Elucidating The Functional Roles of Bidirectionally Transcribed Genes in Immunobiology of Cancers

By

Sheikh Shafin Ahmad
17136009

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 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:

Sheikh Shafin Ahmad
Student ID: 17136009

Approval

The thesis/project titled "**Elucidating the Functional Roles of Bidirectionally Transcribed Genes in Immunobiology of Cancers**" submitted by
Sheikh Shafin Ahmad (Student ID: 17136009)

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 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 fulfill 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, **Md. Abdullah-Al-Kamran Khan** and **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, **Tahsin Anjum Farah** for being with me throughout the study and for her endless help and also **MD Tawsif Ur Rashid, Sajid Mehran Billah, Mayesha Noor** for always encouraging me in all means. Honestly, without them, I would have crumbled during the Covid pandemic period.

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.

Sheikh Shafin Ahmad

ID: 17136009

Abstract

In the realm of correlated regulations in genomic context, bidirectional genes have a spatial and functional connection unlike any other. These are genes originating on the same genomic position, but on opposite strands. Usually sharing a common promoter of hundreds to thousands base pairs, the genes are functionally synchronized in levels of expression. Within the complex pathogenesis of cancer, concerted changes like these can lead to tumorigenesis, or result in tumor cell suppression. In a similar manner, low immune infiltration promotes cancer progression at earlier stages, but acts as an antagonist during metastasis. In this study, we have performed functional enrichment analysis of the bidirectional genes to identify associated cancers. We also looked for the correlation between bidirectional genes and the immune infiltration profiles of different white blood cells. Finally, we checked the differential expression of the disease associated genes in the tumor cells of selected cancer types. Analyzing 5,013 extracted bidirectional gene pairs, cancers like Colorectal Cancer, Low Grade Glioma, Skin Cutaneous Melanoma, Liver Hepatocellular Carcinoma, Kidney Renal Clear Cell Carcinoma could be associated with the deregulation of the pairs. *BCL2L12* and *IRF3* gene pair could be positively correlated in the prognosis of LGG with significant patient survival. Additionally, *PSMB9* and *TAP1* are highly expressed in SKCM and strongly correlated with patient survival. These findings can serve as crucial clues to direct future investigations in cancer immunology and therapeutics.

Keywords: Bidirectional Promoters; lncRNA; protein coding gene; Colorectal Cancer; Low Grade Glioma; Kidney Renal Clear Cell Carcinoma; Liver Hepatocellular Carcinoma; Melanoma.

Table of Contents

Table of Figures

List of Acronyms

COAD	Colorectal Cancer
LGG	Low Grade Glioma
KIRC	Kidney Renal Clear Cell Carcinoma
LIHC	Liver Hepatocellular Carcinoma
SKCM	Skin Cutaneous Melanoma
GDA	Gene Disease Associations
GO	Gene Ontology
GOBP	Gene Ontology Biological Process
GSEA	Gene Set Enrichment Analysis
KEGG	Kyoto Encyclopedia of Gene and Genomes

Chapter 1

Introduction

The functionality of organisms stems from their ability to pass down information through genetic codes in forms of 4 base pair nucleotides. Its complexity thrives upon the regulation of these codes in a particular fashion that facilitates multiple functions of the gene. Both prokaryotes and eukaryotes have adopted their own gene expression and coregulation by various arrangements of the genes as well as by the arrangements of the regulatory DNA elements like- promoters, operators, enhancers, terminators etc. and by their epigenetic modulations [1]. Of them, one interesting and evolutionary conserved genomic organization of the promoter element is the "**Bidirectional Promoter**".

A Promoter is said to be the region in a DNA sequence where proteins bind to initiate the transcription process to produce RNA. Located upstream of the DNA or towards the 5' region of the sense strand, promoters are usually about 100–1000 base pairs long. But when it comes to bidirectional promoter, the same sequence upregulates or downregulates two genes arranged in a head-to-head manner on the opposite strands of DNA; and are within 1000 bp of one another [2].

1.1 Characterization of bidirectional promoters

Bidirectional promoter placement facilitates co-regulation or co-expression in two opposite faced genes by following a certain number of factors like lack of TATA boxes and being both GC-rich and enriched in CpG islands [2], mirroring a sequence composition where Gs and Ts dominate on one side and Cs and As dominate the other [3]. The analyses made by Trinklein et al. suggests the presence of abundant genes whose 5′ ends are on the opposite strands and within 1000 bp, while being

not as much as on genes with 5′ ends on the same strand at this distance. So far, 11% of total genome (1352 pairs) have been identified in the human genome as bidirectional promoter [4].

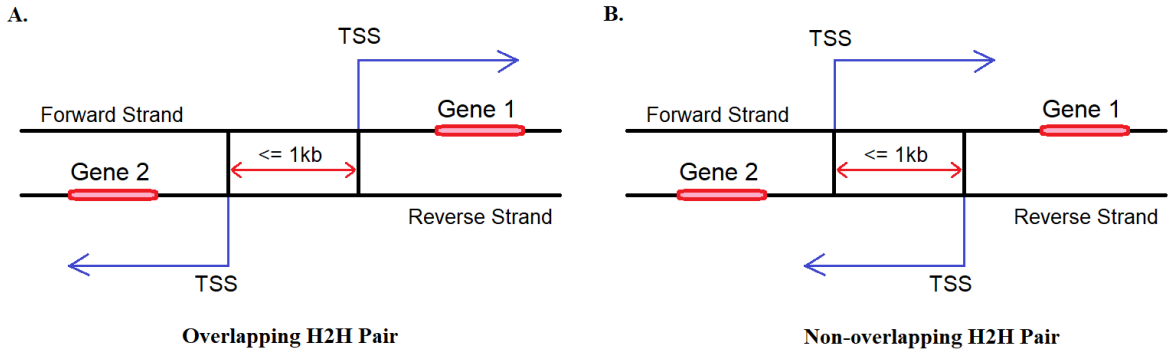


Figure 1: A. *Overlapping bidirectional promoter of head-to-head (H2H) gene pairs.*
B*. Non-overlapping bidirectional promoter of H2H gene pairs.*

A gene configuration where two adjacent genes are located on opposite strands of DNA and situated within 1 kb from the TSS is described as “Head-to-head” or H2H. And the sequences between an H2H gene pair (intra-H2H pair) are called bidirectional promoters [5]. It can be overlapping or non-overlapping bidirectional promoter of H2H gene pairs.

Additionally, bidirectional promoters are rich in GC, with a 66% median GC content with a constant presence of CpG islands near them [6].

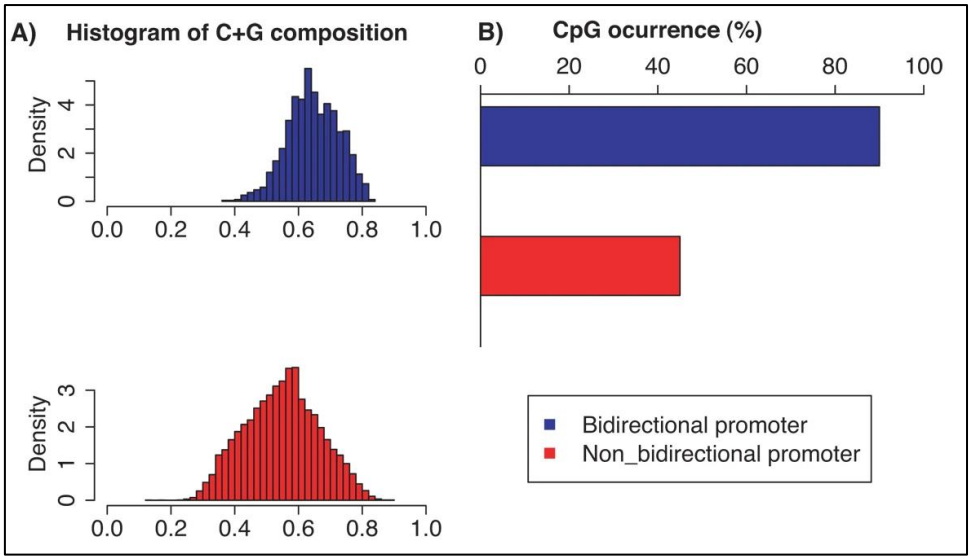


Figure 2: A. *Histogram plot of the GC-Content.* ***B.*** *CpG island comparison in between bidirectional and Unidirectional promoters. Adapted from [7].*

It is essential to understand the function that chromatin structure plays in the assembly of the transcription machinery and how chromatin is dynamically regulated in order to understand how bidirectional transcription functions [8]. In fact, bidirectional transcription is known to be an intrinsic feature of eukaryotic promoters. In DNA repair and other fundamental cellular pathways it is found to be preserved [9].

1.2 Bidirectional Promoter is Evolutionarily Conserved

The structure of bidirectional promoters is strongly conserved which suggests their functional significance. This conservation has been prevalent since the earliest times of gene discovery and it has been observed throughout evolutionary history [10]. For different kinds of species, some portion of the bidirectional promoter remained exactly the same and this kind of retained part is known as consensus sequences. Due to the presence of these consensus sequences, the functionality of associated genes and certain genome structures are notable [11]. Mainly because head-to-head genes tend to perform alike functions [10], which is rational considering the

significant expression correlation of evolutionary conservation of this gene organization. Therefore, head-to-head arrangement helps genes perform functions in the same pathway [12].The shuffling of the bidirectional relation between invertebrates and vertebrates indicates that, through co-regulation, the bidirectional structures are not conserved. Subsequent studies show that the preservation of bidirectional promoters may not seem to be the product of a functional relation between paired genes rather due to functional biasness at the whole genome level. This clearly shows that the conservation of the bidirectional structure involves genome-wide functional restrictions. Along with the selection of bidirectional structure, this functionality preference can intensify [13]. This biased organization, along with the divergence of tetrapod and teleost, was thought to have originated in mammalian ancestry. While non-mammalians have less bimodal arrangements in their genomes, they have a large number of genes in their genomes that are bidirectional. Moreover, the bidirectional promoter would enable the oppositely driven endogenously regulated gene to produce a new gene for this phenomenon, which will ultimately assist in a species' diversification and adaptation as it does not have the previous function [14].

Homology analysis showed that bidirectional organization has appeared in mammalian lineages more recently while the genes in the pairs are more ancient in nature. So, it is safe to say that the correlated genes drive this sort of organization through evolution. Other forms of transcription also facilitates the accumulation of this unique pattern, which includes the sharing of promoter in close vicinity and correlation of two distally positioned genes [15, 16].

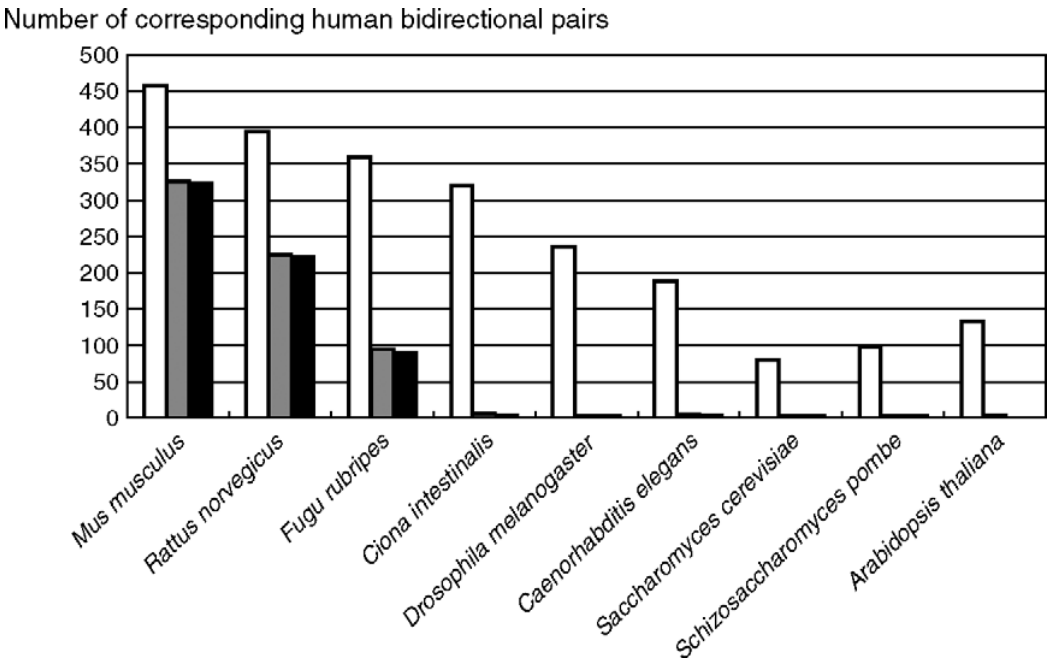


Figure 3: *Conservation of human bidirectional promoters in other eukaryotic genomes. White, grey, black consecutively represent the total homologs, homologs that are adjacent and adjacent genes in a bidirectional pattern respectively. Adapted from [16].*

Since the conservation of the promoter depends heavily on the functional choice of the genes, so genes with the same function as the DNA repair pathway genes and those in the MHC class tend to be close to each other for their mechanical function to be easily regulated. Especially the genes regulated by retained bidirectional promoters are mainly associated with specific cellular functions [13]. A study was conducted on the divergently transcribed lncRNA showed that lncRNA transcribed from Tbx5 is highly evolutionarily conserved. It exhibits a different sequence other than Tbx5, leading to embryonic death by knockdown. Furthermore, it is observed that bidirectional lncRNAs with control functions are enriched in haploinsufficiency genes, indicating that they have functional roles in dose-sensitive gene regulation [17].

Research has shown that CGCG elements developed in vertebrates and serve as an active component of CGI-related promoters. Bidirectional transcription is based

predominantly on the non-methylated form of the factor that was known to be a novel promoter [18]. The architecture of bidirectional promoters to enhance genetic circuit mutational stability can be applied in general to synthetic biology applications [19]. For synthetic biology with future use of gene therapy, compact bidirectional promoters are supposed to be beneficial. Bidirectional promoters are excellent at driving short RNA in transposing systems for sleeping beauty [14]. More research should rely on the bidirectional promoter's consensus sequence for further gene therapy treatment.

1.3 Genomic and epigenetic features of bidirectional promoters

Genomic alteration and epigenetic modifications are the reasons why the cells undergo unregulated continuous cell divisions and proliferation. These changes include failure to DNA damage repair, changes in the DNA sequences, failure of cell cycle regulation, silencing of the tumor suppressor genes, aberrant transcription etc. which ultimately leads to cancer formation [20]. Bidirectional promoters are quite common in DNA repair genes with a frequency of 40% [2]. Additionally, they have a high amount of GC content as opposed to regular promoters [4]. These promoters are said to be more resistant to methylation for protection of essential genes.

Like other promoters, the **CpG islands** are also discussed in bidirectional promoters. They are a short region of DNA in which the frequency of the GC sequence is comparatively less suppressed [21]. Hypermethylation of CpG islands in promoter regions usually results in gene silencing, and several tumor suppressor genes are hypermethylated in their promoter regions in cancers, which can lead to tumorigenesis [4, 22, 23]. Most bidirectional promoters organize the transcription regulation of a gene pair. Now if hypermethylation of CpG islands in such promoters can silence genes in both directions, then a single “hit” within these promoters could

potentially disable two tumor suppressor genes simultaneously. This could accelerate tumor development, according to the multiple hit theory of tumorigenesis [24].

The **Nucleosome depleted region (NDR**) is another site for transcription initiation which assembles the machinery for bidirectional transcription. It is a 80 to 300 bp long region in an active promoter, enclosed by two nucleosomes [25]. Transcription regulation can be controlled by either relocating the nucleosomes from 5' and 3' NDRs [26] or by modifying the size of the NDR [27]. Histone modification like deacetylation, chromatin remodeling plays a key role in controlling transcription regulation [8].

Various **histone markers** like H3K4me2, H3K4me3, H3K9ac, H3K27ac etc. which are located in bidirectional promoter region can also regulate transcription binding [5]. Studies found out H3K4me3 proteins highly promote transcription while markers like H3K27me3, H3K9me3 level are much higher in a silent promoter[28]. But Histone H4 acetylation is found to be less prevalent in the bidirectional promoters. [29].

1.4 Expression pattern of bidirectional promoter driven genes

Since sharing common promoter regions is prevalent for two neighboring genes, expression regulation may not always work in favor of both of the genes. Data collected by measuring transcriptional activity at the start of transcription show that RNA polymerase assembly and initiation occur in almost equal proportions for both the genes in both directions [30]. However, due to post-recruitment regulation, bidirectional transcription rate decreases as RNA polymerase II moves further away

from the promoter [31]. This likely causes an uncontrolled rate of expression regulation.

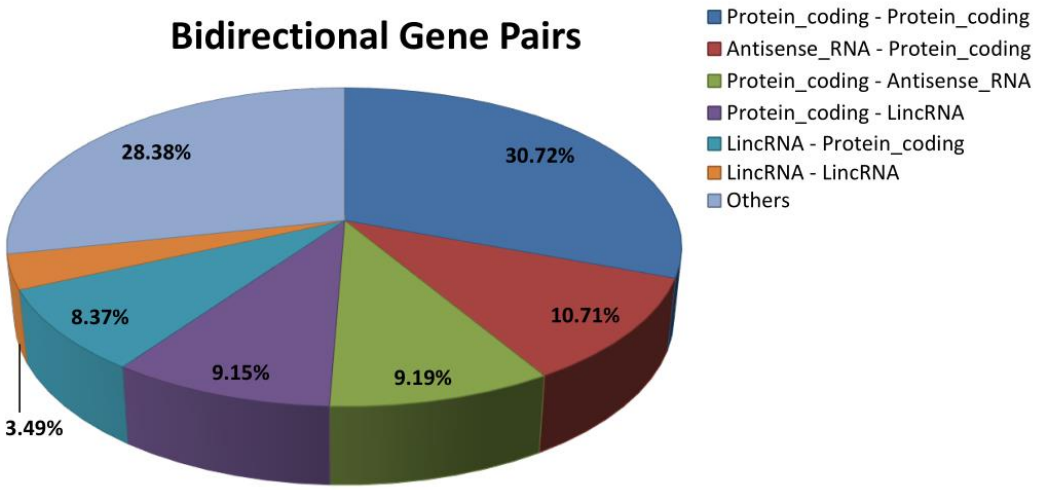


Figure 4: *Percentage of the different functional bidirectional gene pairs. Adapted from [32].*

Although promoters are generally capable of actively initiating transcription in two directions, in most cases productive elongation is seen primarily in one orientation only. Thus, a mechanism must exist to regulate transcription in both directions and thus dictate this asymmetry after bidirectional initiation. This could include:

- **Specific sequence signals** present in the promoter or coding region can lead to a cause. It has been proposed that the nucleotide composition around the promoter affects its bidirectionality [33].
- **Chromatin modifications.** Previous rounds of transcription could mark the orientation favored in subsequent rounds. One example of such epigenetic memory is the co-transcriptional trimethylation of H3K36 that recruits the deacetylase Rpd3S [34]. This deacetylase has been shown to not only repress spurious transcription within the coding region [35], but also to decrease the bidirectionality of a downstream promoter [31]. Another example could be ncRNA transcription in the

proximity of the promoter. This could influence the direction of transcription initiation by inducing chromatin remodeling favoring one orientation.

