BIOINFORMATICS ANALYSIS OF MERKEL CELL CARCINOMA TO IDENTIFY CRITICAL GENES AND THEIR VALIDATION

A PROJECT REPORT

SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS

FOR THE AWARD OF THE DEGREE

OF

MASTER OF TECHNOLOGY
IN
BIOINFORMATICS ENGINEERING

Submitted by:

HIMANSHU KUMAR

(2K18/BIO/07)

Under the supervision of

DR. YASHA HASIJA (Associate Professor)

DEPARTMENT OF BIOTECHNOLOGY

DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering)

Bawana Road, Delhi-110042

JUNE, 2020

DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Bawana Road, Delhi-110042

CANDIDATE'S DECLARATION

I, Himanshu Kumar, 2K18/BIO/07 of M.Tech (Bioinformatics), hereby declare that the project Dissertation titled "**BIOINFORMATICS ANALYSIS OF MERKEL CELL CARCINOMA TO IDENTIFY CRITICAL GENES AND THEIR VALIDATION**" which is submitted by me to the Department of Biotechnology, Delhi Technological University, Delhi in partial fulfilment of the requirements for the award of the degree of Master of Technology, is original and not copied from any source with proper citation. This work has not previously formed the basis for the award of the Degree, Diploma Associateship, Fellowship or other similar title or recognition.

Himanshu

Place: Delhi **HIMANSHU KUMAR**

Date: (2K18/BIO/07)

DEPARTMENT OF BIOTECHNOLOGY
DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Bawana Road, Delhi-110042

CERTIFICATE

I hereby certify that the Project Dissertation titled **“BIOINFORMATICS ANALYSIS OF MERKEL CELL CARCINOMA TO IDENTIFY CRITICAL GENES AND THEIR VALIDATION”** which is submitted by **Himanshu Kumar(2K18/BIO/07)**, Department of Biotechnology, Delhi Technological University, Delhi in partial fulfillment of the requirement for the award of the degree of Master of Technology, is a record of the project work carried out by the student under my supervision. To the best of my knowledge this work has not been submitted in part or full for any Degree or Diploma to this University or elsewhere.

Place: Delhi

Date:

DR. YASHA HASIJA
Associate Professor

ACKNOWLEDGEMENT

I, Himanshu Kumar wish to express my profound gratitude and indebtedness to **Dr. Yasha Hasija,** Department of Bio-Technology, Delhi Technological University, Delhi for introducing the present topic and for her inspiring guidance, constructive criticism and valuable suggestion throughout the project work. And my sincere thanks to all our friends who have patiently extended all sorts of help for accomplishing this undertaking.

Himanshu Kumar (2K18/BIO/07)

Dept. of Bio-Technology,

DTU, Delhi-110042

BIOINFORMATICS ANALYSIS OF MERKEL CELL CARCINOMA TO IDENTIFY CRITICAL GENES AND THEIR VALIDATION

HIMANSHU KUMAR

Delhi Technological University, Delhi, India

Email ID: himanshu97kumar@gmail.com

ABSTRACT

Merkel Cell Carcinoma (MCC) is a destructive form of neuroendocrine carcinoma that represents the 2nd highest cause of skin cancer related deaths despite being a relatively rare form of cancer. Different risk factors have been associated with MCC such as advancement age, immunosuppression, and ultraviolet light exposure, polyomavirus infection and a current, concurrent or previous diagnosis of Chronic Lymphocytic Leukaemia (CLL). However, the exact mechanism that leads to MCC and its inherent biology is yet to be fully researched and explored. The current treatments for MCC are a combination of surgery and radiation or chemo-radiation. So far, only two drugs have been approved by FDA for the treatment of MCC, with both drugs being immune checkpoint inhibitors. However, the drug approval from FDA was accelerated based on promising results from a relatively small number of patients, which means that the more research and trials are undergoing currently considering the fact that a very small cohort was used for clinical trials. Presently there are no biomarkers indicative of MCC and additional research is essential. With the advent of bioinformatics and microarray technology, in silico approaches pertaining to 'omics' data analysis, shall allow identification of hub genes and miRNA to broaden our understanding of MCC biology and their clinical utility.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

MCC	Merkel Cell Carcinoma
UV	Ultraviolet
CLL	Chronic Lymphocytic Leukaemia
FDA	Food and Drug Administration
HIV	Human Immunodeficiency Syndrome
miRNA	Micro Ribonucleic Acid
DEGs	Differential Expression Genes
DEmiRs	Differential Expression microRNA
FC	Fold Change
GEO	Gene Expression Omnibus
GO	Gene Ontology
BP	Biological Processes
MF	Molecular Function
CC	Cellular Components
PPI	Protein-Protein Interaction

1. INTRODUCTION

Merkel Cell Carcinoma (MCC) is destructive form of neuroendocrine carcinoma whose manifestation of occurrence have been on a rise in recent years. It represents the 2nd highest cause of skin cancer related deaths despite being a relatively rare form of cancer [1]. A series of ultrastructural studies revealed the similarity of these cancer cells to that of Merkel Cells (a type of mechanoreceptor cells present in the skin), which led to the renaming of these cells from trabecular carcinoma to Merkel Cell Carcinoma [2]. Different risk factors have been associated with MCC such as advancement in age, immunosuppression, and exposure to UV. A link between MCC and PyV infection has also been established. An existing, concurrent or even a previous diagnosis of Chronic Lymphocytic Leukaemia (CLL) is more frequent in people with MCC [3]. However, the exact biology of MCC and its carcinogenesis is yet to be fully explored.

