in cervical cancer. So here we came to the conclusion that *REL* is the survival gene in CESC which is associated with the patient survival of the cervical cancer.

Chapter 5

Conclusion

Cervical cancer is one of the most often diagnosed malignancies in women globally. While early identification of cervical cancer is associated with improved survival, screening CESC is difficult in impoverished countries with a high prevalence.Thus, a method for testing and an adequate medication can fully cure cervical cancer is very difficult to adopt. This computational study therefore helps to establish the correlations between immune cells and intronic miRNA genes along with their host genes that are involved in cancer forecasting, in a comprehensive research connected to cancer-immunology.

The simultaneous expression of intronic miRNA with their host genes may help pave the way for developing cancer treatments. Similarly to how targeting the *REL* gene can help divert **CESC** away, we can help prevent a plethora of additional cancers by utilizing bioinformatics and computational analysis, as well as greater throughput in their study.

References

1. Waggoner, S.E., *Cervical cancer.* The Lancet, 2003. **361**(9376): p. 2217-2225.
2. Eddy, D.M.J.A.o.I.M., *Screening for cervical cancer.* 1990. **113**(3): p. 214-226.
3. Brameier, M. and C. Wiuf, *Ab initio identification of human microRNAs based on structure motifs.* BMC Bioinformatics, 2007. **8**(1): p. 478.
4. *Intronic MicroRNA: Discovery and Biological Implications.* 2007. **26**(4): p. 195-207.
5. Reddy, K.B., *MicroRNA (miRNA) in cancer.* Cancer Cell International, 2015. **15**(1): p. 38.
6. Iorio, M.V. and C.M. Croce, *MicroRNA dysregulation in cancer: diagnostics, monitoring and therapeutics. A comprehensive review.* EMBO molecular medicine, 2012. **4**(3): p. 143-159.
7. Kim, Y.-K. and V.N. Kim, *Processing of intronic microRNAs.* 2007. **26**(3): p. 775-783.
8. Ramalingam, P., et al., *Biogenesis of intronic miRNAs located in clusters by independent transcription and alternative splicing.* 2014. **20**(1): p. 76-87.
9. Si, W., et al., *The role and mechanisms of action of microRNAs in cancer drug resistance.* Clinical Epigenetics, 2019. **11**(1): p. 25.
10. Saito, Y., et al., *Epigenetic therapy upregulates the tumor suppressor microRNA-126 and its host gene EGFL7 in human cancer cells.* Biochemical and Biophysical Research Communications, 2009. **379**(3): p. 726-731.
11. Hu, X., et al., *A MicroRNA Expression Signature for Cervical Cancer Prognosis.* 2010. **70**(4): p. 1441-1448.
12. Manickam, A., M. Sivanandham, and I.L. Tourkova, *Immunological role of dendritic cells in cervical cancer.* Adv Exp Med Biol, 2007. **601**: p. 155-62.
13. Alves, J.J.P., et al., *Th17 response in patients with cervical cancer (Review).* Oncol Lett, 2018. **16**(5): p. 6215-6227.
14. Bartel, D.P.J.C., *Metazoan micrornas.* 2018. **173**(1): p. 20-51.
15. Kim, V.N., J. Han, and M.C.J.N.r.M.c.b. Siomi, *Biogenesis of small RNAs in animals.* 2009. **10**(2): p. 126-139.
16. Fabian, M.R., N. Sonenberg, and W.J.A.r.o.b. Filipowicz, *Regulation of mRNA translation and stability by microRNAs.* 2010. **79**: p. 351-379.
17. Bartel, D.P.J.c., *MicroRNAs: genomics, biogenesis, mechanism, and function.* 2004. **116**(2): p. 281-297.
18. Khvorova, A., A. Reynolds, and S.D.J.C. Jayasena, *Functional siRNAs and miRNAs exhibit strand bias.* 2003. **115**(2): p. 209-216.
19. Pu, M., et al., *Regulatory network of miRNA on its target: coordination between transcriptional and post-transcriptional regulation of gene expression.* Cellular and Molecular Life Sciences, 2019. **76**(3): p. 441-451.
20. Kozomara, A., M. Birgaoanu, and S.J.N.a.r. Griffiths-Jones, *miRBase: from microRNA sequences to function.* 2019. **47**(D1): p. D155-D162.
21. Kumar, S. and P.H.J.B.e.B.A.-M.B.o.D. Reddy, *Are circulating microRNAs peripheral biomarkers for Alzheimer's disease?* 2016. **1862**(9): p. 1617-1627.
22. Fromm, B., et al., *A uniform system for the annotation of vertebrate microRNA genes and the evolution of the human microRNAome.* 2015. **49**: p. 213-242.