- **3D structure of transcription.** Transcriptional memory could be maintained by using spatial mechanisms, such as DNA looping, linking the promoter with the favored 30 end [36, 37].

The type of expressions that can be observed are briefly described below according to their preferred method, -

a. Shared Transcription Factors and Their Binding Sites in a Bidirectional Gene Pair

Co-Expression of the genes from common bidirectional pairs may involve both common transcription factors and their binding sites. Three probable cases can happen –

i. TFs will be shared in between the genes;
ii. TFs expressed from one gene will regulate the other;
iii. Different TFs will be used for the expression of two different genes.

A significant number of bidirectional gene pairs were found to use the same TFs to facilitate the transcription of both genes. Also, a significant relationship between the shared TFs and the expression correlations of the two genes was found as higher expression correlation was observed for a bidirectional gene pair that uses shared TFs. Even self-regulating gene pairs in which TFs from one gene regulating the expression of other gene was found and this pair showed even higher expression correlation values [38].

Moreover, it is observed that bidirectional promoter has the capacity to drive the expression of both the two reporter proteins, namely eGFP and mCherry

independent of orientation, but never both at the same time. This ultimately indicates the unbiased expression of both reporters by the intergenic region whereas in the case of promoters very negligible biasness is observed [9].

The chromatin landscape at and around the transcription regulatory region between the pair of bidirectional genes was found to modulate in conjunction with the transcriptional behaviour of each gene in the pair in a study of the *NUP26L-PIH1D3* bidirectional gene pair during the Retinoic Acid mediated differentiation of embryonic carcinoma cells. It was thus seen that the expression profile of such genes matched the histone modification profile of marks correlated with successful transcription initiation and elongation for the whole spectrum of intergenic distance separating bidirectional genes [9].

A total of 112 H2H genes targeted by up-regulated eRNAs tended to have significantly higher expression levels in cancer compared with normal, while 69 H2H genes targeted by down-regulated eRNAs showed significantly lower expression levels in cancer than in normal, confirming the role of enhancers in regulating H2H genes expression [39].

b. Functional Similarities of Genes in a Bidirectional Gene Pairs

The genes residing in bidirectional gene pairs are functionally similar resulting in high co-expression and are regulated by shared TFs or by self-regulation. This is observed by analyzing the GO terms of the genes and higher expression correlation was found to be related to the similar function of the genes in a biological pathway [38]. Annotated gene pairs used the GO terms for three different subsystems namely, BP (Biological Process), MF (Molecular Function), CC (Cellular Component) [40]. This helped conclude the existence of similarities of the subsystems.

Genes which function in DNA repair pathways; chromatin maintenance, stability, assembly are over-represented as instability, and failed repair of DNA most likely lead to cancers [41]. This occurs due to the CtBP (C-terminal binding protein) regulated DNA methylation of binding sites of TFs like- GABPA which is enriched in bidirectional promoter regions of the DNA repair genes and tumor suppressor genes resulting in the repression of the gene expression [29].

Also, many factors linked with the CtBP, are transcription factors that are highly enriched in the bidirectional promoters; which leads to the transcriptional regulation of bidirectional promoters of DNA repair genes like BRCA1 [42].

c. Regulation of Transcriptional Direction from a Bidirectional Promoter

Regulation of the direction of transcription can be controlled at different levels like initiation, elongation, termination using various regulatory mechanisms. RNA polymerase II interacts with different chromatin modification and RNA pol II C-terminal domain (CTD) phosphorylation in order to move along [8].

In early elongation step, rapid CTD phosphorylation on Ser5 could lead to the termination of the process due to the interaction with various termination machinery like (Nrd1-Nab3 complex) as well as a hypothetical 5' checkpoint can lead to the selection of either elongation or termination of the process; thus the directionality can be regulated by the termination of the process [15]. But if RNA pol II somehow crosses the 5' checkpoint, then the transcription elongation will occur due to the decreased level of Ser5P and increase the level of Ser2P which can co-transcriptionally regulate the chromatin structure [26]. Also, expression can be post-transcriptionally regulated by regulating the stability of the transcripts [43].

The resultant transcripts are also involved in the regulation process; as various short transcripts of ncRNA can have a functional role to regulate the transcription from its own bidirectional promoters or other protein coding genes which can be its neighbor or even distant [8]. PROMPTs (Promoter Upstream Transcripts) which are small noncoding RNAs with 5'-cap and 3'-adenosine tails [44] can control the expression of the oppositely oriented transcription from the bidirectional promoter [45].

Also, both repression via the ncRNAs and activation via the tissue-specific promoter associated ncRNAs (pancRNA) have been establishing and the major source of these sort of transcripts is the bidirectional promoter mediated transcription [46, 47].

1.5 Associations of bidirectional promoters with different cancers

Cancer is formed when the cells undergo unregulated continuous cell divisions and proliferate enormously due to certain changes in the DNA sequences, failure to DNA damage repair, failure of cell cycle regulation, silencing of the tumor suppressor genes, aberrant transcription etc. All these changes are mostly caused due to both genetic (e.g., DNA damage, mutations) and epigenetic modifications (e.g., DNA methylation, histone modifications etc.) [20].

Since it has been known already that bidirectional promoters maintain a unique transcriptional regulation mechanism to maintain its directionality so any changes in it will result in excess transcriptional activation or repression. This may ultimately lead to tumorigenesis [47]. One such example is seen recently where two adjacent genes whose transcription start sites are neighboring and directed away from each other can form bidirectional gene pairs and have the potential to participate in the development of cancer. The two human oncogenes, *PLGAL2* and *POFUT1* are jointly regulated by an evolutionary bidirectional promoter that has been shown to

lead to a strongly positive colorectal cancer association. Mechanistically, a true bidirectional promoter can boost self-renewal and hinder the differentiation of colorectal cancer stem cells facilitating cancer progression instead of a ceRNA mechanism [48].

Epigenetic variations such as DNA methylation are the core features of cancer, aside from genetic changes. Cancer development is triggered by hyperactivation of growth associated genes or silencing of tumor suppressor genes or DNA repair genes by activity or suppression alteration of the important histones. The most common epigenetic alteration observed in the tumorigenesis system is DNA hypermethylation. Numerous studies have been reported demonstrating that cytosine hypermethylation results in the silencing of the tumor suppressor genes in the CpG islands of the bidirectional promoters. This contributes significantly to tumor development and malignant transformation [20]. This is reported in an experiment that the methylation of CpGs in the CGCG element, for instance, suppresses the function of the promoter. Nearly, 80% of the CpG region of the genome is methylated. CGCG elements in other genome regions would be more methylated, causing transcriptional silencing. However, theoretically it has been seen that DNA methylation of CGCG components could potentially shield the genome from erroneous transcription [18].

Another such example is observed in case of the *BRCA1* gene. The *BRCA1* gene, which is mainly recognized for its involvement in the repair of DNA damage and tumor suppression function, and the *NBR2* gene, which is a lncRNA with a tumor suppressive role, was found to share a similar 218bp bidirectional promoter that produces breast or ovarian cancers if hypermethylated [49, 50]. Table 1 shows

different ovarian cancer genes that are transcribed from bidirectional promoters and the hypermethylation for their aberrant expression pattern in cancer cells.

Table 1: *Bidirectional genes regulated in Ovarian cancer. Adapted from [42].*

Ovarian Cancer Genes	Bidirectional Partner	Record of Aberrant Methylation
BARD1	*DA865307*	No Evidence
BRCA1	*NBR2*	Yes (Wilcox et. al., 2005)
BRCA2	*DR731263*	Yes (Dhillon et. al., 2004)
CHK2	*HSC20*	Yes (Zhang et. al., 2004)
HER2/ERBB2	*PERLD1*	Yes (Fiegl et. al., 2006)
TP53	*AK001247*	Yes (Amatya et. al., 2005)
FANCA	*SPIRE2*	No Evidence
FANCB	*MOSPD2*	No Evidence
FANCD2	*BC043599*	No Evidence
FANCF	*GAS2*	Yes (Dhillon et. al., 2004)

By comparison, a recent study indicates that the role of p300 lysine acetyltransferase and *KDAC1* in the regulation of the promoter is shown to include DNA methylation of the CpG dinucleotides in the promoter. *HDAC1* recruitment to the promoter portion by the corepressor CtBP decreases histone acetylation and *BRCA1* expression, while the activity of its bidirectional promoter in MCF-7 cells is reversed by estrogen induction or *HDAC* inhibition [51].

The hypermethylation of bidirectional promoters in cancer is shown to effectively silence both genes of the pair [11]. The coherent Expression Correlation of H2H Pairs and Their Differential Co-expression in Cancer is observed where "*METTL4-NDC80*," "*C1orf109-CDCA*," and "*TMEM60-PHTF2*" was seen to consistently play protective roles across multiple types of cancer, while pairs like "*ATAD2-WDYHV1*" and "*AURKA-CSTF1*" seemed to be associated with cancer progression and a worse survival across multiple cancer types [39].

The *TP53* gene and the *WRAP53/WDR79* gene share a similar bidirectional promoter that is shown to be hypermethylated in human gliomas as a consequence

of silencing the *TP53* gene expression and losing its activity to suppress the tumor [52, 53]. In some of the cancers, the anti-regulation of *TP53* expression by the *WRAP53* gene that codes for an antisense transcript can also be converted into a protein [54]. The complex gene *ANRIL* transcribes into a lncRNA which is found to be highly expressed in many cancer cells. It shares bidirectionality with some tumor suppressor genes, *p16-CDKN2A, p15-CDKN2B,* and *p14-ARF*. Compared to the other two tumor suppressor genes, the highest positive correlation was identified between *ANRIL* and *p14-ARF*[55]. While other previous research reveals that ANRIL overexpression was predominantly accompanied by *p16-CDKN2A/p15-CDKN2B/p14-ARF* locus transcriptional inactivation via cis direct interaction locus [56].

Two proapoptotic genes, *MAPK10* and *PTPN13* have a significant role in tumor suppression, differentiation, and proliferation [57] and anti-regulation of the *Her2/ErbB2* malignant transformation respectively [58]. They are found to share a bidirectional promoter consisting of 12 CpG islands which is methylated in non-Hodgkin's lymphoma, 50% of Hodgkin's lymphoma, breast cancer, gastric cancer, and hepatocellular carcinoma cells[59, 60]. Because of such methylation either complete or increased gene silencing of *MAPK10* and *PTPN13* is observed [61].

Hypermethylated state of the shared bidirectional promoter between the *PARK2* and *PACRG* genes results in silencing of these genes and inactivation of these genes has been observed in cervical cancer, lung squamous cell cancer, colorectal cancer, gastric cancer, skin cutaneous melanoma, lung adenocarcinoma, and endometrioid cancer [62]. *HSP60* and *HSP10* mitochondrial chaperonin genes are found to be upregulated in the IFN-γ induced astroglioma cells[63] which are responsible for mitochondrial protein homeostasis, including active folding of unfolded proteins and

ATP-dependent proteolysis of denatured or misfolded proteins [64, 65]. *Hsp90* is expressed at 2–10 fold higher levels in cancer cells compared to normal cells and is suggested to be one of the key factors implicated in promotion of cancer cell survival and metastases [66].

In the regulation of transcription initiation in the TATA-containing and TATA-less genes, the transcriptional regulatory factors named MINC have a distinct role, and transcription level and promoter shape showed a definite correlation between the TATA-containing and the TATA-less promoter. Here the study shows that TBP recruits MINC to suppress pervasive transcription and for the precise identification of bona fide TSSs MINC is essential, especially in promoters containing TATA, and histone methylation contribute to the repression of initiation of transcription in coding regions [67].

1.6 Impaired control of mitochondrial bidirectional genes in cancers

Mitochondria is an essential organelle in eukaryotic cells due to their involvement in energy generation. They have a circular genome containing genes encoding mitoribosomal proteins, tRNAs and rRNAs and a set of functional proteins required for energy production in the form of ATP [68]. However, almost 99% of mitochondrial proteins are encoded by the host nuclear genes. So, expression of mitochondrial protein-encoding genes are regulated similarly to genes that are present in the host nuclear chromosomes [69].

Of total bidirectional promoters, 31.6% contains at least one gene for mitochondria associated function and 4.82% contains both mitochondria related genes in the pair [70]. But several genomic DNA repair genes (e.g. TP53/WRAP53 and

APEX1/OSGEP gene pairs) that use a bidirectional promoter can also function in the repair mechanism of the damaged mitochondrial genome [71] and can also contribute in mitochondrial respiration [72]. They also contribute functionally as subunits of the mitochondrial ATP synthase, components of 28S, 39S subunit and NADH dehydrogenase, Mitochondrial aminoacyl- tRNA synthetases, oxoadipate carriers, import proteins etc.

Altered expression pattern of the mitochondria associated genes often found in cancer, including both glycolysis, oxidative phosphorylation related pathways and mTOR/AMPK pathways that vary from normal cells [73, 74].

Apart from this various neurological disorders like Alzheimer's disease, Parkinson's disease, and Huntington's disease are found to be associated with mitochondrial dysfunction [75] and several neuronal genes share characteristic bidirectional promoter [76]. *PINK1* (shared promoter with *PACRG*) which is responsible for the mitochondrial biogenesis induction and reduction of mitochondria induced apoptosis in neurons, is found associated with abnormally expressed in Parkinson's disease [77].

Mrps12 and *Sars2* mitochondrial genes sharing common bidirectional promoter are found in both mouse and human and it has binding preferences for the NF-Y transcription factor which has a role in cellular proliferation [78].

Mitochondrial chaperonin genes *HSP60* and *HSP10*, share a common bidirectional promoter, are found to be involved in various genetic disorders [79].

Apart from mitochondria associated diseases and cancer; genetic diseases like Down syndrome related two gene *DSCR4* and *DSCR8* are found to share an endogenous retroviral bidirectional promoter though their direct link to the phenotype

is yet to be discovered? [80]. Also, human aging-related genes SIRT3 which plays role in mitochondrial function activation, lipid metabolism and PSMD13 which is a proteasome protein subunit; are found to share a 788bp bidirectional promoter [81].

1.7 Role of bidirectional genes in immunobiology of various diseases

The main purpose of our immune system is to gain protection from illness and infection that virus, bacteria, fungi or parasites cause. They tend to get weakened when cancer spreads to parts of the body that are responsible for producing immune cells like bone marrow etc. Innate and acquired immune cells altogether comprising of Neutrophil, Macrophage, NK cells, B cell, T cell etc. are our primary defense mechanism for combatting various diseases including cancer. But cancers are not a mass of transformed cells. Rather they are a new organ composed of various non-malignant cells, fibroblasts, adipocytes, pericytes, vascular endothelial cells and, as main players, immune cells -altogether forming a large portion of the tumor mass, which deviated and lost its ability to maintain the tissue architecture [82].

The bidirectional promoters play a part in the activation or repression of various neighboring genes, controlling patient survival. One gene may contribute in the suppression of cancer formation, while its neighboring gene might enhance it and vice versa. Bidirectionality makes it such that, both will activate simultaneously or being mutually exclusive [83]. Their effect starts from irregulating the immune cells. A study by H. Kambara et al. stated that, interferon-stimulated gene *BST2* was regulated by *BISPR*, a lncRNA transcribed from a shared bidirectional promoter [84] and the IFN-stimulated genes (ISGs) are involved in various aspects of antiviral defense and immune modulatory functions [85]. Again, *BAL1* and *BBAP* genes are regulated by a bidirectional promoter and are overexpressed in Large B-Cell

Lymphomas [86]. The antiviral responses as a tumor suppressor inhibiting cell growth and promoting apoptosis is done with bidirectional promoter between genes *STAT1β* and *EGFP* [87].

1.8 Rationale

Bidirectional promoters and its head-to-head gene organization in the genome have a significant impact on the co-expression or anti-regulation between its neighboring genes. There is also a significant evolutionary history among the vertebrates or more specifically in the mammals, due to its regulation in basic housekeeping genes and some other important genes like tumor suppressor genes, DNA repair genes, growth and differentiation pathway related genes, mitochondrial function associated genes etc. Any changes in the regulatory pattern can cause severe dysfunctions in the cells.

Various studies have found significant interactions of bidirectional promoters with different cancers and other diseases in which some genes are either upregulated or downregulated. Cancer cells also recruit immune cells as discovered during analyzing the hallmarks in order to evade immune destruction via different reaction pathways. Previous studies have already pointed out how low immune infiltration can lead to cancer progression. But whether the bidirectionally transcribed genes have a role in the tumor immune infiltration is yet to be firmly established.

Therefore, in this study, we have targeted to elucidate the correlation between bidirectionally transcribing genes with immune infiltrating cells in different cancers to have a clearer understanding of the immune infiltration level whether these genes are guiding the cancer metastasis or assisting in tumor destruction in tumor microenvironment.

1.9 Hypothesis

In this study, we hypothesized that, -

i. Bidirectional promoter sharing genes influence each other by being up regulated or down regulated together during cancer prognosis.

ii. There is an active correlation between immune cells and gene expression resulting in tumorigenesis influenced by bidirectional genes.

1.10 Objectives

The specific objectives for this study are, -

i. Identification of the bidirectional promoters in the most recent version of the human genome database.

ii. Enrichment analysis of the bidirectional gene pairs to illuminate the functional roles of these genes.

iii. Predicting the correlation of the genes with patient survival in different types of cancer.

iv. Correlating bidirectional genes contributing to immune infiltration for the selected cancers.