The current treatments for MCC are surgery along with radiation or chemo-radiation. The first FDA approved drug for MCC was avelumab based on a clinical trial of eighty-eight patients diagnosed with metastatic MCC who had formerly undergone chemo-radiation for the same [4]. A second drug was approved in the beginning of 2019, called pembrolizumad. Both these drugs are immune checkpoint inhibitors [5]. As a result, there has been a complete sea change in the management of MCC patients in the last two years. Prior to using immunotherapy as a first line of treatment, MCC patients were treated just like any other individual with a neuroendocrine tumor. However, the current approval from FDA has been an accelerated one, based on promising results from a relatively small number of patients, which means that the more research and trials are undergoing currently considering the fact that a very small cohort was used for clinical trials. Although, 56% of the drug recipients have shown a positive response, many stopped using the drug due to severe side effects. A further issue is that many patients with MCC may not be ideal candidates for immunotherapy because their immune systems have been suppressed—due to organ/graft transplantation, lymphoma or HIV [5].

At the moment, there is a lack of definitive biomarkers that could aid in a more effective selection of patients for new therapies, thereby creating a crucial need for further research. The present study is therefore an attempt to identify key genes and microRNAs in MCC by various in silico methodologies.

2. MATERIALS AND METHODS

2.1 Microarray data processing and screening for DEGs and DEmiRs

In this study, the expression profile GSE39612 from the GEO database was used to identify DEGs. The series GSE39612 provides transcription data of 30 tumors from 27 different patients from which samples of MCC and normal skin were acquired. DEGs were identified by using Subio platform (Subio Inc., Japan) where a p-value of < 0.05 and $|FC| > 1.5$ was considered as the cut-off. Furthermore, the series GSE45146 was used to identify DEmiRNAs between normal skin samples and MCC. This analysis was performed by using GEO2R tool where DEmiRNAs are considered on the basis of statistical measures that met adjusted P-value (adj. P) <0.01 and |log fold change| >1.5 of miRNAs.

2.2 Pathway and GO enrichment analysis of DEGs

To understand the functional changes of the DEGs, their cellular components (CC), biological processes (BP) and molecular function (MF) were analysed by DAVID [9]. Only GO terms with p-value < 0.05 and count > 2 were measured statistically important.

Enrichment of pathways was also performed by using DAVID. The DEGs were used to perform a Fisher’s exact test followed by multiple test correction using BH’s False Discovery Rate correction. A revised p-value of < 0.05 was considered statistically important to determine the overrepresentation of certain biological pathways.

2.3 PPI network construction

Cytoscape is used along with STRING database [25] to construct a PPI network. Only query proteins were displayed and a lowest required interaction score of 0.4 (medium confidence) was kept as the primary parameter. Subsequently, significant cluster screening was done by using another Cytoscape plugin, MCODE, using their pre-set cut- off criteria [2].

2.4 Identification of hub genes using cytoHubba

Five algorithms, namely, MCC, MNC, EPC, Degree and EcCentricity were determined using cytoHubba plug-in to categorize hub genes from the PPI network. The overlapped genes of the five algorithms mentioned above were further enriched in terms of KEGG pathways, GO molecular functions, GO biological processes and GO cellular components using DAVID 6.8

2.5 Target gene prediction of DE microRNA

To identify putative target genes of the DEmiRs, three online resources were used. TargetScan Human version 7.2 and miRDB were used to identify predictive miRNA-gene pairs while miRTarBase was used to identify experimentally validated miRNA-gene pairs.

2.6 Construction and analysis of DEG-DEmiR network

The DEGs and DEmiRs were used towards constructing a dysregulation network using Cytoscape. Using Network Analyser tool of Cytoscape, the above network was analysed for the purpose of finding hub miRNAs. A degree cut-off of > 20 was set as the criteria. Hub node annotation in terms of GO molecular function, GO biological process, GO cellular components and KEGG pathway were again performed by DAVID.

3. RESULTS

3.1 Identification of DEGs and DEmiRs from datasets

To find DEGs between normal skin samples and MCC, GSE39612 series was used in Subio platform. DEGs are evaluated by fold change of 1.5 and p-value < 0.05. A total of 449 genes were recognized which included 285 downregulated genes and 164 upregulated genes. Whereas, the series GSE45146 was used to identify DEmiRNAs between normal skin samples and MCC using GEO2R. DEmiRNAs are considered on the basis of statistical measures that met adjusted P-value (adj. P) <0.01 and |log fold change| >1.5 of miRNAs. A total of 17 DEmiRNAs were identified which included 1 upregulated miRNA and 16 downregulated miRNAs.

3.2 Functional enrichment analysis of DEGs

To recognize the function of the DEGs, their biological processes (BP), cellular components (CC) and molecular function (MF) were analysed by DAVID. The DEGs were mainly enhanced in oxidation-reduction process (GO: 0055114), keratinocyte differentiation (GO: 0030216), epidermis development (GO: 0008544) and positive regulation of cell proliferation (GO: 0008284) in the BP group. In CC group, the DEGs were mostly enriched in plasma membrane (GO: 0005886), cytoplasm (GO: 0005737) and extracellular space (GO: 0005615). Finally, in the MF group, the genes enriched in calcium ion binding (GO: 0005509), protein homo-dimerization activity (GO: 0042803) and structural molecule activity (Table 3.1, 3.2). Meanwhile, significantly enriched pathways highlighted were cAMP signalling pathway (hsa04024), Calcium signalling pathway (hsa04020), Oocyte meiosis (hsa04114), Arachidonic acid metabolism (hsa00590), Vascular smooth muscle contraction (hsa04270), Inflammatory mediator regulation of TRP channels (hsa04750), GnRH signalling pathway (hsa04912), p53 signalling pathway (hsa04115), Aldosterone synthesis and secretion (hsa04925) and HIF-1 signalling pathway (hsa04066).