23. Wang, J.-Y. and L.-J. Chen, *The role of miRNAs in the invasion and metastasis of cervical cancer.* Bioscience reports, 2019. **39**(3): p. BSR20181377.
24. Rodriguez, A., et al., *Identification of mammalian microRNA host genes and transcription units.* 2004. **14**(10a): p. 1902-1910.
25. Lin, S.-L., J.D. Miller, and S.-Y. Ying, *Intronic microRNA (miRNA).* Journal of biomedicine & biotechnology, 2006. **2006**(4): p. 26818-26818.
26. Cai, X.J.H.m.a.p.f.c., polyadenylated transcripts that can also function as mRNAs. RNA, *Hagedorn CH, Cullen BR.* 2004. **10**: p. 1957-1966.
27. Baskerville, S. and D.P.J.R. Bartel, *Microarray profiling of microRNAs reveals frequent coexpression with neighboring miRNAs and host genes.* 2005. **11**(3): p. 241-247.
28. Barik, S.J.N.a.r., *An intronic microRNA silences genes that are functionally antagonistic to its host gene.* 2008. **36**(16): p. 5232.
29. Liu, B., et al., *Interplay between miRNAs and host genes and their role in cancer.* Brief Funct Genomics, 2018. **18**(4): p. 255-266.
30. Lee, Y., et al., *MicroRNA maturation: stepwise processing and subcellular localization.* 2002. **21**(17): p. 4663-4670.
31. Pawlicki, J.M. and J.A.J.T.J.o.c.b. Steitz, *Primary microRNA transcript retention at sites of transcription leads to enhanced microRNA production.* 2008. **182**(1): p. 61-76.
32. Bernstein, E., et al., *Role for a bidentate ribonuclease in the initiation step of RNA interference.* 2001. **409**(6818): p. 363-366.
33. Okamura, K., et al., *The mirtron pathway generates microRNA-class regulatory RNAs in Drosophila.* 2007. **130**(1): p. 89-100.
34. Ma, N., et al., *Coexpression of an intronic microRNA and its host gene reveals a potential role for miR-483-5p as an IGF2 partner.* Mol Cell Endocrinol, 2011. **333**(1): p. 96-101.
35. Dill, H., et al., *Intronic miR-26b controls neuronal differentiation by repressing its host transcript, ctdsp2.* Genes Dev, 2012. **26**(1): p. 25-30.
36. Gromak, N., *Intronic microRNAs: a crossroad in gene regulation.* Biochemical Society Transactions, 2012. **40**(4): p. 759-761.
37. Han, J., et al., *Posttranscriptional crossregulation between Drosha and DGCR8.* Cell, 2009. **136**(1): p. 75-84.
38. Shukla, G.C., J. Singh, and S. Barik, *MicroRNAs: Processing, Maturation, Target Recognition and Regulatory Functions.* Molecular and cellular pharmacology, 2011. **3**(3): p. 83-92.
39. Gromak, N., *Intronic microRNAs: a crossroad in gene regulation.* Biochem Soc Trans, 2012. **40**(4): p. 759-61.
40. Hinske, C., et al., *Alternative Polyadenylation Allows Differential Negative Feedback of Human miRNA miR-579 on Its Host Gene ZFR.* PLOS ONE, 2015. **10**: p. e0121507.
41. Danin-Kreiselman, M., C.Y. Lee, and G. Chanfreau, *RNAse III-mediated degradation of unspliced pre-mRNAs and lariat introns.* Mol Cell, 2003. **11**(5): p. 1279-89.
42. Gruber, A.R., et al., *Means to an end: mechanisms of alternative polyadenylation of messenger RNA precursors.* Wiley Interdiscip Rev RNA, 2014. **5**(2): p. 183-96.
43. Mehta, A. and D. Baltimore, *MicroRNAs as regulatory elements in immune system logic.* Nat Rev Immunol, 2016. **16**(5): p. 279-94.