Chapter 2

Methodology

As, the study mainly focuses on the associations of cancers with the bidirectionally transcribed genes; the initial steps are involved in identifying the bidirectional promoter regions; associating bidirectional promoter sharing genes with different cancer types on the basis of how they are mediated (up or down regulated) and correlating the gene expression in each cancer types with immune infiltration. Accordingly, the entire study is designed into the 5 stages as follows, -

i. Retrieval of the bidirectional promoters: Bidirectional promoter is defined by a common region shared between two head-to-head oriented genes and its size ranges withing 1kb. Firstly, from the genome browser *Ensembl 103, BioMart* which contained updated data till February 2021 [88], all the long non-coding RNA (lncRNA) and protein coding genes were collected for Human Genes (GRCh38.p13) along with their chromosome number and transcription start and end location for both the strands. Then using *BEDTools* [89], genes of opposite strands that share a common upstream region ranging within 1kb in size and with a minimum promoter sequence length of 100bp were selected along with the pair of genes and their functional types.

ii. Functional Enrichment Analysis of Bidirectional genes: Gene set enrichment analysis (GSEA) or functional enrichment analysis was done to identify the genes that suggest possibility of an association with disease phenotypes. The retrieved genes which are transcribed from the bidirectional promoters, were annotated by their functional roles in different cellular pathways modules notably KEGG [90], GOBP [91] by enrichment analysis using *Gitools-1.8.4* [92]. We also

mark the important cancer hallmarks for the study. During the enrichment analysis, we opted to use the multiple test correction approach of Benjamini-Hochberg's FDR (False Discovery Rate) to reduce the false-positive results. We only considered those enriched terms/pathways statistically significant which have a corrected p-value < 0.05.

iii. Identification of significant Cancer types to target: The Protein coding genes were checked for *DisGeNET* Enrichment Analysis using KEGG module of *Gitools-1.8.4* [92], whereas the lncRNAs were searched by their survival on various cancer diseases using lncSEA [93]. Common cancers were selected based on the presence of their responsible genes in bidirectional fashion. We only considered those enriched terms statistically significant which have a corrected p-value < 0.05.

iv. Survival Analysis: The survival analysis gave a comparative statistical study on the condition of patients for whom gene are highly expressed for a given time period. From Gepia2 [94], we check the survival log rank p-value for each of the protein coding gene with its corresponding cancer along, with their expression value. A Kaplan-Meier plot is constructed which compares the survival rate of two groups based on the corresponding gene expression [95]. Additionally, a box plot is also extracted from the same site following similar methodology to compare the tumor cell expression opposed to the normal cells [94]. For the lncRNAs, we utilized UALCAN [96] in order to derive our KM plot and box plot for the same purpose. From the survival analysis, we filtered and selected only those gene pairs where both of the genes have significant correlation with patient survival in the associated cancer type. We then searched for the correlation between the significant survival associated bidirectional genes (under the same promoter) using TIMER tool [97]. We only concluded those results as statistically significant which have p-values < 0.05.

v. Immune Infiltration

For our selected cancer types, we have identified the association between the tumor infiltrating immune cells with the prognosis of the cancer patients using “Immune Outcome” function of the TIMER2.0 [98] tool which incorporates six immune deconvolution algorithm models, namely- TIMER [97], CIBERSORT [99], quanTIseq [100], xCell [101], MCP-counter [102] and EPIC [103] to calculate the normalized coefficient of the infiltrate. Next, to obtain the correlation between the expression of a specific survival associated bidirectional gene and the individual immune infiltrating cells in a cancer-type, we utilized the “Gene Module” of TIMER [97] tool to find out the purity-corrected partial Spearman’s rho value and statistical significance in p-values. We took only those correlation values as statistically significant which have p-values < 0.05.

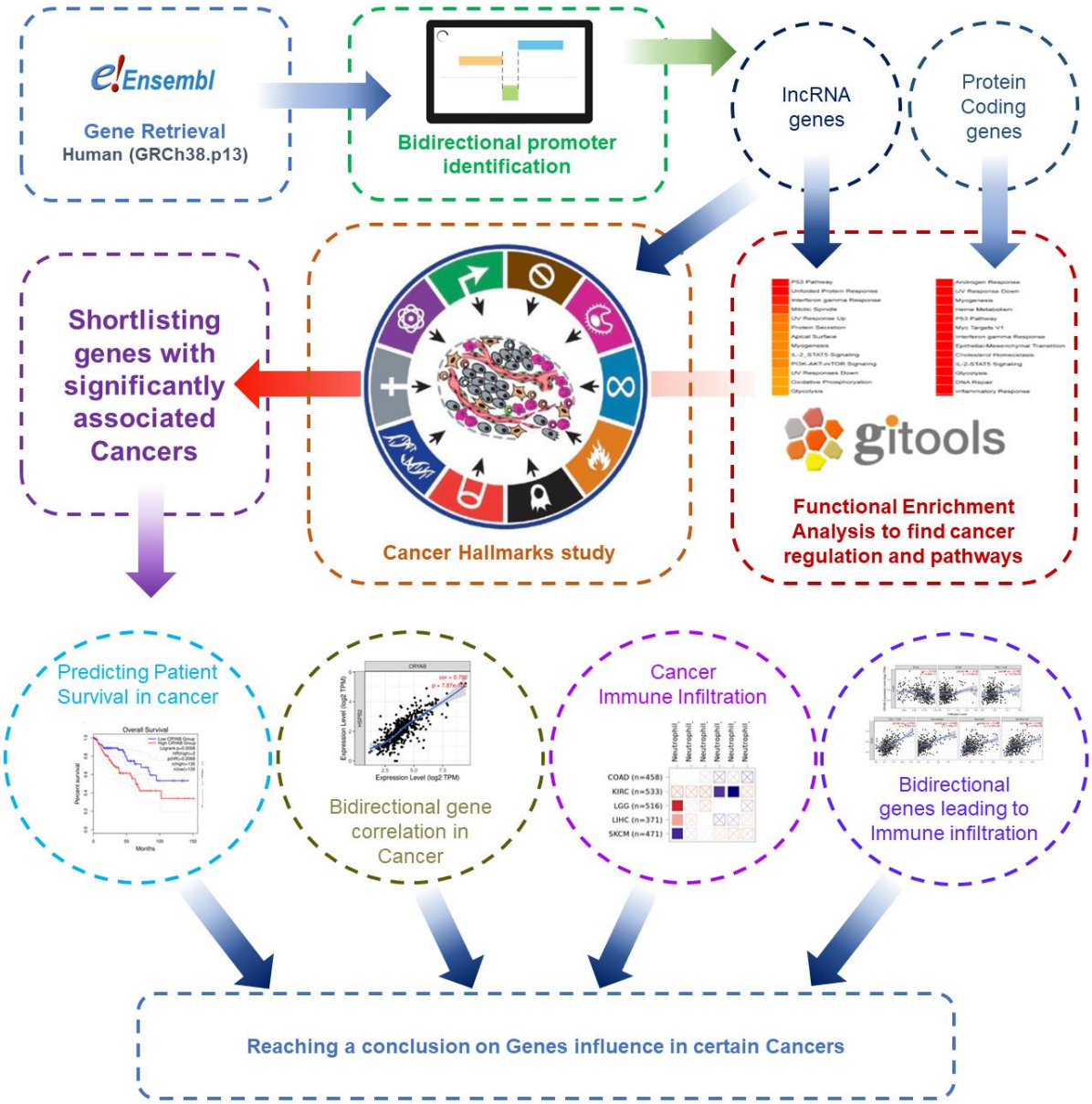


Figure 5: *Overview of the entire workflow*

Chapter 3

Results

3.1 Identification of Bidirectional Promoters and Bidirectional genes

The BEDTools intersect presented us with a list of genes sharing a common promoter, within 1000bp of each other and in the opposite strand whose promoter length is of 100bp (Figure 6).

Gene sequence 1						Gene sequence 2					
Chromosome	Strand	Start Sequence	End Sequence	Gene Name	Gene Type	Chromosome	Strand	Start Sequence	End Sequence	Gene Name	Gene Type
Chr1	+	826635	827635	*LINC01128*	lncRNA	Chr1	-	827522	828522	*LINC00115*	lncRNA
Chr1	+	58932643	58933643	*LINC01358*	lncRNA	Chr1	-	58931897	58932897	*LINC02777*	lncRNA
Chr3	+	72060061	72061061	*AC105265.2*	lncRNA	Chr3	-	72060242	72061242	*LINC00877*	lncRNA
Chr3	+	84880984	84881984	*LINC02025*	lncRNA	Chr3	-	84881679	84882679	*LINC00971*	lncRNA
Chr1	+	1274223	1275223	*LINC01786*	lncRNA	Chr1	-	1273853	1274853	*UBE2J2*	protein_coding
Chr2	+	174486512	174487512	*AC010894.1*	lncRNA	Chr2	-	174487029	174488029	*GPR155*	protein_coding
Chr2	+	214809229	214810229	*SNHG31*	lncRNA	Chr2	-	214809683	214810683	*BARD1*	protein_coding
Chr3	+	98901424	98902424	*AC091212.1*	lncRNA	Chr3	-	98901171	98902171	*DCBLD2*	protein_coding

Figure 6: *Extracted list of bidirectional gene pairs*

A total of 5,013 gene pairs were obtained out of which 401 were lncRNA-lncRNA pairs; 2993 pairs of lncRNA-protein coding pairs and 1619 were protein coding-protein coding genes pairs (Figure 7).

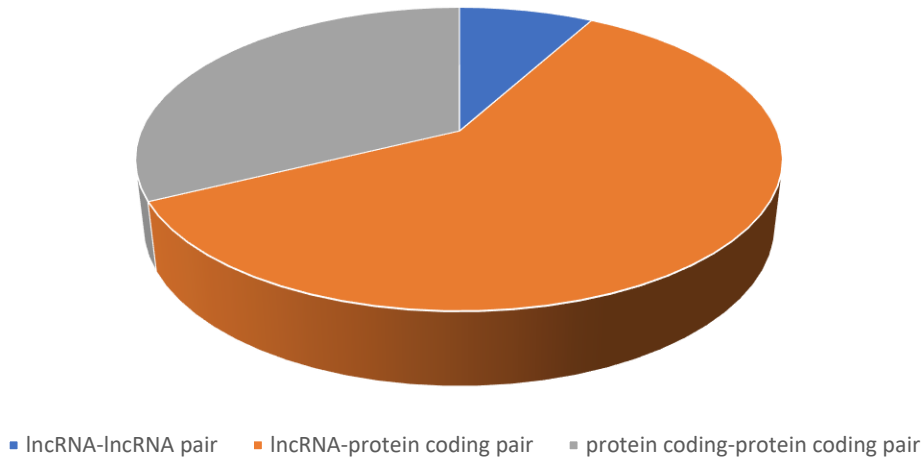


Figure 7: *Number of extracted bidirectional gene pairs*

3.2 Bidirectional gene pairs observed to be associated with different cancers

For the **protein coding genes** that we obtained already, DisGeNET enrichment analysis has been done to find correlation of the genes with certain diseases. Gene set enrichment analysis (GSEA) or functional enrichment analysis allowed us to know the gene's association with disease phenotypes (Figure 8). They were further sorted with the Corrected Right P-value in order to know their statistical significance.

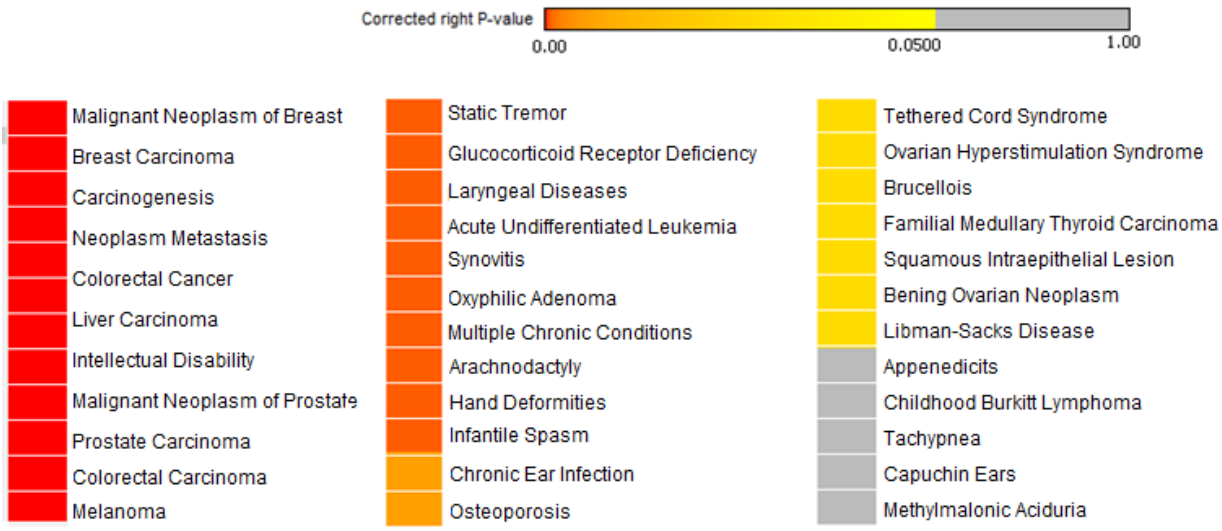


Figure 8: *Enrichment Analysis of Protein Coding genes causing diseases. Red indicates higher enrichment values and few from each scale are presented here from the sets.*

Again, for the list of genes coding for **long non-coding RNA** or lncRNA, we check their survival analysis on different cancer types using lncSEA (Figure 9). Therefore, we were able to figure out the most significant cancer types which are regulated by the bidirectional promoter transcribed genes.

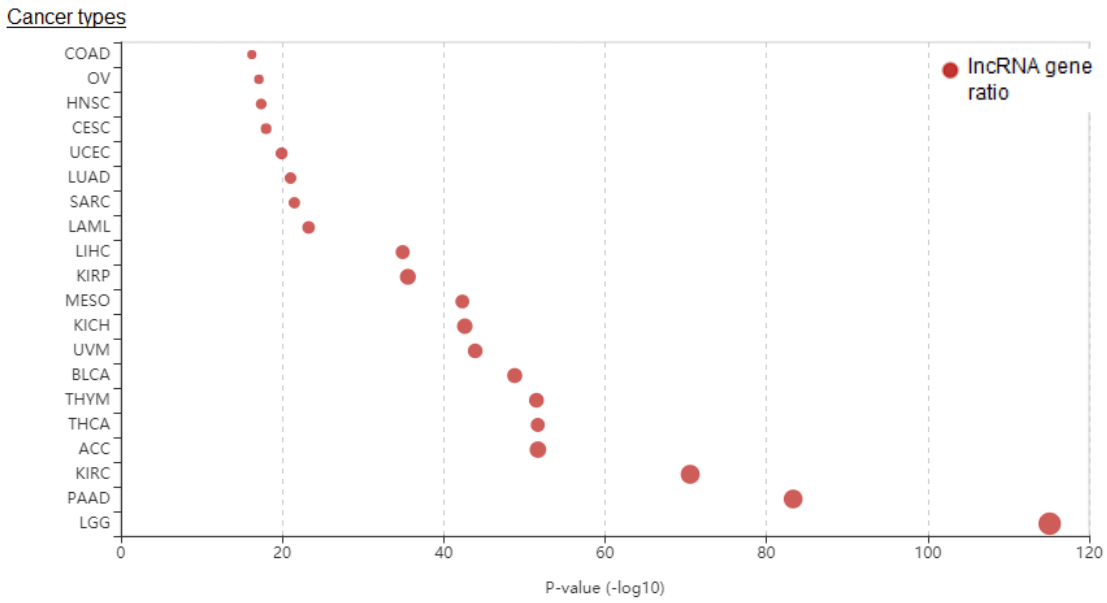


Figure 9: *Enrichment Analysis of lncRNA genes associated with certain Cancer types*

Our retrieved protein coding genes and lncRNA 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 10) from where we can see they have been involved in Metabolic pathways, pathways in cancer, B cell and T cell receptor signaling pathway, Thyroid hormone signaling pathway, MAPK signaling pathway, NF-κB signaling pathway to name a few. An inflammation is triggered when immune cells detect infection or tissue injury which triggers activation of NF-κB, AP1, CREB, c/EBP, and IRF transcription factors. Therefore, most of the retrieved genes actively try to follow the immune pathway.

Figure 10: Enrichment Analysis of bidirectional gene pairs using **A. KEGG Module, B. GOBP Module.**

Despite the obvious signaling where the genes have been leading to immune pathway, we cannot rule out the possibility of immune infiltration. The idea that inflammation and cancer may be linked is not entirely new. The **Hallmarks of Cancer** indicate the stages of development of human tumors. The major hallmarks for both lncRNA gene targets and protein coding genes have been identified (Figure 11), which includes Hypoxia, DNA Repair, TNF alpha signaling via NF-κB, P53 pathway, Epithelial-mesenchymal Transition (EMT), Apoptosis, Prognosis, Proliferation etc. to name a few. The gradual progression of cancer can be known which acts out as an influence of the genes. Protein coding genes cover majority of the hallmarks needed to clarify cancer stage than the lncRNA target genes.

Figure 11: *Cancer Hallmark for* ***A.*** *lncRNA gene targets and* ***B.*** *Protein Coding genes*

3.3 Selection of cancers associated with bidirectional genes

Multiple common targets of gene pairs have been identified where both of the genes are significantly associated with various types of cancer (Table 2). The significant level was determined from their previously derived p-value with the disease. A p-value < 0.05 determines highly significant, which we use as our baseline for disease selection.

Table 2: Selected Diseases along with their bidirectional gene pairs

Gene1	Gene2	Disease	Short Form
HSPB2	*CRYAB*	Colorectal Cancer	COAD
ATF5	*NUP62*	Low Grade Glioma	LGG
BCL2L12	*IRF3*	Low Grade Glioma	LGG
SLC12A9	*EPHB4*	Low Grade Glioma	LGG
HOXA-AS3	*HOXA3*	Low Grade Glioma	LGG
NTRK1	*INSRR*	Kidney Renal Clear Cell Carcinoma	KIRC
NUF2	*RGS5*	Kidney Renal Clear Cell Carcinoma	KIRC
BCL2L12	*IRF3*	Liver Hepatocellular Carcinoma	LIHC
PSMB9	*TAP1*	Skin Cutaneous Melanoma	SKCM

The majority of the bidirectional genes that were significantly correlated with the associated diseases are the *protein coding* genes (Table 2), such as, Colorectal Cancer causing ***HSPB2*** and ***CRYAB***; then Low-Grade Glioma causing ***ATF5*** and ***NUP62***, ***BCL2L12*** and ***IFR3***, ***SLC12A9*** and ***EPHB4***; Kidney Renal Clear Cell Carcinoma causing genes ***NTRK1*** and ***INSRR***, ***NUF2*** and ***RGS5***; Liver Hepatocellular Carcinoma causing genes ***BCL2L12*** and ***IRF3***; Melanoma causing ***PSMB9*** and ***TAP1*** gene. Only ***HOXA-AS3*** which is expressed in Low Grade Glioma is found to be a *lncRNA* gene but its bidirectional partner ***HOXA3*** is a *protein coding* gene.