Table 3.1: A tabular representation of analysis of significant enrichment GO terms of DEGs present in MCC

Term	Description	Count	P Value	Genes
Biological Process				
GO:0045944	RNA polymerase II promoter positive regulation of transcription	46	494E-05	FGFR,HLF,IL18,FOXM1,F2RL1,SOX2,BEX1,PAX6,
GO:0008284	cell proliferation positive regulation	29	1.53E-05	FGFR2,KRT6A,CCK,IL6ST, FOXM1,TTK,SOX9,
GO:0006366	transcription from RNA polymerase II promoter	27	0.000471	HLF,FOXM1,SOX2,PAX6, PAX5,TP63,EHF,MYBL1,
GO:0055114	oxidation-reduction process	25	0.012791	ME1,TYRP1,HSD3B1, CYP1B1,EGLN3,BBOX1,
GO:0008544	epidermis development	24	5.00E-18	KLK7,KLK5,LCE2B,KRT31, GJB5,CDSN,SCEL
GO:0007155	cell adhesion	24	0.001142	CLCA2,TNXB,CYP1B1, SLURP1,ADAM23,CTNND2
GO:0045893	DNA-templated, Positive regulation of transcription	24	0.004885	KLF5,SOC1,FOXM1,PSRC1, SOX2,PAX6,IGF1,
Cellular Components				
GO:0005737	Cytoplasm	158	0.000572	PALMD,DCN,KRT5,NCAPG, HJURP,POU2F3,
GO:0005886	plasma membrane	135	5.74E-05	TBC1D30,FCER1A,KLF5, VAV3,ADAM23,IL1RN,
GO:0070062	extracellular exosome	126	1.19E-12	CLEC3B,FBLN5,SPRR1B, AOX1,DSC2,DSC1,CP
GO:0005829	Cytosol	99	0.014265	CENPE,CAPN3,RGS13,SCG, AOX1,RAB38,HPGD
GO:0005576	extracellular region	90	5.16E-14	SCGN,OMD,FBLN1,CD55, CCL14,NPY,SFRP1,
GO:0005615	extracellular space	85	3.42E-16	FBLN1,OMD,CCL14,NPY,

				SFRP1,CLEC3B,FBLN5,
GO:0005887	integral component of plasma membrane	47	0.022283	P2RY14,NTRK2,KCNH6, CLDN1,PDGFRA,TM4SF1,
GO:0005794	Golgi apparatus	28	0.098523	NPY,DUSP26,ATP8A2, RAB38,PERP,KIF20A
Molecular Function				
GO:0005515	protein binding	249	0.001438	KIF18A,CENPF,DPYSL4, CENPE,DACH1,ISL1,
GO:0005509	calcium ion binding	48	1.12E-09	DSC2,LRP8,DSC1,NCAN, DST,PROS1,CACNA1A,
GO:0042803	protein homo-dimerization activity	32	0.002538	FBLN5,KCNN2,NTRK2, PDGFRA,CLIP1,CRRBA2,
GO:0005198	structural molecule activity	27	5.44E-10	KRT15,KRT14,CLDN1,DSP, CSTA,ADD2,
GO:0001077	RNA polymerase II core promoter, transcriptional activator activity	21	1.84E-06	POU4F2,POU3F2,ZNF750, POU4F1,KLF4,PITX2,
GO:0005200	structural constituent of cytoskeleton	19	2.10E-10	KRT14,DSP,ACTL6B,NEFH, KRT2,NEFL,NEFM,
GO:0008201	heparin binding	14	0.000178	FGFR2,TNXB,ECM2,GREM2,APLP1,PRELP,

Table 3.2: A tabular representation of significant KEGG pathways identified for the DEGs in this dataset. Terms with p-value < 0.05 were considered as significant.

Term	Description	p-Value	Genes
hsa04024	cAMP signaling pathway	0.006951985	ATP2B2,ADCY1,PLD1,VAV3, NPY,CALML3,PLN,
hsa04020	Calcium signaling pathway	0.008673811	EGFR,GNAL,ATP2B2,AGTR1, ADCY1,CALML3,PLN,
hsa05146	Amoebiasis	0.007513759	GNAL,IL1R2,ARG1,LAMB4, ADCY1,SERPINB2,
hsa04114	Oocyte meiosis	0.009824526	CCNE2,ADCY1,CALM3,BUB1, FBXO5,IGF1,AURKA,
hsa00590	Arachidonic acid metabolism	0.001161382	AKR1C3,PLA2G4A,PTGIS, ALOX15B,PLA2G2A,EPHX2,
hsa05214	Glioma	0.001693434	EFGR,CDKN2A,CALML3, PDGFRA,IGF1,CAMK2B
hsa04750	Inflammatory mediator regulation of TRP channels	0.016071416	ADCY1,PLA2G4A,CALML3, F2RL1,IGF1,CAMK2B,
hsa04270	Vascular smooth muscle contraction	0.037827198	ACTG2,AGTR1,ADCY1, PLA2G4A,CALML3,CALD1,
hsa04115	p53 singaling pathway	0.008943032	CCNE2,CDKN2A,RRM2, RPRM,IGF1,PERP,GTSE1
hsa04925	Adosterone synthesis and secretion	0.021401112	AGTR1,ADCY1,HSDB1, CALML3,CAMK2B,CALML5,
hsa04912	GnRH signaling pathway	0.035359873	EFGR,ADCY1,PLA2G4A, PLD1,CALML3,CAMK2B,
hsa04066	HIF-1 signaling pathway	0.044130154	EFGR,TF,EGLN3,ENO2,IGF1, CAMK2B,PIK3R1
hsa04913	Ovarian steroidogenesis	0.009751094	AKR1C3,ADCY1,PLA2G4A, HSD3B1,CYP1B1,IGF1
hsa04924	Renin secretion	0.028291683	AGTR1,CLCA2,CLCA4, CALML3,CALML5,AQP1
hsa03320	PPAR signaling pathway	0.033623759	CPTIB,NPY,LEPR,ADIPOQ, ACSL6,ACSBG1
hsa04920	Adipocytokine signaling pathway	0.039542173	CPT1B,NPY,LEPR,ADIPOQ, ACSL6,ACSBG1
hsa05218	Melanoma	0.041647285	EFGR,CDKN2A,PDGFRA,IGF1, CDH1,PIK3R1
hsa00350	Tyrosine metabolism	0.013828406	DCT,TYRP1,AOX1,ADH1B, ALDH3B2
hsa00565	Ether lipid metabolism	0.032014452	PLA2G4A,PLD1,PLA2G2A, PAFAH1B1,PLPP3
hsa01212	Fatty acid metabolism	0.039315876	CPT1B,FLOVL2,SCD5,ACSL6, ACSBG1
hsa04614	Renin-angiotensin system	0.022816032	AGTR1,CPA3,CMA1,CTSG