44. Squadrito, M.L., et al., *MicroRNA-mediated control of macrophages and its implications for cancer.* Trends Immunol, 2013. **34**(7): p. 350-9.
45. Trotta, R., et al., *miR-155 regulates IFN-γ production in natural killer cells.* Blood, 2012. **119**(15): p. 3478-85.
46. Rodriguez, A., et al., *Requirement of bic/microRNA-155 for normal immune function.* Science, 2007. **316**(5824): p. 608-11.
47. Kuo, G., C.-Y. Wu, and H.-Y. Yang, *MiR-17-92 cluster and immunity.* Journal of the Formosan Medical Association, 2019. **118**(1, Part 1): p. 2-6.
48. Mi, S., et al., *MicroRNA expression signatures accurately discriminate acute lymphoblastic leukemia from acute myeloid leukemia.* 2007. **104**(50): p. 19971-19976.
49. Jin, P., R.S. Alisch, and S.T. Warren, *RNA and microRNAs in fragile X mental retardation.* Nat Cell Biol, 2004. **6**(11): p. 1048-53.
50. Eberhart, D.E., et al., *The fragile X mental retardation protein is a ribonucleoprotein containing both nuclear localization and nuclear export signals.* Hum Mol Genet, 1996. **5**(8): p. 1083-91.
51. Schmitt, D.C., et al., *ErbB2-intronic MicroRNA-4728: a novel tumor suppressor and antagonist of oncogenic MAPK signaling.* Cell Death & Disease, 2015. **6**(5): p. e1742-e1742.
52. Lin, S.L., et al., *A novel RNA splicing-mediated gene silencing mechanism potential for genome evolution.* Biochem Biophys Res Commun, 2003. **310**(3): p. 754-60.
53. Takahashi, R.-u., H. Miyazaki, and T. Ochiya, *The Roles of MicroRNAs in Breast Cancer.* 2015. **7**(2): p. 598-616.
54. Barrett, T., et al., *NCBI GEO: mining millions of expression profiles—database and tools.* 2005. **33**(suppl_1): p. D562-D566.
55. Love, M.I., W. Huber, and S. Anders, *Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2.* Genome Biology, 2014. **15**(12): p. 550.
56. Team, R.C., *R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Version 4.0. 3*. 2020.
57. Wickham, H.J.W.I.R.C.S., *ggplot2.* 2011. **3**(2): p. 180-185.
58. Wickham, H., et al. *Dplyr*. in *useR! Conference*. 2014.
59. Hinske, L.C., et al., *miRIAD—integrating microRNA inter-and intragenic data.* 2014. **2014**.
60. Oliveros, J.C.J.h.b.c.c.e.t.v.i.h., *VENNY. An interactive tool for comparing lists with Venn Diagrams.* 2007.
61. Tang, Z., et al., *GEPIA2: an enhanced web server for large-scale expression profiling and interactive analysis.* 2019. **47**(W1): p. W556-W560.
62. Chandrashekar, D.S., et al., *UALCAN: a portal for facilitating tumor subgroup gene expression and survival analyses.* 2017. **19**(8): p. 649-658.
63. Karagkouni, D., et al., *DIANA-TarBase v8: a decade-long collection of experimentally supported miRNA–gene interactions.* 2018. **46**(D1): p. D239-D245.
64. Wong, N. and X.J.N.a.r. Wang, *miRDB: an online resource for microRNA target prediction and functional annotations.* 2015. **43**(D1): p. D146-D152.
65. Zhang, Q., et al., *hTFtarget: a comprehensive database for regulations of human transcription factors and their targets.* 2020. **18**(2): p. 120-128.
66. Zheng, R., et al., *Cistrome Data Browser: expanded datasets and new tools for gene regulatory analysis.* 2019. **47**(D1): p. D729-D735.