3.4 Bidirectional Genes significantly associated with Patient Survival in the selected cancers

Survival analysis is an area of statistics for modeling clinical data by dividing them into two groups. The application of Kaplan–Meier (KM) plot is used to understand the patient's survival among highly gene expressed group and lower gene expressed group of the specified disease. Additionally, a box plot of the genes along with their correlated disease is also constructed. The signature score is calculated by taking the mean value of expression level (log_2 (TPM + 1)) of each gene in Th1-like signature gene set. One box indicates the tumor samples while the other one represents the normal tissues. Comparing the box plots, we can get an estimate of the expression signature of each gene in the corresponding disease.

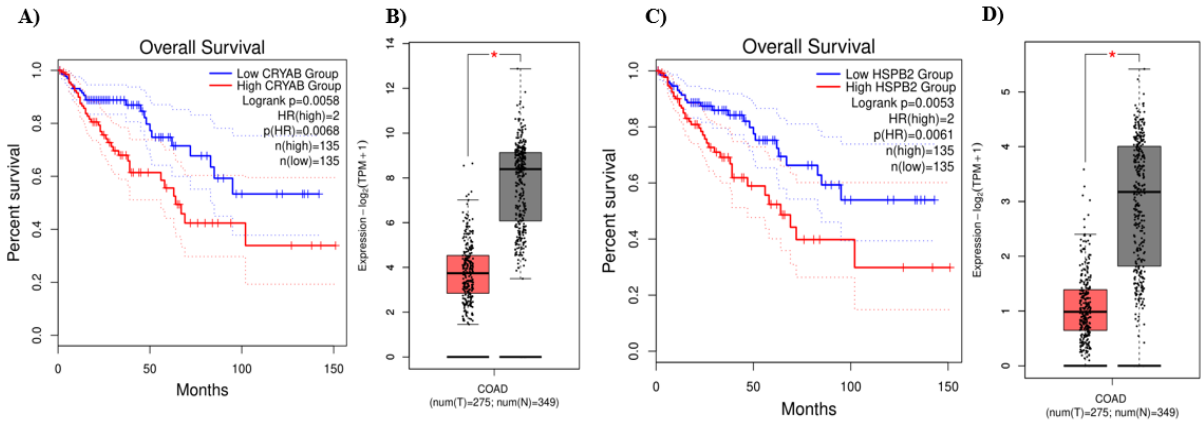


Figure 12: For Colorectal Cancer or COAD (A), (C) *represent Kaplan–Meier (KM) plot for survival of CRYAB and HSPB2 respectively;* ***(B), (D)*** *represent box plot for expression of CRYAB and HSPB2 respectively.*

The Kaplan–Meier (KM) plot for **Colorectal Cance**r or COAD shows 2 groups for *CRYAB* and *HSPB2* (Figure 12) 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 Colorectal Cancer patients. Again, the box plot is

showing lower expression of tumor cells than opposed to normal cells for both the genes in COAD patients (Figure 12).

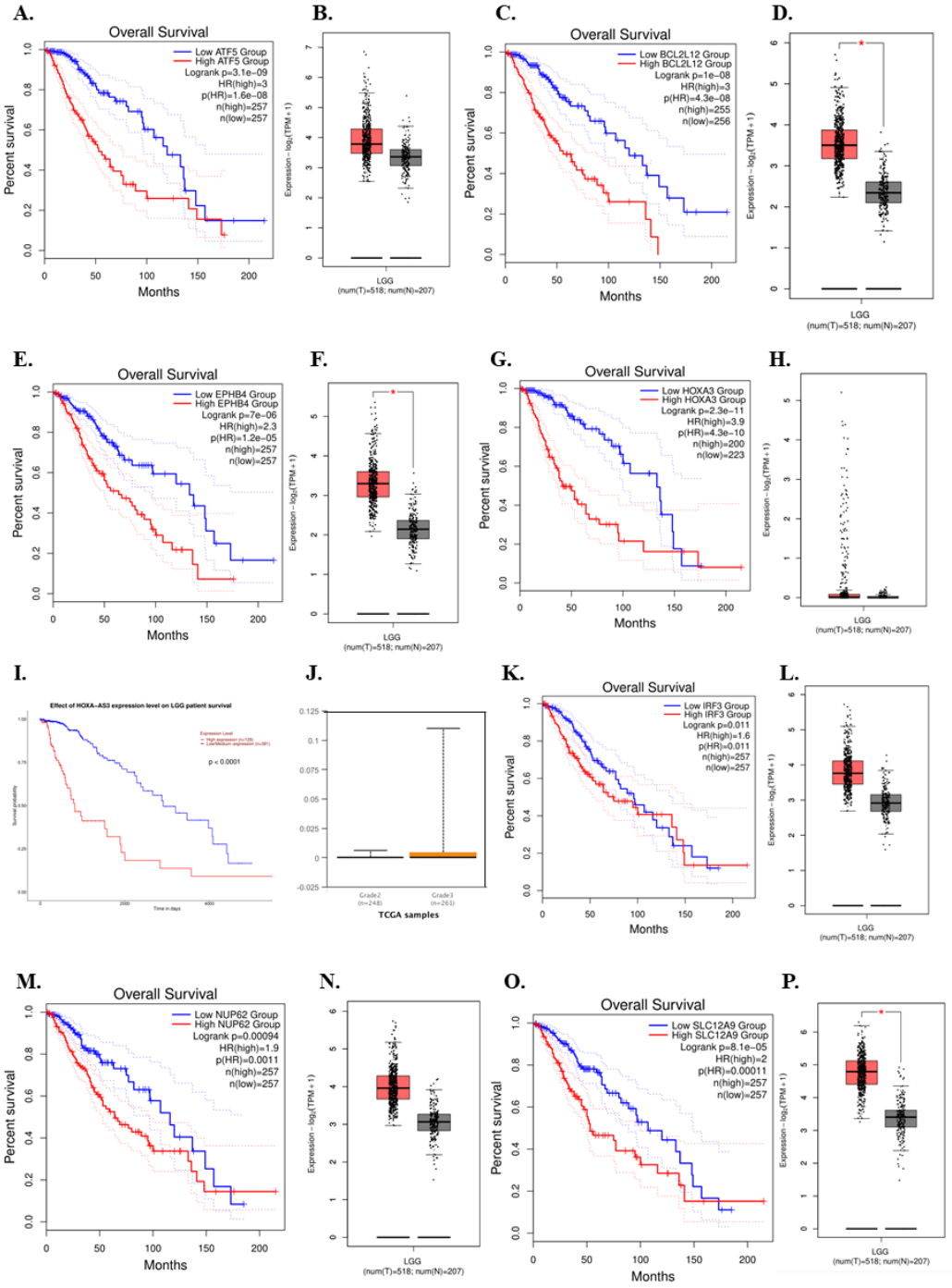


Figure 13:For Low Grade Glioma or LGG (A), (C), (E), (G), (I), (K), (M), (O) *represent Kaplan–Meier (KM) plot for survival of ATF5, BCL2L12, EPHB4, HOXA3, HOXA-AS3, IRF3, NUP62 and SLC12A9 respectively;* ***(B), (D), (F), (H), (J), (L), (N), (P)*** *represent box plot for expression of ATF5, BCL2L12, EPHB4, HOXA3, HOXA-AS3, IRF3, NUP62 and SLC12A9 respectively.*

The Kaplan–Meier (KM) plot for **Low Grade Glioma or LGG** also shows 2 groups for each of the genes with one group having a high gene expression and the other with low gene expression (Figure 13). As the time goes, the survival rate for both groups gets lowered in each of the high and low gene expressed group. Despite having an overlapping among the groups for *ATF5*, *HOXA3, IRF3, NUP62* and *SLAC12A9*, all of the highly expressed group ultimately has the least rate of survival. Again, the box plot of gene expression is higher in tumor cells than the normal cells in all of the disease groups (Figure 13).

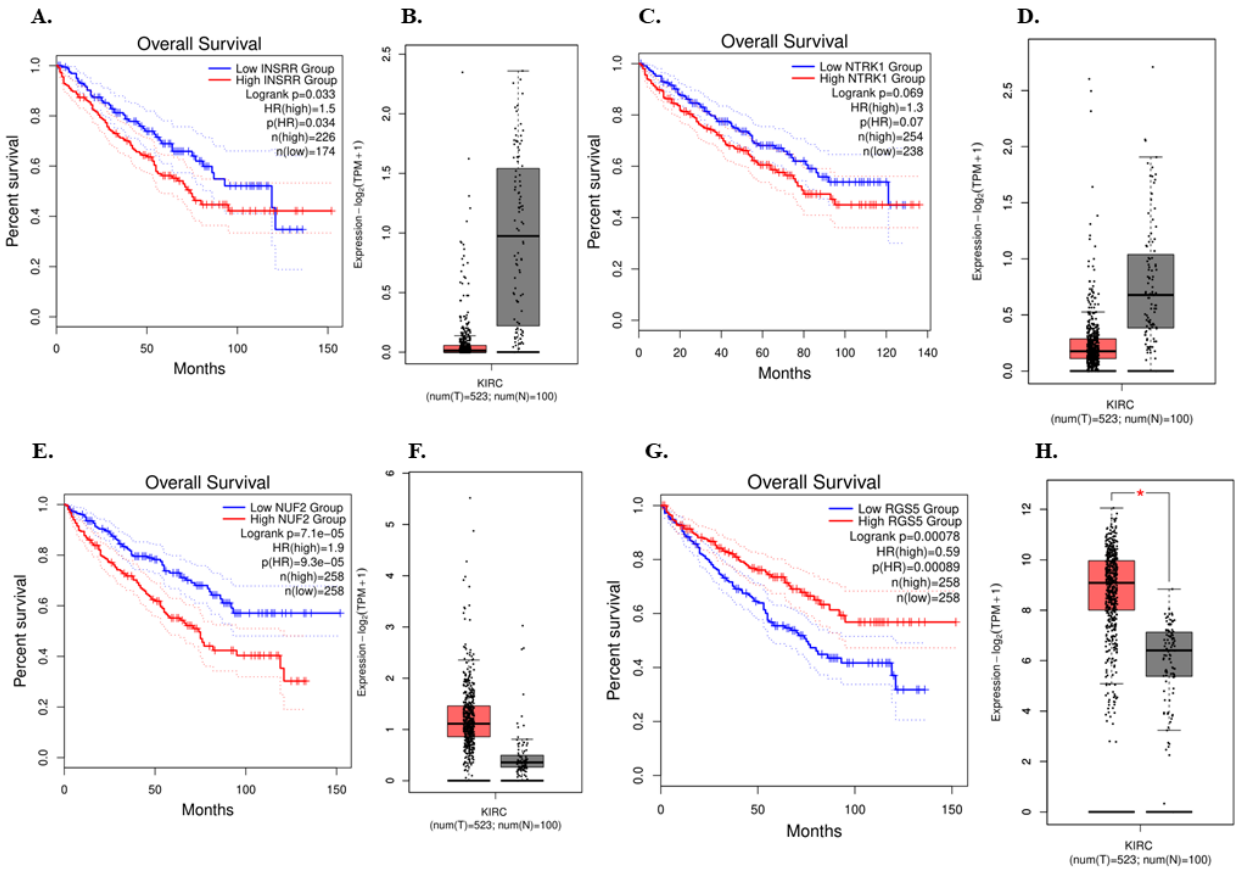


Figure 14: For Kidney Renal Clear Cell Carcinoma or KIRC (A), (C), (E) *and* ***(G)*** *represent Kaplan–Meier (KM) plot for survival of INSRR, NTRK1, NUF2 and RGS5 respectively;* ***(B), (D), (F)*** *and* ***(H)*** *represent box plot for expression of INSRR, NTRK1, NUF2 and RGS5 respectively.*

For **KIRC**, the low gene expression group of *INSRR* has lower rate of survival whereas the high gene expression group has higher chance of survival; again, the high gene expression group of *NTRK1*, *NUF2* and *RGS5* has lower survival rate

(Figure 14). And the tumor cell producing group for *NUF2* and *RGS5* has higher expression level than the normal cells; but the opposite happens for *INSRR* and *NTRK1* (Figure 14).

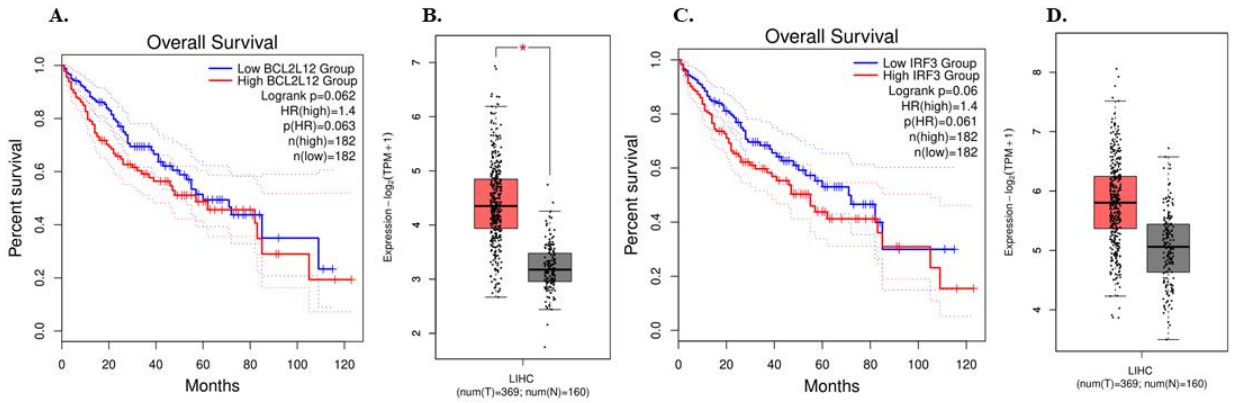


Figure 15: For Liver Hepatocellular Carcinoma or LIHC (A), (C) *represent Kaplan–Meier (KM) plot for survival of BCL2L12 and IRF3 respectively;* ***(B), (D)*** *represent box plot for expression of BCL2L12 and IRF3 respectively.*

Among the groups of **LIHC** patients, high expression of *BCL2L12* and *IRF3* genes group have lower survival rate than the other group of low expression according to the KM plot (Figure 15). Whereas, significantly higher amount of tumor cell expression level than normal cells in LIHC is observed for both the genes (Figure 15).

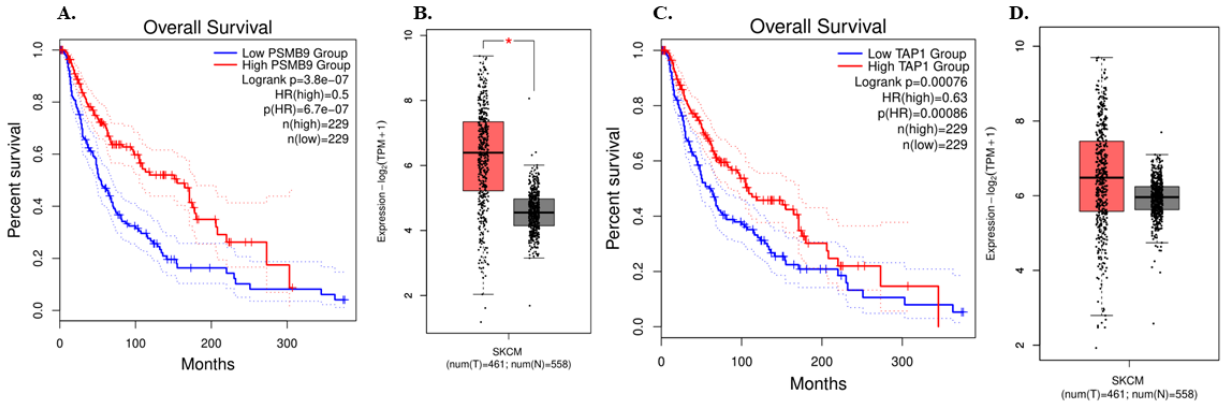


Figure 16: For Melanoma or SKCM (A), (C) *represent Kaplan–Meier (KM) plot for survival of PSMB9 and TAP1 respectively;* ***(B), (D)*** *represent box plot for expression of PSMB9 and TAP1 respectively.*

Patients with **Skin cutaneous melanoma** or **SKCM** has a lower survival rate for low *PSMB9* gene expression group (Figure 16). And the tumor expression level is higher than the normal ones for *PSMB9* in **SKCM** (Figure 16). On the other hand, the high expression group of *TAP1* has lower survival than the low gene expression group (Figure 16). Yet, tumor expression level of *TAP1* is higher in **SKCM** (Figure 16).

3.5 Correlation of Bidirectional Gene pair in selected Cancer

The obtained genes are checked for their correlation with their neighboring genes in whichever cancers they influence. High correlation can be indicative of cancer prognosis [104].