3.3 Construction of PPI network and its module analysis

With the help of STRING database on Cytoscape, PPI network was constructed with a pre-set criteria of displaying only query proteins and a minimum required interaction score of 0.4 as primary parameters. A network of 448 nodes and 2190 edges were included in the DEGs network. Subsequently, significant module selection was done by using, MCODE, using their pre-set cut- off criteria. The clusters were primarily enriched in cell cycle checkpoints, cell cycle, structural constituent of epidermis, regulation of cell cycle, cornified envelope, cell-cell adhesion and immune system.

3.4 Identification of Hub Genes and Pathways through analysis of PPI network

CytoHubba was employed to observe the intersection of five algorithms (degree, MCC, MNC, EPC and EcCentricity) to identify the hub genes present in the network. As seen from Figure 3.1, 16 genes were found in the intersection of the five algorithms. These genes include: AURKA, FLG, KIF4A, DSG1, PKP1, LOR, BUB1, CDC6, CENPE, CXCR4, FOXM1, CXCL12, CDSN, DSG3, IVL and DSP.

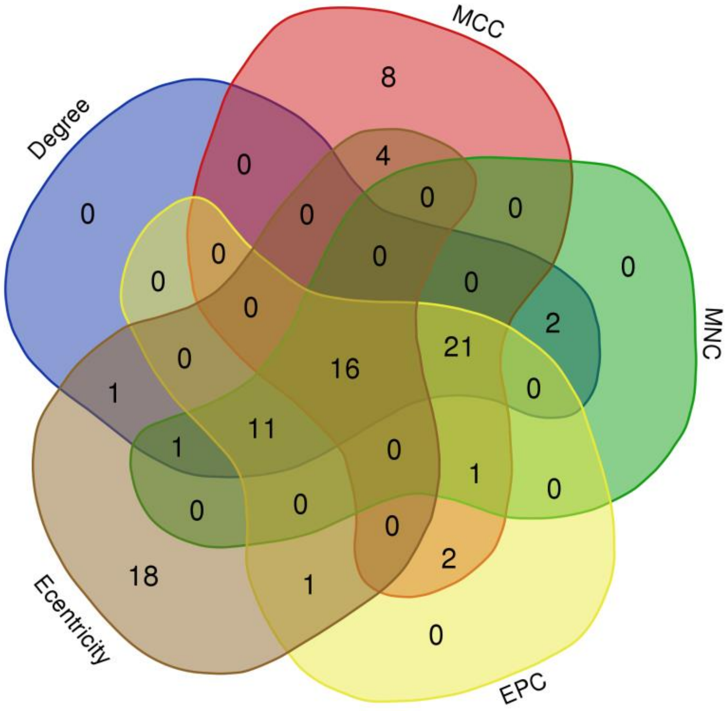

Figure 3.1: Representation of Venn plot by intersecting 5 algorithms to recognize significant hub genes. Different algorithms are represented by different colours.

Commonly collected DEGs is the cross section all colours. The divisions in coexisting areas are the 16 hub genes (AURKA, FLG, KIF4A, DSG1, PKP1, LOR, BUB1, CDC6, CENPE, CXCR4, FOXM1, CXCL12, CDSN, DSG3, IVL and DSP).

The 16 hubs genes included 7 upregulated and 9 downregulated genes and were used towards constructing a network with 34 edges and 16 nodes. A network of these 16 hub genes revealed that the upregulated and downregulated genes were part of two individual clusters, as shown in Figure 3.2. A STRING enrichment analysis (GO components, GO processes, GO functions, KEGG pathways and Reactome pathways) was performed for the hub genes, keeping the FDR < 0.05. In the GO components group, the genes were mostly enriched in cornified envelope (GO.0001533), desmosome (GO.0030057), cytoskeletal part (GO.0044430), plasma membrane (GO.0005886) and cytosol (GO.0005829). In the GO process group, most genes were enriched in programmed cell death (GO.0012501), cell differentiation (GO.0030154), animal organ development (GO.0048513), cornification (GO.0070268), anatomical structure development (GO.0048856) and system development (GO.0048731). Further, genes were mainly enriched in protein binding (GO.0005515) and binding (GO.0005488) in the GO functions group. For pathway enrichment, results from both KEGG and Reactome were considered. Eight KEGG pathways were highlighted which included Cell cycle (hsa04110), Leukocyte trans-endothelial migration (hsa04670), Oocyte meiosis (hsa04114), Progesterone-mediated oocyte maturation (hsa04914), Chemokine signaling pathway (hsa04062), Intestinal immune network for IgA production (hsa04672), Axon guidance (hsa04360) and Cytokine-cytokine receptor interaction (hsa04060). On the other hand, reactome highlighted eighteen pathways, with the same statistical conditions intact. The major pathways highlighted were formation of the cornified envelope (HSA-6809371), developmental biology (HSA-1266738), immune system (HSA-168256) and cell cycle, mitotic (HSA-69278).