67. Szklarczyk, D., et al., *The STRING database in 2017: quality-controlled protein–protein association networks, made broadly accessible.* 2016: p. gkw937.
68. Kohl, M., S. Wiese, and B. Warscheid, *Cytoscape: software for visualization and analysis of biological networks*, in *Data mining in proteomics*. 2011, Springer. p. 291-303.
69. Perez-Llamas, C. and N.J.P.o. Lopez-Bigas, *Gitools: analysis and visualisation of genomic data using interactive heat-maps.* 2011. **6**(5): p. e19541.
70. Pathan, M., et al., *FunRich: An open access standalone functional enrichment and interaction network analysis tool.* 2015. **15**(15): p. 2597-2601.
71. Kanehisa, M. and S.J.N.a.r. Goto, *KEGG: kyoto encyclopedia of genes and genomes.* 2000. **28**(1): p. 27-30.
72. Subramanian, A., et al., *GSEA-P: a desktop application for Gene Set Enrichment Analysis.* 2007. **23**(23): p. 3251-3253.
73. Zhang, D., et al., *CHG: a systematically integrated database of cancer hallmark genes.* 2020. **11**: p. 29.
74. Thissen, D., et al., *Quick and easy implementation of the Benjamini-Hochberg procedure for controlling the false positive rate in multiple comparisons.* 2002. **27**(1): p. 77-83.
75. Lynn, D.J., et al., *InnateDB: facilitating systems-level analyses of the mammalian innate immune response.* 2008. **4**(1): p. 218.
76. Boterenbrood, A., *TIMER 2.0 and the price of oil: Addressing the mismatch between TIMER oil simulations and historic world oil developments between 2007 and 2016*. 2016.
77. Li, T., et al., *TIMER: a web server for comprehensive analysis of tumor-infiltrating immune cells.* 2017. **77**(21): p. e108-e110.
78. Chen, B., et al., *Profiling tumor infiltrating immune cells with CIBERSORT*, in *Cancer systems biology*. 2018, Springer. p. 243-259.
79. Finotello, F., et al., *quanTIseq: quantifying immune contexture of human tumors.* 2017: p. 223180.
80. Aran, D., Z. Hu, and A.J.J.G.b. Butte, *xCell: digitally portraying the tissue cellular heterogeneity landscape.* 2017. **18**(1): p. 1-14.
81. Petitprez, F., et al., *Transcriptomic analysis of the tumor microenvironment to guide prognosis and immunotherapies.* 2018. **67**(6): p. 981-988.
82. Vincent, J.-L., et al., *The prevalence of nosocomial infection in intensive care units in Europe: results of the European Prevalence of Infection in Intensive Care (EPIC) Study.* 1995. **274**(8): p. 639-644.
83. Weingaertner, I.R., S. Koutnik, and H.J.P.o. Ammer, *Chronic morphine treatment attenuates cell growth of human BT474 breast cancer cells by rearrangement of the ErbB signalling network.* 2013. **8**(1): p. e53510.
84. Gonzalvez, F. and A.J.O. Ashkenazi, *New insights into apoptosis signaling by Apo2L/TRAIL.* 2010. **29**(34): p. 4752-4765.
85. Goldstein, B.J., et al., *Role of insulin-induced reactive oxygen species in the insulin signaling pathway.* 2005. **7**(7-8): p. 1021-1031.
86. Stacker, S.A. and M.G.J.C.j.o.c. Achen, *The VEGF signaling pathway in cancer: the road ahead.* 2013. **32**(6): p. 297.
87. Biswas, Subhra K., *Metabolic Reprogramming of Immune Cells in Cancer Progression.* Immunity, 2015. **43**(3): p. 435-449.

88. Hanahan, D. and Robert A. Weinberg, *Hallmarks of Cancer: The Next Generation.* Cell, 2011. **144**(5): p. 646-674.
89. Rhyasen, G.W. and D.T. Starczynowski, *IRAK signalling in cancer.* British Journal of Cancer, 2015. **112**(2): p. 232-237.
90. Guillamot, M. and I. Aifantis, *Splicing the innate immune signalling in leukaemia.* Nature Cell Biology, 2019. **21**(5): p. 536-537.