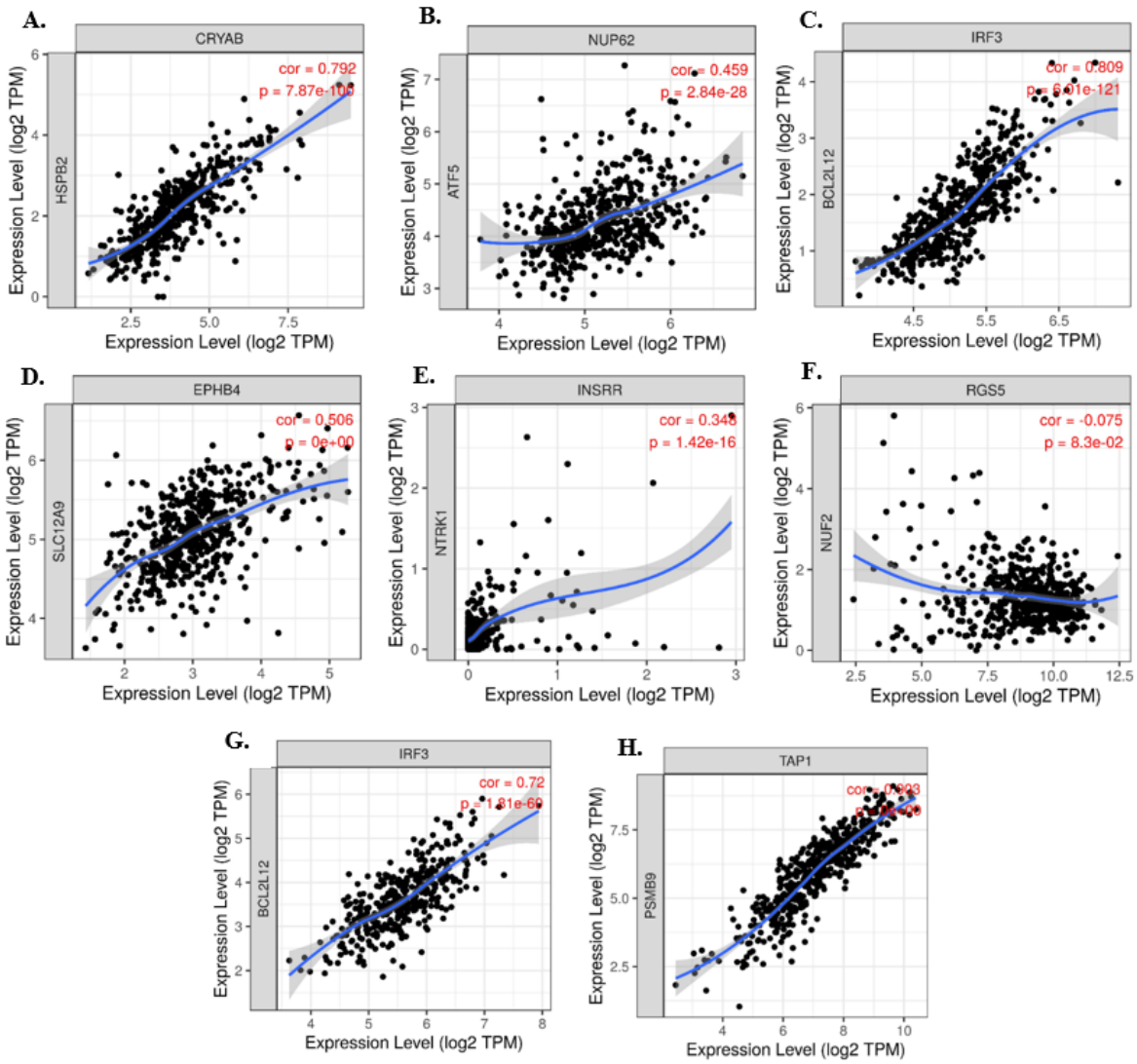


Figure 17: *Correlation between* ***(A)*** *HSPB2 and CRYAB in* ***COAD; (B)*** *ATF5 and NUP6 in* ***LGG; (C)*** *BCL2L12 and IRF3 in* ***LGG; (D)*** *SLC12A9 and EPHB4 in* ***LGG; (E)*** *NTRK1 and INSRR in* ***KIRC; (F)*** *NUF2 and RGS5 in* ***KIRC; (G)*** *BCL2L12 and IRF3 in* ***LIHC****; (H)PSMB9and TAP1 in* ***SKCM***

There lies a strong positive correlation between gene pairs *HSPB2* and *CRYAB* for **Colorectal Cancer** (Figure 17.A), implying they can serve as potential biomarkers for COAD. Again, among the genes of **Low Grade Glioma** or LGG only *BCL2L12* and *IRF3* has a strong positive correlation (Figure 17.C) whereas both *ATF5* and *NUP6* and *SLCL12A9* and *EPHB4* has a moderate positive correlation (Figure 17.B, 17.D). Implying *BCL2L12* and *IRF3* to be more accountable for LGG cancer prognosis. The genes of KIRC or **Kidney Renal Clear Cell Carcinoma** has a weak correlation value (Figure 17.E, 17.F) indicating there is absolutely no correlation among the genes for KIRC. The value of correlation of the expression levels among the promoter sharing gene pairs *BCL2L12* and *IRF3* is again a strong positive correlation (Figure 17.G) for LIHC or **Liver Hepatocellular Carcinoma**. Expression levels' correlation value among the gene pairs sharing bidirectional promoter *HSPB2* and *CRYAB* in **Skin Cutaneous Melanoma** or SKCM is very strong and positive (Figure 17.H). Thus, these genes can be implicative of good prognosis in their respective cancer types.

3.6 Association of Immune infiltration level with Patient Survival

The rise in immune cells most notably B cell, CD4+ T cell, CD8+ T cell, Dendritic Cell, Neutrophil and Macrophage may pose risk a risk in certain Cancer patients [105]. We predict the significance on the selected cancer types with their p-value and z-score.

For patients with increasing **B cell**, the survival rate for patients with COAD and KIRC is in the risk prone region. Whereas it is safe for SKCM patients (Figure 18.A). Again, **CD4+ T cell** increase might be a problem for patients having KIRC, LGG and LIHC with a reduced risk for SKCM patients (Figure 18.B). Abundance of **CD8+ T cell** in KIRC and LGG patients might pose a risk but not an issue for COAD, LIHC or

SKCM patients (Figure 18.C). **Dendritic Cell** increase in patients with COAD, KIRC, LIHC and SKCM patients have reduced to no risk but concerning for LGG patients (Figure 18.D). Patients with elevated **Macrophage** count are relatively in lower risk zone for all cancer patients (Figure 18.E) but LGG, KIRC and LIHC patients might often find surge in their disease prognosis due to the rise. COAD, KIRC and SKCM patients with increased **Neutrophil** are relatively in reduced to no risk area but mildly concerning for LGG and LIHC patients (Figure 18.F).

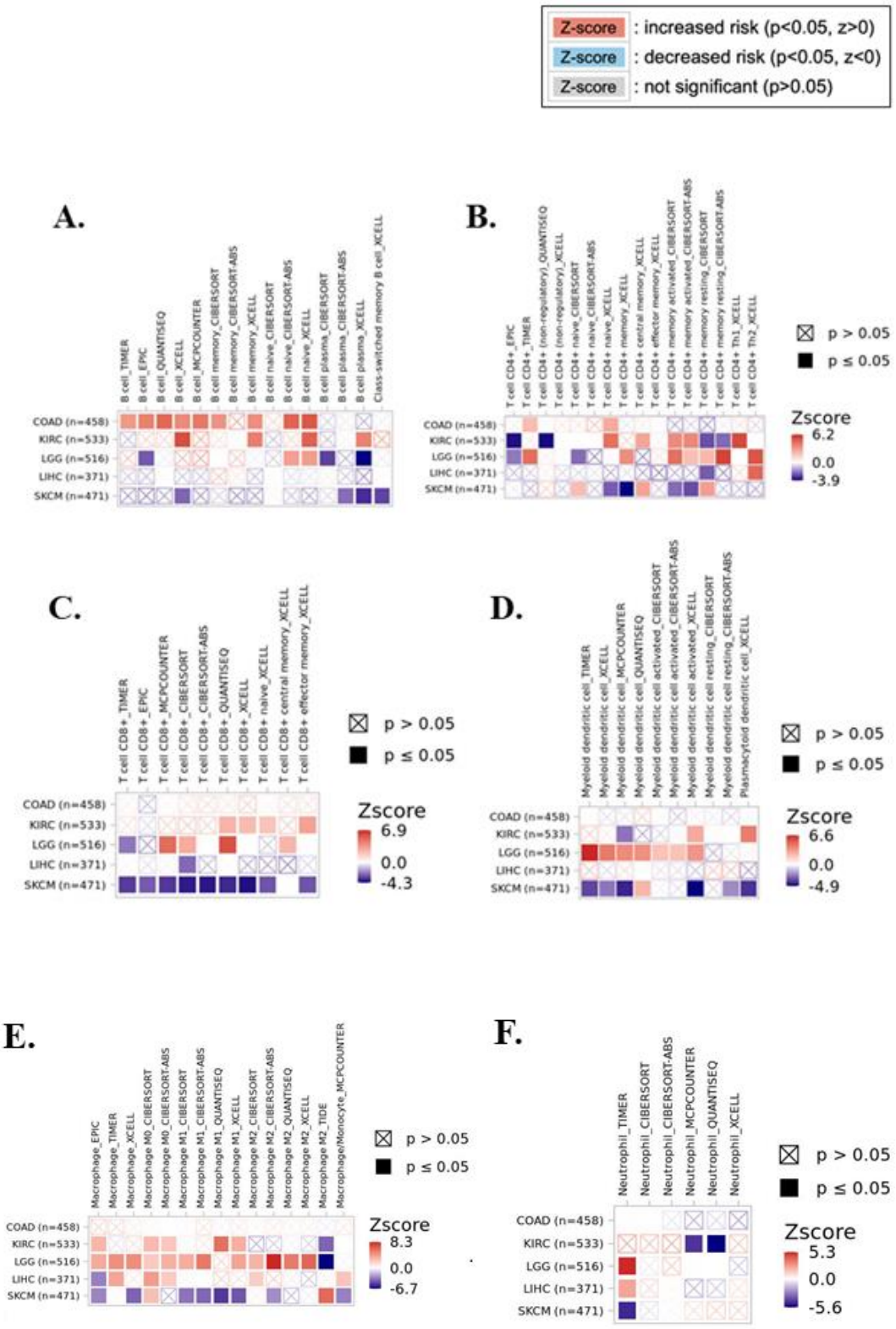


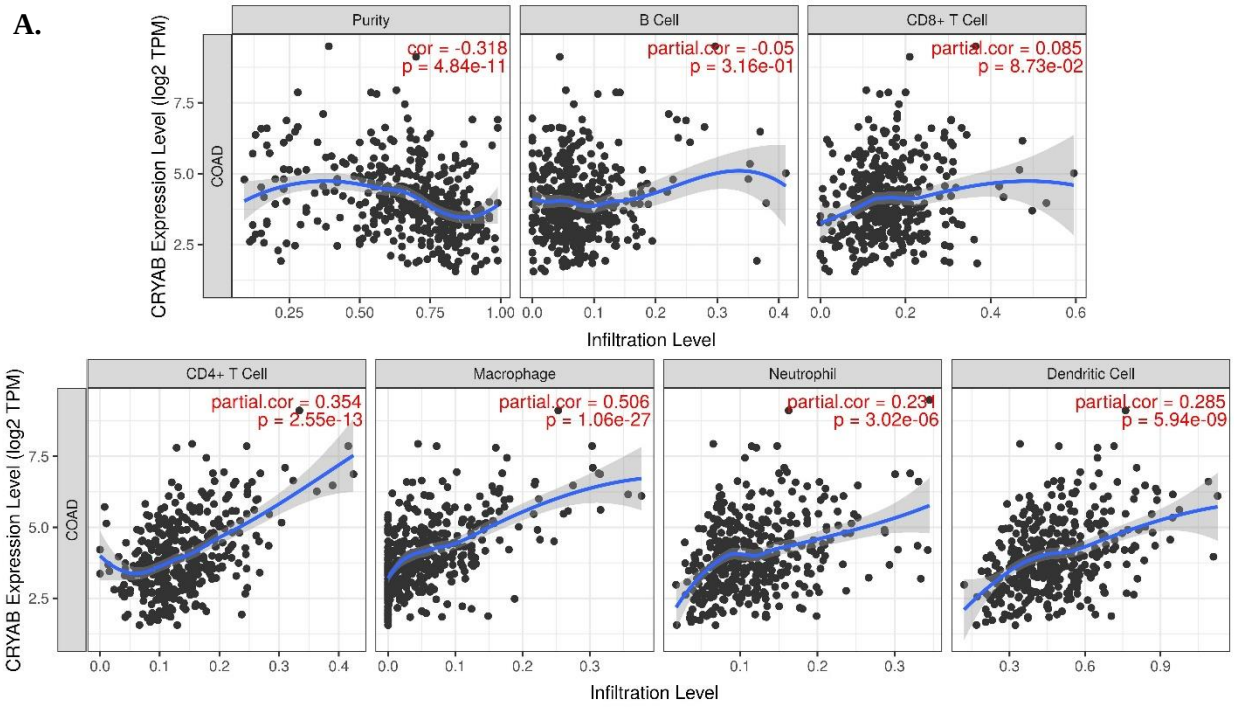
Figure 18: *Association of Patients Survival in selected Cancer types with Immune Cells:* ***(A)*** *B cell,* ***(B)*** *CD4+ T cell,* ***(C)*** *CD8+ T cell,* ***(D)*** *Dendritic cell,* ***(E)*** *Macrophage,* ***(F)*** *Neutrophil*

3.7 Correlating Bidirectional genes contributing to Immune infiltration in the selected Cancer types

Since a correlation between the bidirectional promoter sharing genes have already been determined for their associated cancer types and the risk level of those cancer with increased immune cells have been predicted, we move on to check the correlation between the immune cells and each gene individually. This will help determine evidence if that gene is assisting in the recruitment of the cancer cells.

i. Colorectal Cancer or COAD

As we have already seen that a rise in B cell can be labelled as risk in patient survival for someone affected with said cancer type (Figure 19), therefore, even though B cell had relatively higher risk of survival, the correlation is weak negative for either of *CRYAB* or *HSPB2* gene (Figure 19). And other immune cells do not bear high risk so their correlation value is not put into a factor here.



B.

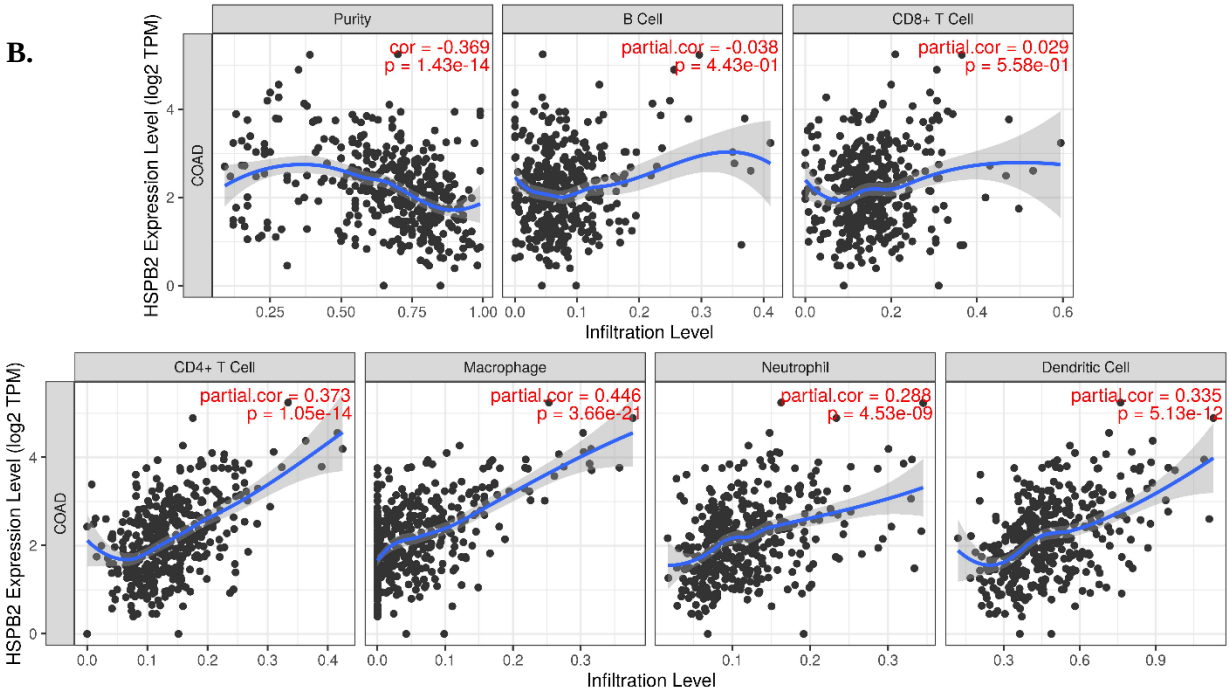
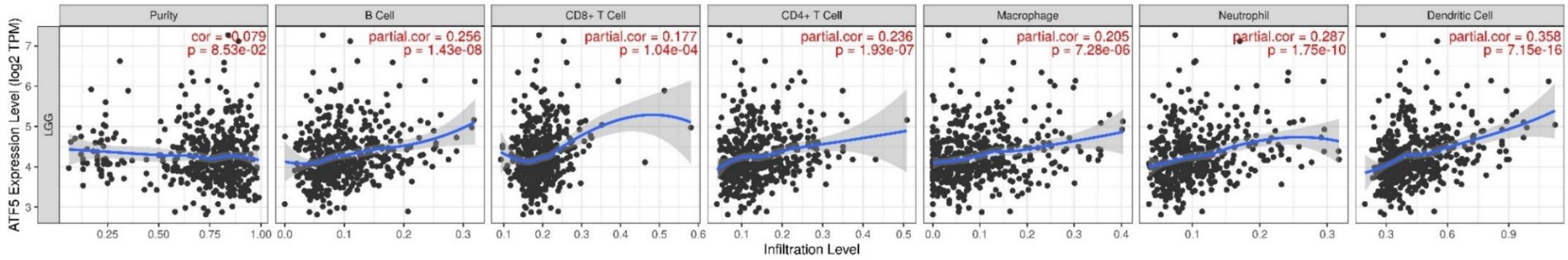


Figure 19: *Correlation value of* ***COAD*** *for* ***(A)*** *CRYAB gene and* ***(B)*** *HSPB2 gene with immune cells*

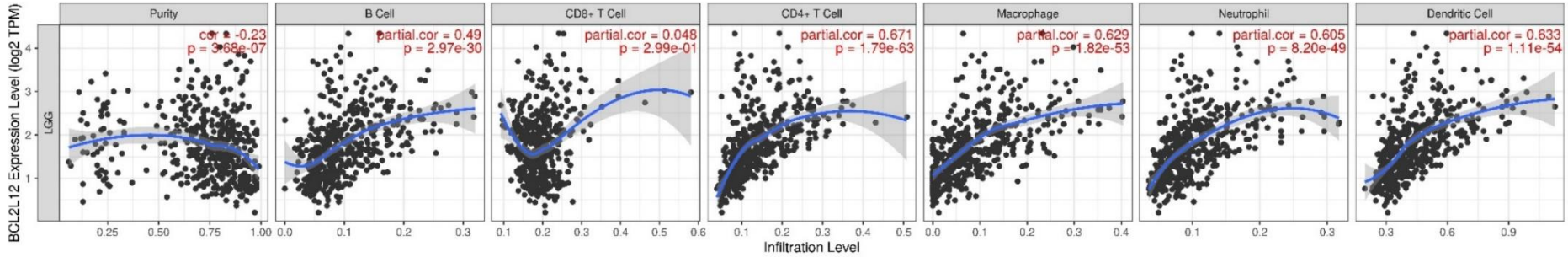
ii. Low Grade Glioma or LGG

Due to increase of T cell, Dendritic Cell, Macrophage, Neutrophil, patient survival goes in risk prone zone (Figure 18). *ATF5* and *EPHB4* has weak positive correlation in the immune cell types (Figure 20.A, 20.C). Whereas, *BCL2L12*, *IRF3* and *NUP62* has moderate positive correlation for CD4+ T cell, Dendritic, Macrophage and Neutrophils (Figure 20.B, 20.D, 20.E). Lastly, *SLC12A9* has weak positive correlation for all of the concerned immune cells (20.F). Therefore, only the protein coding genes, *BCL2L12*, *IRF3* and *NUP62* directly points towards recruiting immune cells during cancer prognosis for LGG.