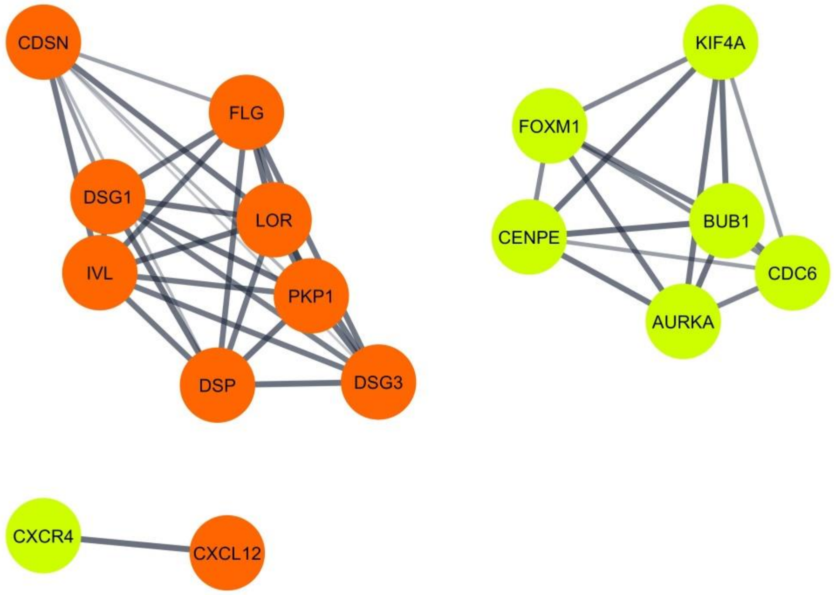

Figure 3.2: PPI Network of the hub genes with 16 nodes and 34 edges. The 16 hub genes were used to construct a network using STRING database in Cytoscape. The ‘orange’ nodes specify the downregulated genes while the ‘green’ nodes specify upregulated genes.

3.5 Target gene prediction of DE microRNA

A total of 55,684 target interactions (experimental and predictive) were identified as putative targets for the 17 DEmiRs gathered in this study. The data acquired was then used to identify validated and predictive pairs of DEGs-DEmiRs from this study. Out of a possible of 956 individual genes, 28 genes were found to be overlapping with the DEGs as well. These include ABHD5, CA12, CD164, CDH1, CLIP1, COLEC12, CXCR4, DEPDC1, EGLN3, ELAVL2, EREG, HIPK3, IGF1, IGF2BP3, KLF4, KLF5, MBNL1, MBNL2, NAB1, PICALM, RECK, RORA, SOX9, TM4SF1, TWF1, VAMP3, VSNL1 and WDR1. 56 pairs of validated DEGs-DEmiRs and 221 pairs of predicted DEGs-DEmiRs were curated from miRTarBase and TargetScan & miRDB respectively.

3.6 Construction and analysis of DEG-DEmiR network

The DEGs and DEmiRs were used towards constructing a network using Cytoscape which consisted of 47 nodes and 267 edges. Using Network Analyser tool of Cytoscape, the above network was analysed for the purpose of finding hub miRNAs. A degree cut-off of > 20 was set as the criteria, and subsequently, three major hubs identified were hsa-miR-19a, hsa-miR-9-5p and hsa-miR-1. The hubs were primarily enriched in positive regulation of transcription, positive regulation of mitotic nuclear division, cadherin binding involved in cell-cell adhesion, positive regulation of fibroblast proliferation, protein kinase B signalling and pathways in cancer.

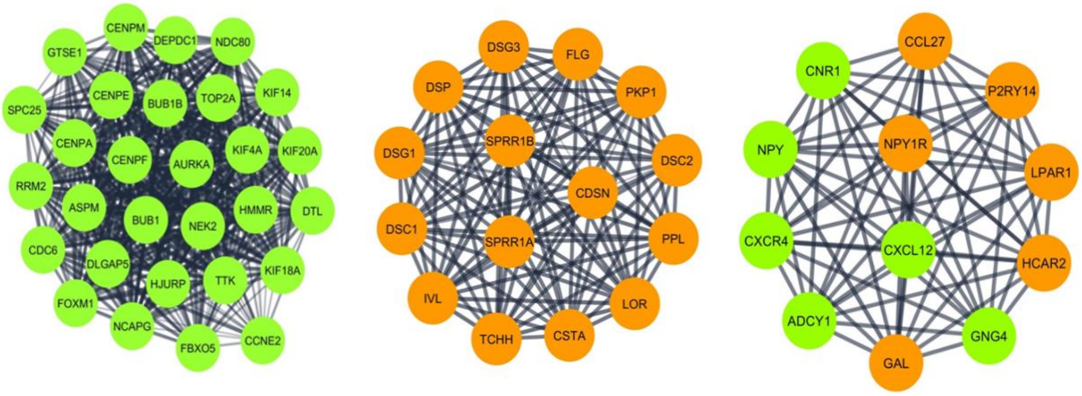

Figure 3.3: Significant clusters identified from PPI using MCODE plugin. The circular orange nodes represent downregulated genes while the circular green nodes represent upregulated genes.

4. VALIDATION

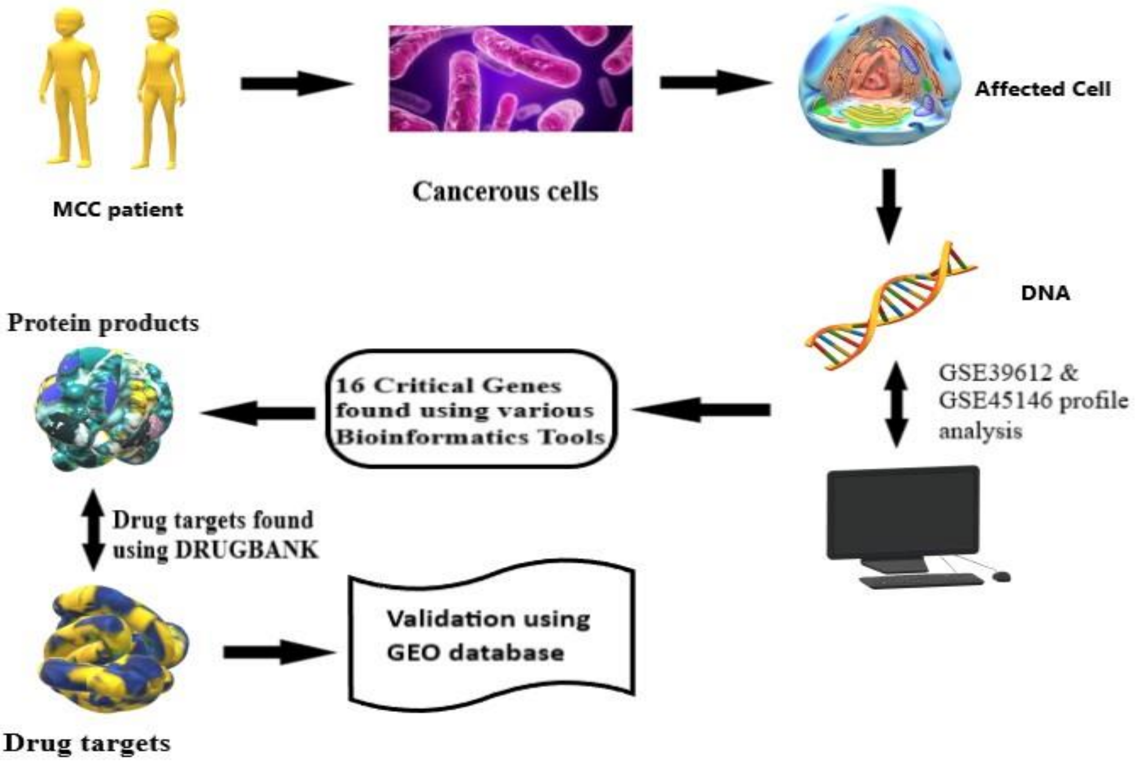

Figure 4.1: Overall strategy applied to find out critical genes and validate them for Merkel Cell Carcinoma (MCC)

16 Critical genes were validated through disease association microarray study approach. UniProt was employed to find out protein products of all these 16 critical genes. Then, DRUGBANK was used to find out the drug targets associated with these protein products (Table 4.1).

Table 4.1: Tabular representation of critical genes, their proteins, drug targets, and GEO dataset.

GENE	PROTEIN	DRUG TARGETS	DRUGBANK ID	GEO ID
AURKA	Aurora kinase A	Phosphonothreonine, 2-(1H-pyrazol-3-yl)-1H-benzimidazole	DB02482, DB08065	
FLG	Filaggrin	Benzamidine	DB03127	
KIF4A	Chromosome-associated kinesin KIF4A			
DSG1	Desmoglein-1	Dasatinib	DB01254	GSE39073
PKP1	Plakophilin-1			
LOR	Loricrin			
BUB1	Mitotic checkpoint serine/threonine-protein kinase BUB1	N-[(1S)-2-amino-1-phenylethyl]-5-(1H-pyrrolo[2,3-b]pyridin-4-yl)thiophene-2-carboxamide, 4-(4-CHLOROPHENYL)-4-[4-(1H-PYRAZOL-4-YL)PHENYL]PIPERIDINE	DB07812, DB07859	
CDC6	Cell division control protein 6 homolog			
CENPE	Centromere-associated protein E	GSK-923295	DB06097	GSE35200
CXCR4	C-X-C chemokine receptor type 4	Framycetin, AMD-070, Plerixafor	DB00452, DB05501, DB06809	
FOXM1	Forkhead box protein M1			
CXCL12	Stromal cell-derived factor 1	Tinzaparin,	DB06822	
CDSN	Corneodesmosin			
DSG3	Desmoglein-3			
IVL	Involucrin			
DSP	Desmoplakin	Zinc, Zinc acetate	DB01593, DB14487	GSE27165, GSE2964

4.1 MICROARRAY ANALYSIS

Through disease association study microarray dataset for each approved drug target was identified using GEO Microarray database. 3 datasets were found that show association with carcinoma disease. These 3 microarray datasets were analysed by GEO2R tool available on GEO (Gene Expression Omnibus) website.

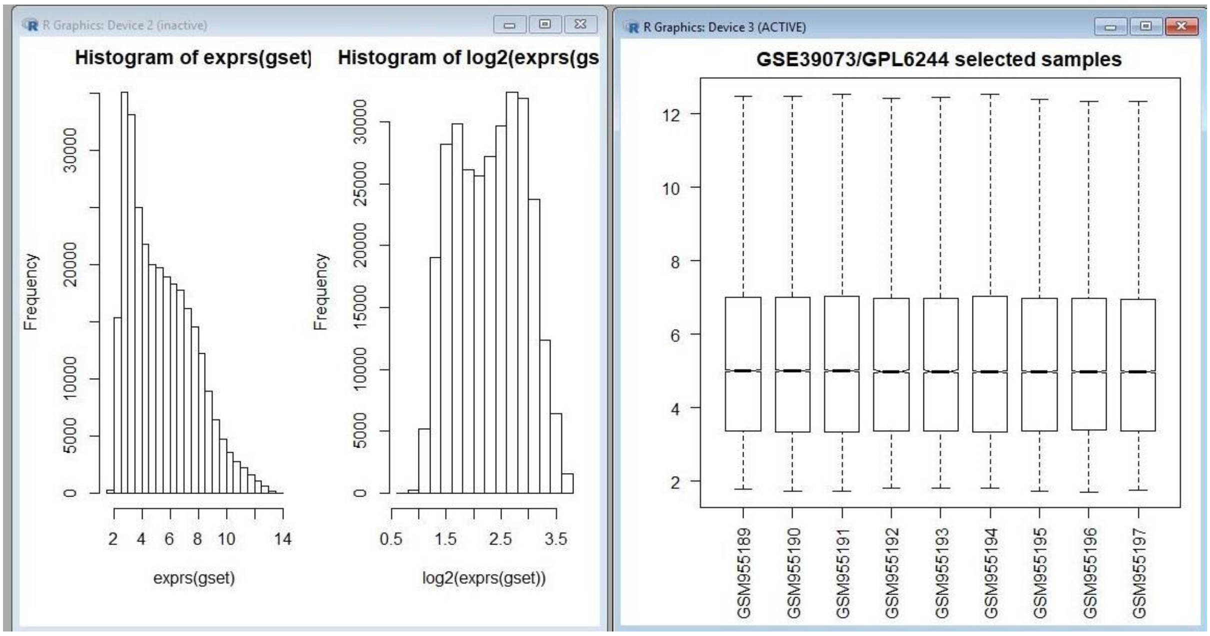

Figure 4.2: Representation of GSE39073 dataset normalisation and validation on R platform.