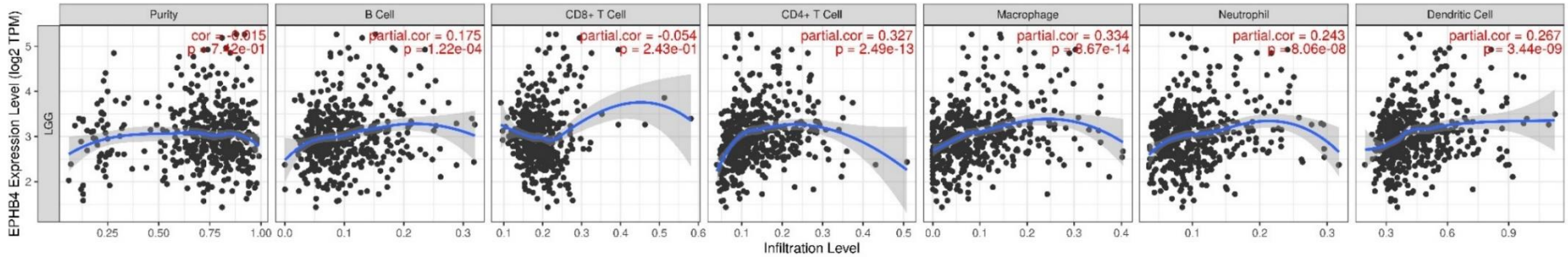
A.


Purity
cor = 0.079
p = 8.53e-02
B Cell
partial.cor = 0.256
p = 1.43e-08
CD8+ T Cell
partial.cor = 0.177
p = 1.04e-04
CD4+ T Cell
partial.cor = 0.236
p = 1.93e-07
Macrophage
partial.cor = 0.205
p = 7.28e-06
Neutrophil
partial.cor = 0.287
p = 1.75e-10
Dendritic Cell
partial.cor = 0.358
p = 7.15e-16
ATF5 Expression Level (log2 TPM)
LGG
Infiltration Level

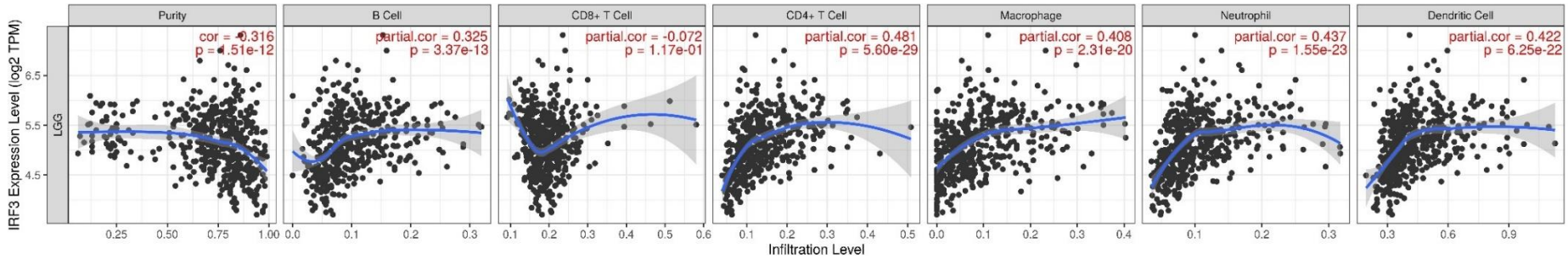
B.


Purity
cor = -0.23
p = 3.68e-07
B Cell
partial.cor = 0.49
p = 2.97e-30
CD8+ T Cell
partial.cor = 0.048
p = 2.99e-01
CD4+ T Cell
partial.cor = 0.671
p = 1.79e-63
Macrophage
partial.cor = 0.629
p = 1.82e-53
Neutrophil
partial.cor = 0.605
p = 8.20e-49
Dendritic Cell
partial.cor = 0.633
p = 1.11e-54
BCL2L12 Expression Level (log2 TPM)
LGG
Infiltration Level

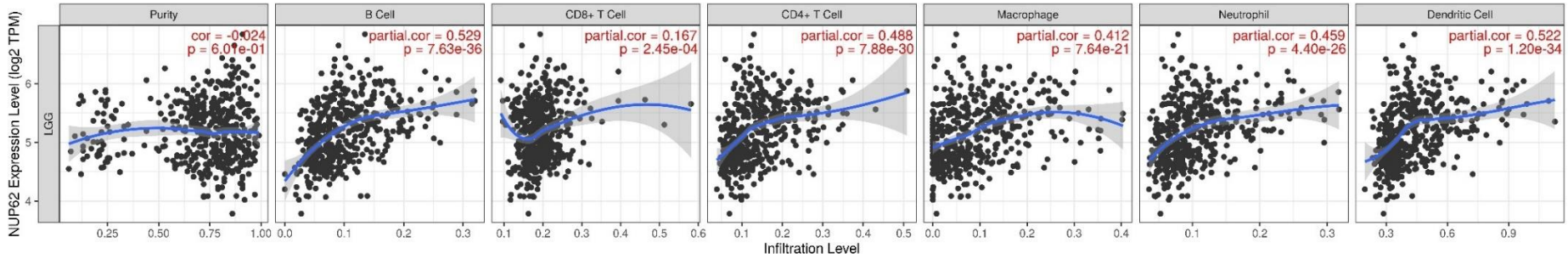
C.


Purity
cor = -0.015
p = 7.42e-01
B Cell
partial.cor = 0.175
p = 1.22e-04
CD8+ T Cell
partial.cor = -0.054
p = 2.43e-01
CD4+ T Cell
partial.cor = 0.327
p = 2.49e-13
Macrophage
partial.cor = 0.334
p = 6.67e-14
Neutrophil
partial.cor = 0.243
p = 8.06e-08
Dendritic Cell
partial.cor = 0.267
p = 3.44e-09
EPHB4 Expression Level (log2 TPM)
LGG
Infiltration Level

D.



E.



F.

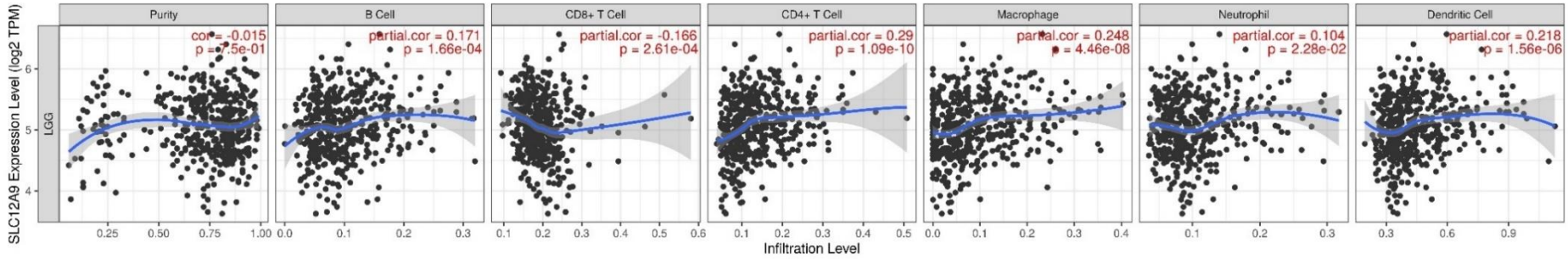
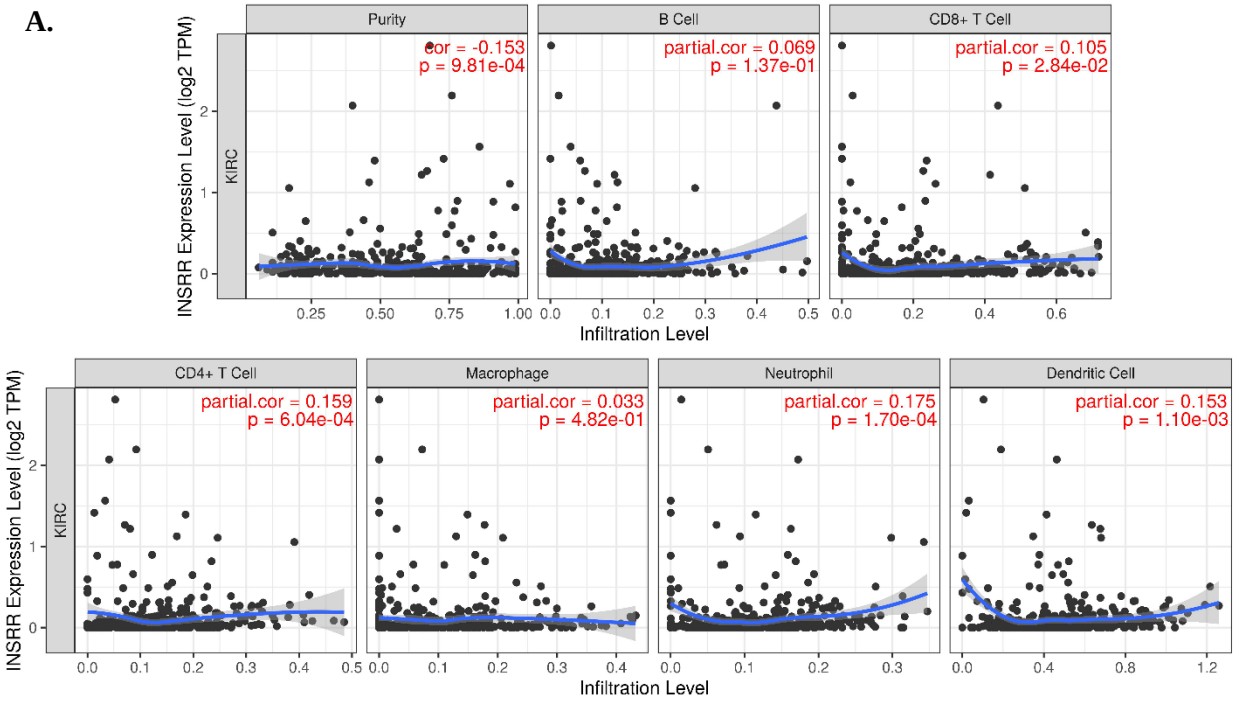


Figure 20: *Correlation value of* ***LGG*** *for* ***(A)*** *ATF5,* ***(B)*** *BCL2L12,* ***(C)*** *EPHB4,* ***(D)*** *IRF3,* ***(E)*** *NUP62,* ***(F)*** *SLC12A9 genes with immune cells*

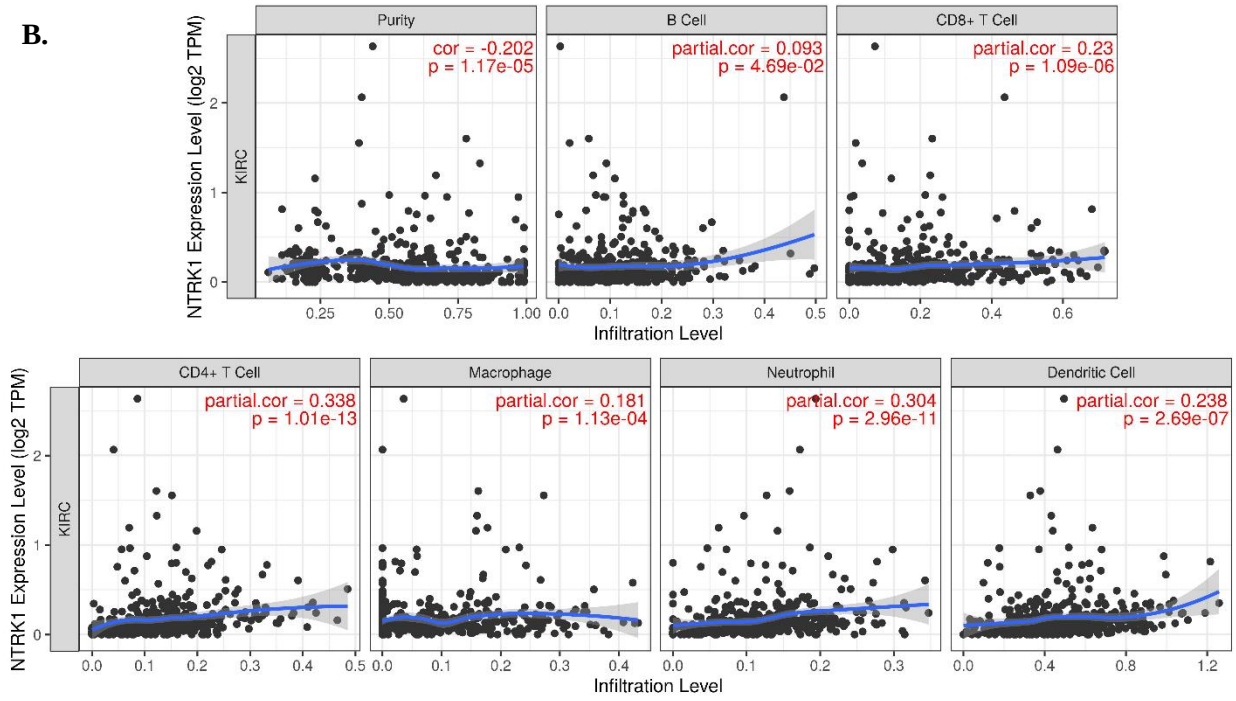
iii. Kidney Renal Clear Cell Carcinoma or KIRC

The immune cells that pose a higher risk in patient survival for KIRC are B cell naive, CD8+ T cell, Plasmacytoid Dendritic Cell and Macrophage (Figure 18).

A.



B.



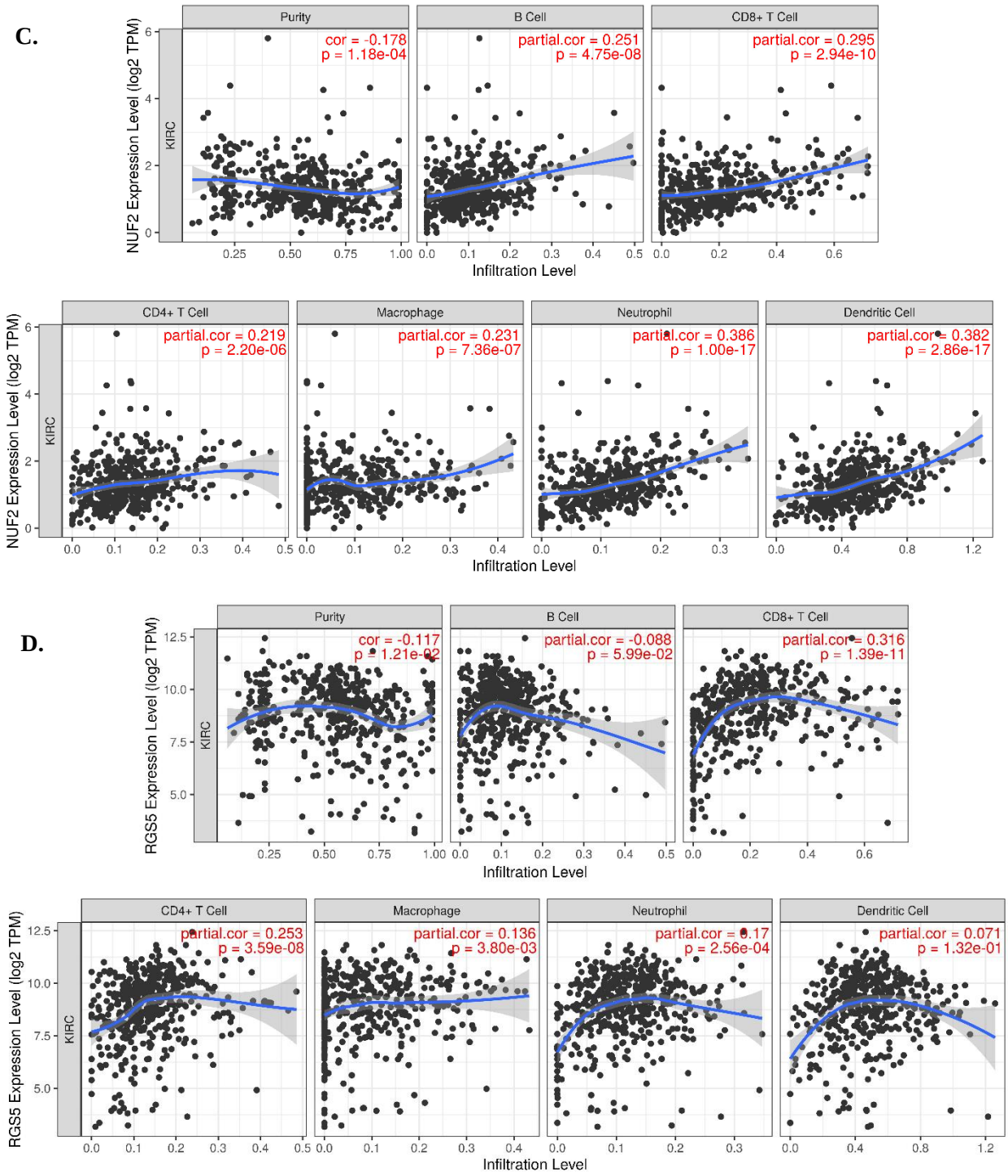


*Figure 21: Correlation value of **KIRC** in **(a)INSRR** gene, **(b)NTRK1** gene, **(c)NUF2** gene, **(d) RGS5** gene with immune cells*

All the genes show weak correlation in our concerned immune cells for KIRC (Figure 21). So, these genes are less likely to recruit immune cells for the prognosis of KIRC.

iv. Liver Hepatocellular Carcinoma or LIHC

LIHC patients are less likely to be associated with immune cell infiltration. However, CD4+ T cell, Macrophage and Neutrophil elevation can cause a risk in the patients' survival (Figure 18).