First datasets were studied and normalized (Figure 4.2) on R platform, then differential gene expressions were obtained. Gene expression of all 16 critical genes, were studied and identified on all 3 Microarray databases. There are 3 main approved drugs for MCC namely Vemurafenib, Avelumab, Pembrolizumab. Out of these 3 Vemurafenib is used for treatment of last stage of melanoma. Vemurafenib is the inhibitor of BRAF gene mutation that causes melanoma. So, the microarray date of Vemurafenib (GSE104869) is used to find out the gene expression before and after the drug usage. This data is further employed to find out the ‘fold change’ of 16 critical genes. Fold change showing the negative values in upregulated genes indicating that these genes are affected by the BRAF mutation inhibition which is responsible for Merkel Cell Carcinoma (MCC). Hence, this microarray analysis study validates our 16 critical genes that are responsible for Merkel Cell Carcinoma (MCC).

Table 4.2: A tabular representation showing fold change values in different Microarray dataset GSE22396, GSE50451 and GSE104869. Green and red colour represents upregulated and downregulated genes respectively.

Gene symbol	GSE22396 (Log FC)	GSE50451 (Log FC)	GSE104869 (Log FC)
AURKA	0.48525592	0.19211608	-0.69955388
KIF4A	0.49459202	0.07702051	-0.76814444
PKP1	-1.18842001	0.09340696	Not found
BUB1	0.65614758	0.09906461	-0.85706888
CDC6	0.3905455	0.06685074	-0.80892104
CENPE	0.62753201	0.0191033	-0.9252193
FOXM1	0.65953432	0.03433299	-0.70192919
DSG3	-1.86577326	0.11021616	Not found
DSP	-1.60072796	0.08411476	-0.68181818

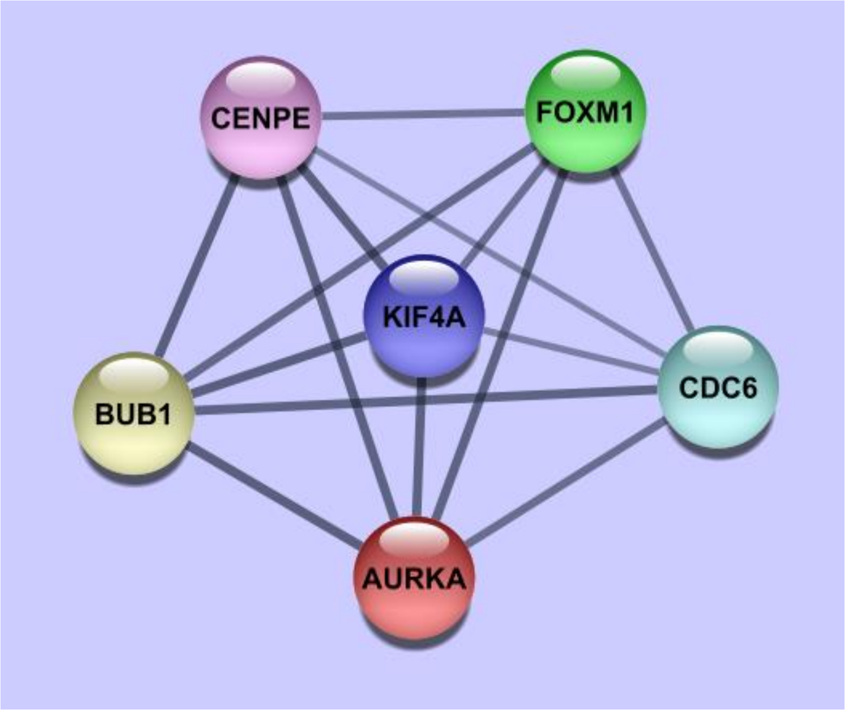

Figure 4.3: Interaction network between validated genes using Cytocape

4.2 RELEVANT CANCER STUDIES RELATED TO VALIDATED GENES IN HOMO SAPIENS:

- **Miriam N.C. et al** [6] studied the obesity as risk factor for breast cancer. In this research work they compared the protein-protein interaction (PPI) between obese and non-obese patients. In their study they used GSE78958 microarray dataset in which they reduce it to 269 gene expression data of primary breast tumors. Total of 16 potential candidates were found that might be responsible for amplification of breast cancer. PPI network for obese patients shows **BUB1** and CDK1 showing more interaction that means it is associated in cell proliferation and cell differentiation.

- **Bhairavi Tolani et al** [7] worked on Malignant pleural mesothelioma (MPM) tumors which is a very aggressive form of cancer because it does not depend on stage of the tumor development. MPM only depends on patient characteristics so it becomes necessary to know to potential genes responsible for it. For this they studied the molecular pathways, immunological response and progress at the genomic level of MPM cancer. In their case study they show **AURKA, KIF4A** and **FOXM1** genes are related to MPM cell cycle effectors. This means these genes are related to cell proliferation and crucial for MPM cancer development.