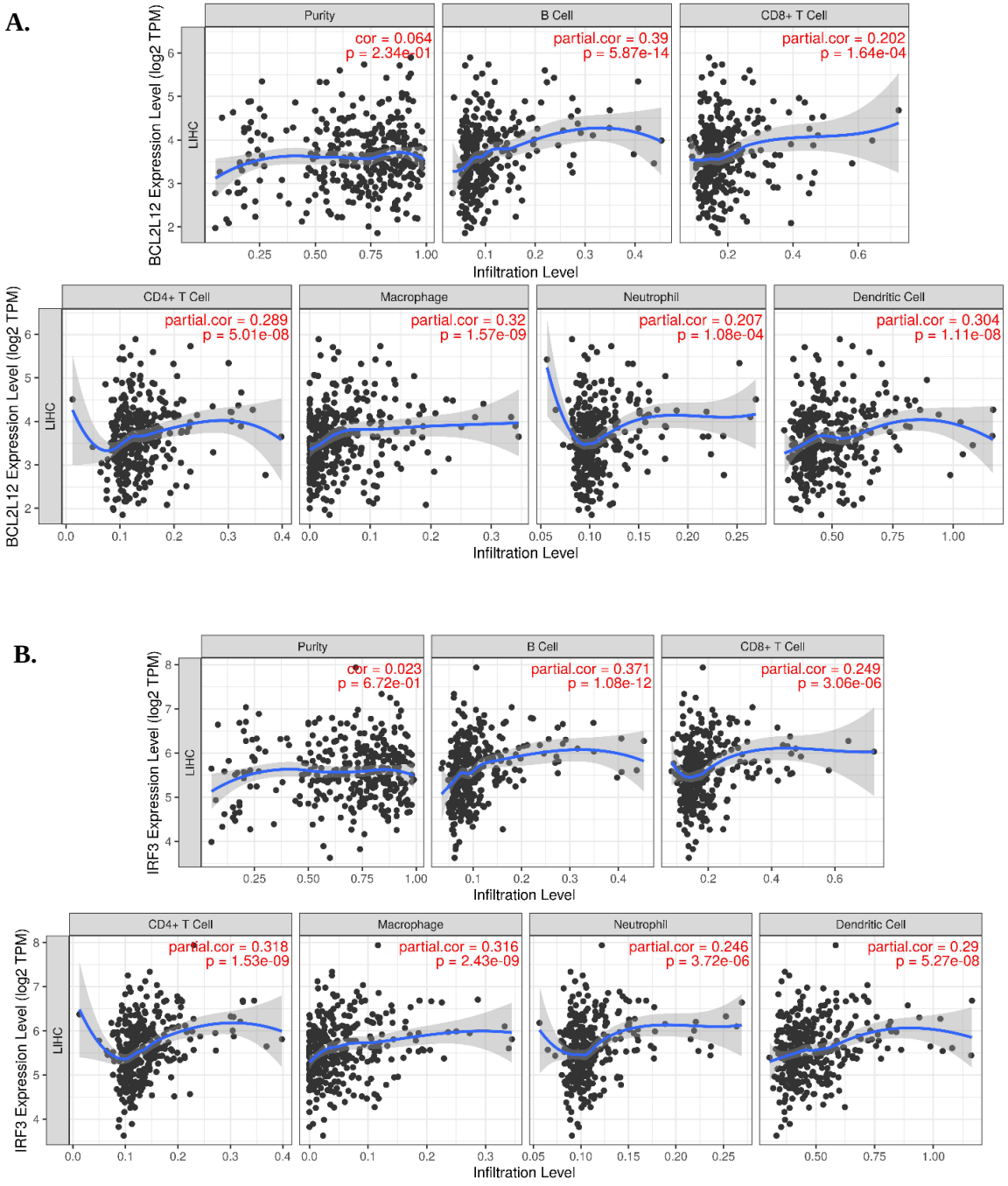


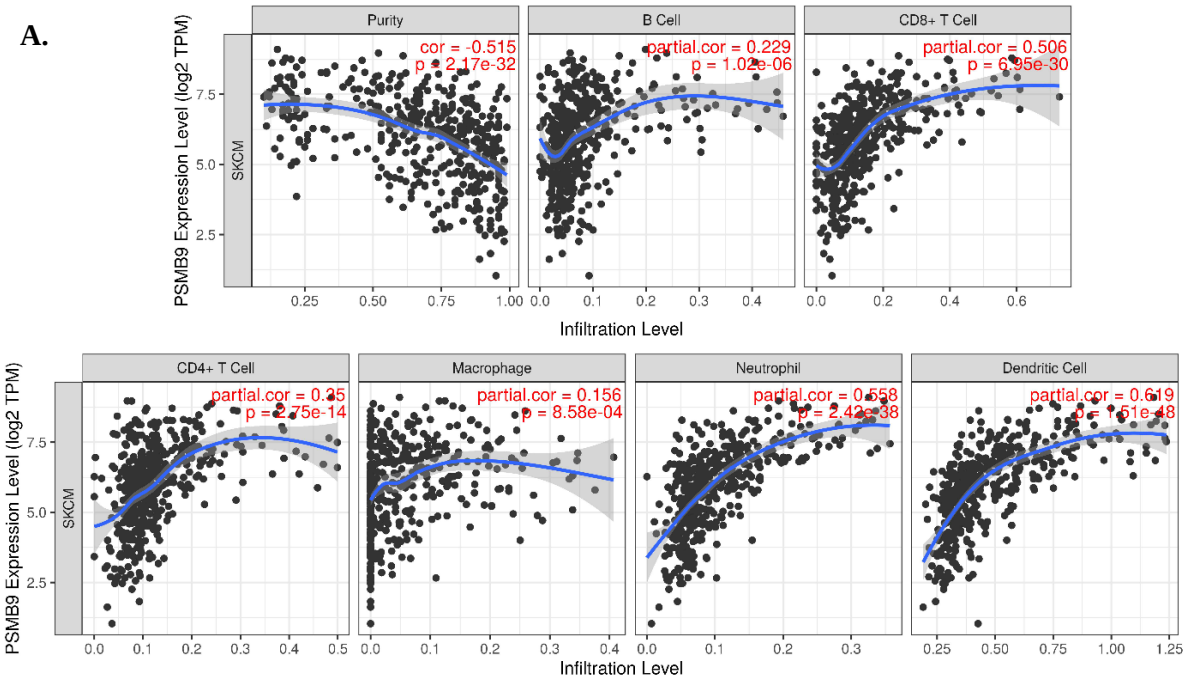
Figure 22: *Correlation value for LIHC in* ***(A) BCL2L12*** *gene and* ***(B) IRF3*** *gene with immune cells*

Again, all the genes of LIHC are seen to have weak positive correlation among each of the genes and our concerned immune cells (Figure 22), once again implicative of the mentioned genes to not recruit immune cells for cancer prognosis.

v. Skin Cutaneous Melanoma or SKCM

There is reduced or very low risk due to almost all of the immune cells that might infiltrate in the case of a SKCM patient. However, CD4+ T cell and Macrophage pose slight risk in patient survival (Figure 18).

A.



B.

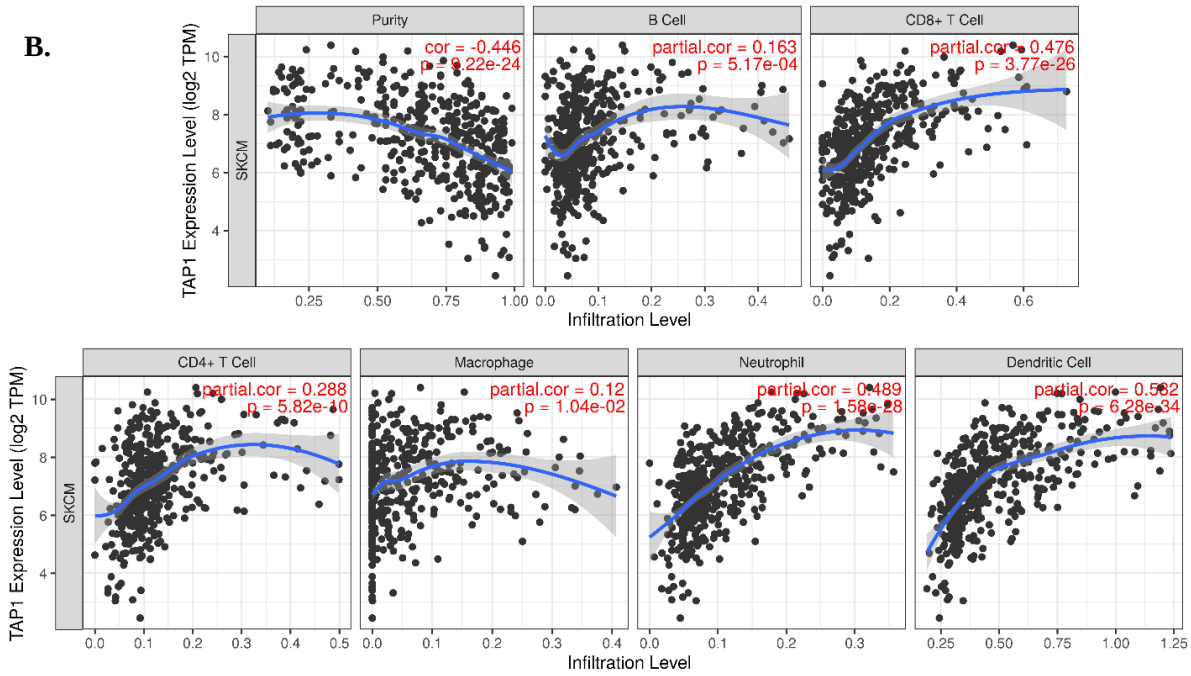


Figure 23: *Correlation value in SKCM for* ***(A) PSMB9*** *gene and* ***(B) TAP1*** *gene with immune cells*

Once again, both the genes for SKCM, *PSMB9* and *TAP1* are seen to have weak positive correlation between each of the concerned immune cells, CD 4+ T cell and Macrophage (Figure 23). Which still points to the fact that, the mentioned genes do not recruit immune cells for cancer prognosis. However, CD 8+ T cell, Dendritic and Neutrophil happens to have a moderate positive correlation but the genes elevation post reduced risk in the cancer patients (Figure 18).

Chapter 4

Discussion

Cancer is a state of uncontrolled cellular growth mediated by some genes gone rogue. Even though we have associated cancer with certain gene names, unbeknownst to most people, certain other genes may indirectly have a hand on this. In this study, we take a wider look at the bidirectional promoter sharing gene pairs who have a correlation in cancer prognosis with immune cells.

Bidirectional promoters were shortlisted from the entire extracted human genes based on the distance between two neighboring genes distance from transcription start site being 1000bp. Studies have stated that, the promoter length is within 100-1000bp long [106]. Therefore, the ones less than 100bp are also filtered out. Moreover, while we could've chosen among a wider array of RNA genes, such as siRNAs, miRNAs and piRNAs, we only stuck to lncRNA and Protein Coding genes omitting the smaller RNA transcripts. Evidence points towards lncRNAs having key functions in regulating diverse biological processes, such as imprinting control, cell differentiation, development, immune response, cell cycle and apoptosis [107, 108]. Again, lncRNAs act as cis-regulators because their expression is significantly correlated with their neighboring protein-coding genes [109].

The retrieved genes have a clear association with the immune related cancer hallmarks such as, Hypoxia, DNA Repair, TNF alpha signaling via NF-kB, P53 pathway, Epithelial-mesenchymal Transition (EMT), Apoptosis, Prognosis, Proliferation etc. (Figure 11). While we think of our immune system as an ally, our immune system also plays a role in promoting tumor growth as studies have pointed out [41]. Tumor-promoting inflammation done by immune infiltration during early

carcinogenesis promotes angiogenesis and tissue remodeling without triggering anti-tumor immunity. This mechanism is actively promoted by the tumor through production of cytokines/chemokine and modulators of metabolic pathways rather than neoantigen recognition and lymphocyte involvement which we can see our retrieved genes have an active role in such as, pathways in cancer, transcriptional misregulation in cancer, T cell receptor signaling pathway, B cell receptor signaling pathway, cAMP signaling pathway, inflammatory mediator regulation of TRP channels, apoptotic process, angiogenesis, negative regulation of translation, axonogenesis (Figure 10). And to know which disease each of the genes highly correlate to are easily found from the DisGeNET based enrichment as it is one of the largest and comprehensive repositories of human gene-disease associations (GDAs) currently available.

Finding differential correlation is a good approach for detecting functional relationships among gene pairs. The general hypothesis is that genes behaving differently in different disease conditions are more likely to be related to a specific disease mechanism [110]. With a large sample number, the correlation coefficients become a good approximation to the true values and thereby ensuring significant functional relationships [111].

Among the gene pairs that we narrowed down, *BCL2L12* and *IRF3* are highly expressed in LGG (Figure 17.C). Each of them has a high correlation value with the immune cells CD4+ T cell, Dendritic, Macrophage and Neutrophils (Figure 20.B, 20.D). These cells when highly elevated in a LGG patient will enter an increased risk zone (Figure 18). And also, the rate of tumor cells increases drastically as opposed to normal cells for these genes during a case of LGG (Figure 13.D, 13.L). Therefore, it supports the idea that the immune cells here are being recruited by the

aforementioned genes which is promoting cancer metastasis. A study by Stegh et. al concluded that *BCL2L12* gene directly inhibits the tumor suppressor gene p53 leading to Glioma [112]. Whereas, Pattwell and Holland in their study talked about its promoter sharing gene IRF3 which upon increasing may reduce the tumorigenicity of Glioma [113].

Again, the gene pairs *PSMB9* and *TAP1* are highly expressed in SKCM (Figure 17.H). These genes also have a strong positive correlation value with CD 8+ T cell, Dendritic and Neutrophil (Figure 23). Even the expression correlation between the genes is significantly high in the case of SKCM (Figure 17.H). This suggests possible cancer prognosis however, there is reduced risk in SKCM patients for the mentioned immune cells (Figure 18). Studies by Wang et. al stated the high expression of *PSMB9* and *TAP1* can lead to CD8+ T cell infiltration [114]. Cancer cells gain immunity by the infiltration of CD8+ T cells in the tumor microenvironment [115]. Therefore, due to the reduced risk, we can altogether find an opportunity to improve the prognosis of SKCM patients.

Bidirectional promoter sharing genes *HSBP2* and *CRYAB* with the potential to induce COAD have rather low correlation with immune infiltration (Figure 19). But still the high expression (Figure 17.A) may indicate lack of immune system recruitment towards the tumor microenvironment. Similarly, the paired genes *NTRK1*, *INSRR* and *NUF2*, *RGS5* which are highly expressed in KIRC, has a weaker immune infiltration correlation (Figure 21) as well as a weaker gene expression correlation among them (Figure 17). So, it may not be targeting the immune approach to cancer. Again, genes *BCL2L12* and *IRF3* while discussing LIHC might have a stronger correlation between their gene expression (Figure 17), but their weak immune

infiltration correlation (Figure 22) again points towards the fact that it might be missing the immune system recruitment towards the tumor microenvironment.

These findings altogether can be utilized in cancer therapeutics or cancer-immunology related researches since we have a clear understanding of which genes recruit immune cells for tumorigenesis. For example, by inhibiting the expression of genes *BCL2L12* and *IRF3* for LGG patients, we can drastically reduce the risk of survival for those patients. Same thing holds true for other genes and their immunotherapeutic approaches.

Chapter 5

Conclusion

Cancers have been one of the deadliest diseases of the modern world, and our quest for effective treatments continue adamantly. The hallmarks of cancer clarify how they utilize immune infiltration for their own benefit and growth. We need to gather every information we can get to combat this life-threatening disease. In this study, we try solve a piece of this intricate puzzle by establishing correlation between immune cells and bidirectional genes involved in cancer prognosis.

Bidirectionally transcribed gene promoters have not remained hidden from the world for long but rather overlooked due to their feeble presence throughout the whole genome. The simultaneous expression of neighboring genes can bring forth newer approaches to how we design cancer therapeutics.

References

1. Kozak, M.J.G., *Initiation of translation in prokaryotes and eukaryotes.* 1999. **234**(2): p. 187-208.
2. Adachi, N. and M.R. Lieber, *Bidirectional gene organization: a common architectural feature of the human genome.* Cell, 2002. **109**(7): p. 807-9.
3. Carninci, P., et al., *Genome-wide analysis of mammalian promoter architecture and evolution.* 2006. **38**(6): p. 626-635.
4. Trinklein, N.D., et al., *An abundance of bidirectional promoters in the human genome.* Genome Res, 2004. **14**(1): p. 62-6.
5. Chen, Y., et al., *Transcriptional regulation and spatial interactions of head-to-head genes.* 2014. **15**(1): p. 1-9.
6. Wang, G., et al., *Identification of regulatory regions of bidirectional genes in cervical cancer.* BMC medical genomics, 2013. **6 Suppl 1**(Suppl 1): p. S5-S5.
7. Yang, M.Q. and L.L.J.B.g. Elnitski, *Diversity of core promoter elements comprising human bidirectional promoters.* 2008. **9**(2): p. 1-8.
8. Wei, W., et al., *Functional consequences of bidirectional promoters.* Trends in genetics : TIG, 2011. **27**(7): p. 267-276.
9. Jangid, R.K., et al., *Bidirectional promoters exhibit characteristic chromatin modification signature associated with transcription elongation in both sense and antisense directions.* BMC genomics, 2018. **19**(1): p. 313-313.
10. Li, Y.Y., et al., *Systematic analysis of head-to-head gene organization: evolutionary conservation and potential biological relevance.* PLoS Comput Biol, 2006. **2**(7): p. e74.
11. Shu, J., et al., *Silencing of bidirectional promoters by DNA methylation in tumorigenesis.* Cancer Res, 2006. **66**(10): p. 5077-84.
12. Heikkilä, P., R. Soininen, and K.J.J.o.B.C. Tryggvason, *Directional regulatory activity of cis-acting elements in the bidirectional alpha 1 (IV) and alpha 2 (IV) collagen gene promoter.* 1993. **268**(33): p. 24677-24682.
13. Xu, C., J. Chen, and B. Shen, *The preservation of bidirectional promoter architecture in eukaryotes: what is the driving force?* BMC Systems Biology, 2012. **6**(1): p. S21.
14. He, K., et al., *Compact Bidirectional Promoters for Dual-Gene Expression in a Sleeping Beauty Transposon.* Int J Mol Sci, 2020. **21**(23).
15. Seila, A.C., et al., *Divergent transcription: a new feature of active promoters.* 2009. **8**(16): p. 2557-2564.
16. Koyanagi, K.O., et al., *Comparative genomics of bidirectional gene pairs and its implications for the evolution of a transcriptional regulation system.* 2005. **353**(2): p. 169-176.
17. Hori, Y., et al., *Important cardiac transcription factor genes are accompanied by bidirectional long non-coding RNAs.* BMC Genomics, 2018. **19**(1): p. 967.
18. Mahpour, A., et al., *A methyl-sensitive element induces bidirectional transcription in TATA-less CpG island-associated promoters.* PLoS One, 2018. **13**(10): p. e0205608.
19. Yang, S., S. Sleight, and H. Sauro, *Rationally designed bidirectional promoter improves the evolutionary stability of synthetic genetic circuits.* Nucleic acids research, 2012. **41**.
20. Dawson, M.A. and T. Kouzarides, *Cancer epigenetics: from mechanism to therapy.* Cell, 2012. **150**(1): p. 12-27.
21. Jones, P.A. and D.J.S. Takai, *The role of DNA methylation in mammalian epigenetics.* 2001. **293**(5532): p. 1068-1070.
22. Herman, J.G. and S.B.J.N.E.J.o.M. Baylin, *Gene silencing in cancer in association with promoter hypermethylation.* 2003. **349**(21): p. 2042-2054.
23. Baylln, S.B., et al., *Alterations in DNA methylation: a fundamental aspect of neoplasia.* 1997. **72**: p. 141-196.

24. Compagni, A. and G.J.B.j.o.c. Christofori, *Recent advances in research on multistage tumorigenesis.* 2000. **83**(1): p. 1-5.
25. Venters, B.J., B.F.J.C.r.i.b. Pugh, and m. biology, *How eukaryotic genes are transcribed.* 2009. **44**(2-3): p. 117-141.
26. Whitehouse, I., et al., *Chromatin remodelling at promoters suppresses antisense transcription.* 2007. **450**(7172): p. 1031-1035.
27. Yadon, A.N., et al., *Chromatin remodeling around nucleosome-free regions leads to repression of noncoding RNA transcription.* 2010. **30**(21): p. 5110-5122.
28. Barski, A., et al., *High-resolution profiling of histone methylations in the human genome.* 2007. **129**(4): p. 823-837.
29. Wakano, C., et al., *The dual lives of bidirectional promoters.* 2012. **1819**(7): p. 688-693.
30. Core, L.J., J.J. Waterfall, and J.T.J.S. Lis, *Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters.* 2008. **322**(5909): p. 1845-1848.
31. Churchman, L.S. and J.S.J.N. Weissman, *Nascent transcript sequencing visualizes transcription at nucleotide resolution.* 2011. **469**(7330): p. 368-373.
32. Aken, B.L., et al., *The Ensembl gene annotation system.* 2016. **2016**.
33. Engström, P.G., et al., *Complex loci in human and mouse genomes.* 2006. **2**(4): p. e47.
34. Carrozza, M.J., et al., *Histone H3 methylation by Set2 directs deacetylation of coding regions by Rpd3S to suppress spurious intragenic transcription.* 2005. **123**(4): p. 581-592.
35. Lickwar, C.R., et al., *The Set2/Rpd3S pathway suppresses cryptic transcription without regard to gene length or transcription frequency.* 2009. **4**(3): p. e4886.
36. Tan-Wong, S.M., et al., *Gene loops function to maintain transcriptional memory through interaction with the nuclear pore complex.* 2009. **23**(22): p. 2610-2624.
37. Lainé, J.-P., et al., *A physiological role for gene loops in yeast.* 2009. **23**(22): p. 2604-2609.
38. Chen, Y.-Q., et al., *Sorting out inherent features of head-to-head gene pairs by evolutionary conservation.* 2010. **11**(11): p. 1-9.
39. Chen, Y., et al., *Pan-Cancer Analysis of Head-to-Head Gene Pairs in Terms of Transcriptional Activity, Co-expression and Regulation.* Frontiers in genetics, 2021. **11**: p. 560997-560997.
40. Ashburner, M., et al., *Gene ontology: tool for the unification of biology.* 2000. **25**(1): p. 25-29.
41. Hanahan, D. and R.A.J.c. Weinberg, *Hallmarks of cancer: the next generation.* 2011. **144**(5): p. 646-674.
42. Yang, M.Q., L.M. Koehly, and L.L.J.P.C.B. Elnitski, *Comprehensive annotation of bidirectional promoters identifies co-regulation among breast and ovarian cancer genes.* 2007. **3**(4): p. e72.
43. Xu, Z., et al., *Bidirectional promoters generate pervasive transcription in yeast.* 2009. **457**(7232): p. 1033-1037.
44. Preker, P., et al., *PROMoter uPstream Transcripts share characteristics with mRNAs and are produced upstream of all three major types of mammalian promoters.* 2011. **39**(16): p. 7179-7193.
45. Ntini, E., et al., *Polyadenylation site–induced decay of upstream transcripts enforces promoter directionality.* 2013. **20**(8): p. 923.
46. Chen, Y., et al., *Principles for RNA metabolism and alternative transcription initiation within closely spaced promoters.* 2016. **48**(9): p. 984-994.
47. Uesaka, M., et al., *Bidirectional promoters are the major source of gene activation-associated non-coding RNAs in mammals.* BMC Genomics, 2014. **15**(1): p. 35.
48. Li, D., et al., *PLAGL2 and POFUT1 are regulated by an evolutionarily conserved bidirectional promoter and are collaboratively involved in colorectal cancer by maintaining stemness.* EBioMedicine, 2019. **45**: p. 124-138.

49. Baldwin, R.L., et al., *BRCA1 promoter region hypermethylation in ovarian carcinoma: a population-based study.* Cancer Res, 2000. **60**(19): p. 5329-33.
50. Xu, C., et al., *Isolation and characterisation of the NBR2 gene which lies head to head with the human BRCA1 gene.* Human molecular genetics, 1997. **6**: p. 1057-62.
51. Chatterjee, B., K. Ghosh, and S.R. Kanade, *Resveratrol modulates epigenetic regulators of promoter histone methylation and acetylation that restores BRCA1, p53, p21CIP1 in human breast cancer cell lines.* 2019. **45**(5): p. 818-829.
52. Amatya, V.J., et al., *TP53 promoter methylation in human gliomas.* Acta Neuropathol, 2005. **110**(2): p. 178-84.
53. Reisman, D. and P. Zeigler, *The Bi-directional Nature of the Promoter of the p53 Tumor Suppressor Gene.* Journal of Leukemia, 2015. **03**.
54. Mahmoudi, S., et al., *WRAP53 promotes cancer cell survival and is a potential target for cancer therapy.* Cell Death Dis, 2011. **2**(1): p. e114.
55. Drak Alsibai, K., et al., *High Positive Correlations between ANRIL and p16-CDKN2A/p15-CDKN2B/p14-ARF Gene Cluster Overexpression in Multi-Tumor Types Suggest Deregulated Activation of an ANRIL-ARF Bidirectional Promoter.* Noncoding RNA, 2019. **5**(3).
56. Yap, K.L., et al., *Molecular interplay of the noncoding RNA ANRIL and methylated histone H3 lysine 27 by polycomb CBX7 in transcriptional silencing of INK4a.* Mol Cell, 2010. **38**(5): p. 662-74.
57. Davis, R.J., *Signal transduction by the JNK group of MAP kinases.* Cell, 2000. **103**(2): p. 239-52.
58. Zhu, J.H., et al., *Protein tyrosine phosphatase PTPN13 negatively regulates Her2/ErbB2 malignant signaling.* Oncogene, 2008. **27**(18): p. 2525-31.
59. Yeh, S.H., et al., *Genetic characterization of fas-associated phosphatase-1 as a putative tumor suppressor gene on chromosome 4q21.3 in hepatocellular carcinoma.* Clin Cancer Res, 2006. **12**(4): p. 1097-108.
60. Ying, J., et al., *Epigenetic disruption of two proapoptotic genes MAPK10/JNK3 and PTPN13/FAP-1 in multiple lymphomas and carcinomas through hypermethylation of a common bidirectional promoter.* Leukemia, 2006. **20**(6): p. 1173-1175.
61. Li, L.X., et al., *Lysine methyltransferase SMYD2 promotes cyst growth in autosomal dominant polycystic kidney disease.* The Journal of clinical investigation, 2017. **127**(7): p. 2751-2764.
62. Xu, L., et al., *An emerging role of PARK2 in cancer.* Journal of Molecular Medicine, 2014. **92**(1): p. 31-42.
63. Kim, S.W., et al., *Interferon-gamma-induced expressions of heat shock protein 60 and heat shock protein 10 in C6 astroglioma cells: identification of the signal transducers and activators of transcription 3-binding site in bidirectional promoter.* Neuroreport, 2007. **18**(4): p. 385-9.
64. Knowlton, A.A. and U. Srivatsa, *Heat-shock protein 60 and cardiovascular disease: a paradoxical role.* Future Cardiol, 2008. **4**(2): p. 151-61.
65. Henderson, B., M.A. Fares, and P.A. Lund, *Chaperonin 60: a paradoxical, evolutionarily conserved protein family with multiple moonlighting functions.* Biological Reviews, 2013. **88**(4): p. 955-987.
66. Hanahan, D. and R.A.J.c. Weinberg, *The hallmarks of cancer.* 2000. **100**(1): p. 57-70.
67. Zhang, H., et al., *Distinct roles of nucleosome sliding and histone modifications in controlling the fidelity of transcription initiation.* RNA Biology, 2020: p. 1-11.
68. Vafai, S.B. and V.K.J.N. Mootha, *Mitochondrial disorders as windows into an ancient organelle.* 2012. **491**(7424): p. 374-383.
69. Scarpulla, R.C.J.P.r., *Transcriptional paradigms in mammalian mitochondrial biogenesis and function.* 2008. **88**(2): p. 611-638.
70. Uchiumi, F., et al., *Implication of bidirectional promoters containing duplicated GGAA motifs of mitochondrial function-associated genes.* 2014. **1**(1): p. 1-26.

71. Szczesny, B., et al., *Long patch base excision repair in mammalian mitochondrial genomes.* 2008. **283**(39): p. 26349-26356.
72. Matoba, S., et al., *p53 regulates mitochondrial respiration.* 2006. **312**(5780): p. 1650-1653.
73. Scatena, R., P. Bottoni, and B. Giardina, *Advances in mitochondrial medicine*. Vol. 942. 2012: Springer Science & Business Media.
74. Zong, W.-X., J.D. Rabinowitz, and E.J.M.c. White, *Mitochondria and cancer.* 2016. **61**(5): p. 667-676.
75. Lezi, E. and R.H.J.A.i.M.M. Swerdlow, *Mitochondria in neurodegeneration.* 2012: p. 269-286.
76. Hu, H.Y., L. He, and P.J.B.g. Khaitovich, *Deep sequencing reveals a novel class of bidirectional promoters associated with neuronal genes.* 2014. **15**(1): p. 1-16.
77. Deng, H., et al., *Genetic analysis of parkin co-regulated gene (PACRG) in patients with early-onset parkinsonism.* 2005. **382**(3): p. 297-299.
78. Zanotto, E., Z.H. Shah, and H.T.J.N.a.r. Jacobs, *The bidirectional promoter of two genes for the mitochondrial translational apparatus in mouse is regulated by an array of CCAAT boxes interacting with the transcription factor NF-Y.* 2007. **35**(2): p. 664-677.
79. Hansen, J.J., et al., *Genomic structure of the human mitochondrial chaperonin genes: HSP60 and HSP10 are localised head to head on chromosome 2 separated by a bidirectional promoter.* 2003. **112**(1): p. 71-77.
80. Dunn, C.A., et al., *Transcription of two human genes from a bidirectional endogenous retrovirus promoter.* 2006. **366**(2): p. 335-342.
81. Bellizzi, D., et al., *Characterization of a bidirectional promoter shared between two human genes related to aging: SIRT3 and PSMD13.* 2007. **89**(1): p. 143-150.
82. Galli, F., et al., *Relevance of immune cell and tumor microenvironment imaging in the new era of immunotherapy.* Journal of Experimental & Clinical Cancer Research, 2020. **39**(1): p. 89.
83. Seila, A.C., et al., *Divergent transcription from active promoters.* 2008. **322**(5909): p. 1849-1851.
84. Kambara, H., et al., *Regulation of interferon-stimulated gene BST2 by a lncRNA transcribed from a shared bidirectional promoter.* 2015. **5**: p. 676.
85. Schneider, W.M., M.D. Chevillotte, and C.M.J.A.r.o.i. Rice, *Interferon-stimulated genes: a complex web of host defenses.* 2014. **32**: p. 513-545.
86. Juszczynski, P., et al., *BAL1 and BBAP are regulated by a gamma interferon-responsive bidirectional promoter and are overexpressed in diffuse large B-cell lymphomas with a prominent inflammatory infiltrate.* 2006. **26**(14): p. 5348-5359.
87. Baran-Marszak, F., et al., *Differential roles of STAT1α and STAT1β in fludarabine-induced cell cycle arrest and apoptosis in human B cells.* 2004. **104**(8): p. 2475-2483.
88. Howe, K.L., et al., *Ensembl 2021.* Nucleic acids research, 2021. **49**(D1): p. D884-D891.
89. Quinlan, A.R. and I.M.J.B. Hall, *BEDTools: a flexible suite of utilities for comparing genomic features.* 2010. **26**(6): p. 841-842.
90. Kanehisa, M. and S.J.N.a.r. Goto, *KEGG: kyoto encyclopedia of genes and genomes.* 2000. **28**(1): p. 27-30.
91. Botstein, D., et al., *Gene Ontology: tool for the unification of biology.* 2000. **25**(1): p. 25-9.
92. 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.
93. Chen, J., et al., *LncSEA: a platform for long non-coding RNA related sets and enrichment analysis.* 2021. **49**(D1): p. D969-D980.
94. Tang, Z., et al., *GEPIA2: an enhanced web server for large-scale expression profiling and interactive analysis.* 2019. **47**(W1): p. W556-W560.
95. Stel, V.S., et al., *Survival analysis I: the Kaplan-Meier method.* 2011. **119**(1): p. c83-c88.

96. Chandrashekar, D.S., et al., *UALCAN: a portal for facilitating tumor subgroup gene expression and survival analyses.* 2017. **19**(8): p. 649-658.
97. Li, T., et al., *TIMER: a web server for comprehensive analysis of tumor-infiltrating immune cells.* 2017. **77**(21): p. e108-e110.
98. Li, T., et al., *TIMER2. 0 for analysis of tumor-infiltrating immune cells.* 2020. **48**(W1): p. W509-W514.
99. Chen, B., et al., *Profiling tumor infiltrating immune cells with CIBERSORT*, in *Cancer systems biology*. 2018, Springer. p. 243-259.
100. Finotello, F., et al., *quanTIseq: quantifying immune contexture of human tumors.* 2017: p. 223180.
101. 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.
102. Becht, E., et al., *Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression.* 2016. **17**(1): p. 1-20.
103. Racle, J. and D. Gfeller, *EPIC: A Tool to Estimate the Proportions of Different Cell Types from Bulk Gene Expression Data.* Methods Mol Biol, 2020. **2120**: p. 233-248.
104. Ling, B., et al., *Gene Expression Correlation for Cancer Diagnosis: A Pilot Study.* BioMed Research International, 2014. **2014**: p. 253804.
105. Pages, F., et al., *Immune infiltration in human tumors: a prognostic factor that should not be ignored.* 2010. **29**(8): p. 1093-1102.
106. Busby, S. and R.H.J.C. Ebright, *Promoter structure, promoter recognition, and transcription activation in prokaryotes.* 1994. **79**(5): p. 743-746.
107. Nagano, T. and P.J.C. Fraser, *No-nonsense functions for long noncoding RNAs.* 2011. **145**(2): p. 178-181.
108. Lee, J.T.J.S., *Epigenetic regulation by long noncoding RNAs.* 2012. **338**(6113): p. 1435-1439.
109. Paraskevopoulou, M.D., et al., *DIANA-LncBase: experimentally verified and computationally predicted microRNA targets on long non-coding RNAs.* 2013. **41**(D1): p. D239-D245.
110. Blazer, D.G. and L.M. Hernandez, *Genes, behavior, and the social environment: Moving beyond the nature/nurture debate.* 2006.
111. Kuo, W.P., et al., *Functional relationships between gene pairs in oral squamous cell carcinoma.* AMIA ... Annual Symposium proceedings. AMIA Symposium, 2003. **2003**: p. 371-375.
112. Stegh, A.H., et al., *Glioma oncoprotein Bcl2L12 inhibits the p53 tumor suppressor.* Genes & development, 2010. **24**(19): p. 2194-2204.
113. Pattwell, S.S. and E.C. Holland, *Putting Glioblastoma in Its Place: IRF3 Inhibits Invasion.* Trends in molecular medicine, 2017. **23**(9): p. 773-776.
114. Wang, Y., et al., *CD8+ T Cell Co-Expressed Genes Correlate With Clinical Phenotype and Microenvironments of Urothelial Cancer.* Frontiers in oncology, 2020. **10**: p. 553399-553399.
115. Xiao, J., et al., *Integrated analysis of physiological, transcriptomic and metabolomic responses and tolerance mechanism of nitrite exposure in Litopenaeus vannamei.* 2020. **711**: p. 134416.