- **Yuping Han et al** [8] studied the molecular mechanism of bladder cancer to find the key genes in bladder cancer using gene expression profiling. For this they used GSE52519 microarray dataset that contains 3 normal bladder tissue and 9 bladder cancer samples. They find AURKA as the potential DEG in bladder cancer, also FOXM1 was upregulated in their study. BUB1 along with 4 other genes is one of the highest degree gene in their PPI network with 53 degree score. So, this shows the involvement of **AURKA, FOXM1** and **BUB1** in bladder cancer that means they might be involved in Cell proliferation, cell differentiation and cell cycle.

- **Alvaro Aytes et al** [9] worked on genome wide cross-species approach to find the potential targets for prostate cancer. In this study they used genetically engineered mouse model and join in with human prostate cancer data expression to find the common cancer associated genes that are preserved between both species. They found

FOXM1 and CENPF have synergistic interaction as the regulators that promote the growth of prostate cancer. They also experimentally validated these genes association with prostate cancer growth and also found to co-expression of these genes results in metastasis and poor survival rate.

Association of our critical genes in various forms of cancer like bladder cancer, breast cancer, MPM cancer and prostate cancer shows their involvement in cell proliferation, cell cycle, cell differentiation and melanoma. Hence, this microarray analysis and reviewed research study validated our critical genes AURKA, KIF4A, BUB1, CDC6, CENPE, FOXM1 for Merkel Cell Carcinoma.

5. DISCUSSION

Merkel Cell Carcinoma (MCC) is a destructive form of neuroendocrine carcinoma that represents the 2nd highest cause of skin cancer related deaths despite being a relatively rare form of cancer [1]. Different risk factors have been associated with MCC such as advancement age, immunosuppression, and ultraviolet light exposure, polyomavirus infection and a current, concurrent or previous diagnosis of Chronic Lymphocytic Leukaemia (CLL) [3]. However, the exact mechanism that leads to MCC and its inherent biology is yet to be fully researched and explored. The current treatments for MCC are a combination of surgery and radiation or chemo-radiation. So far, only two drugs have been approved by FDA for the treatment of MCC, with both drugs being immune checkpoint inhibitors. However, the drug approval from FDA was accelerated based on promising results from a relatively small number of patients, which means that the more research and trials are undergoing currently considering the fact that a very small cohort was used for clinical trials. A further issue is that many patients with MCC may not be ideal candidates for immunotherapy because their immune systems have been suppressed—due to organ/graft transplantation, lymphoma or HIV [5]. Presently there are no biomarkers indicative of MCC and additional research is essential. With the advent of bioinformatics and microarray technology, in silico approaches pertaining to 'omics' data analysis, shall allow identification of hub genes and miRNA to broaden our understanding of MCC biology and their clinical utilites.

For comparing the differeneces in expression of normal skin and MCC, one dataset was identified. Eventually, 164 upregulated and 285 downregulated genes were screened. Functional analysis revealed significant enrichment in positive regulation of cell proliferation, oxidation-reduction, epidermis development in the BP group. IN CC, cytoplasm, plasma membrane and extracellular exosome were highlighted. Structural molecule activity, Protein binding, homo-dimerization activity, calcium ion binding and structural constituent of cytoskeleton were the enriched MF terms. Additionally, pathways that were enriched included calcium signalling pathway, AAM, cAMP signalling pathway, p53 signalling pathway and GnRH signalling pathway.

CytoHubba plugin was used to identify hub nodes in the PPI network by observing an intersection of five algorithms (degree, MCC, MNC, EPC and EcCentricity). These include 16 genes: AURKA, FLG, KIF4A, DSG1, PKP1, LOR, BUB1, CDC6, CENPE, CXCR4, FOXM1, CXCL12, CDSN, DSG3, IVL and DSP. The 16 hubs genes included 7 upregulated and 9 downregulated genes and a network of these 16 hub genes revealed that the upregulated and downregulated genes were part of two individual clusters. An enrichment study showed that these genes were mostly enhanced in cornified envelope, cytosol, programmed cell death, cornification, cell cycle, immune system and chemokine signalling pathway. The involvement and implication of all these GO terms have already been discussed, barring cornification and chemokine signalling pathway. While a relationship between cornification and MCC hasn't been reported in literature, it is an important aspect of terminal differentiation and programmed cell death in epidermal keratinocytes. The identification of mechanisms controlling cornification may help in directing non-differentiating cancerous keratinocytes towards their eventual cell death. Similarly, a relation between chemokine signalling and MCC has yet not been reported however, chemokine receptor, $CXCR_3$ and its ligands reportedly play a role in skin cancers [23]. According to Human Protein Atlas (HPA), genes FLG, DSG1, LOR, CDSN and IVL show aberrant expression in skin cancers in comparison to normal skin samples. The rest of the genes identified in this study haven't been reported previously in relation to MCC studies.

Out of these 16 genes either downregulated or upregulated, 6 gene are further validated by microarray analysis. For validation three microarrays are used, two microarrays (GSE22396, GSE50451) are related to Merkel Cell Carcinoma (MCC) and one microarray (GSE104869) is for Vemurafenib drug which is FDA approved drug for Merkel Cell Carcinoma. After that several related literatures is also studied for validation of critical gene for carcinoma.

6. CONCLUSION

MCC is an aggressive form of neuroendocrine carcinoma that represents the second most common cause of skin cancer related deaths. MCC is often linked with a weak prognosis, as more than 33.33% of patients die from the disease compared to ~15% for malignant melanoma. Due to the rarity of the disease and of the occurrence of Merkle cells, MCC remains a cancer type that is yet to be fully researched and understood. This is of immediate consequence now as the incidences of MCC have been on a rise and the disease is often associated with fatality. In fact, only two FDA approved drugs exist in the market for treatment of MCC. Further, there is a lack of definitive biomarkers that could aid in a more effective selection of patients for new therapies, thereby creating a crucial need for further research The current study identified hub genes and hub miRNAs associated with MCC using integrated bioinformatics along with their functional analysis, PPI network construction, module selection and module enrichment analyses. Out of these 16 genes either downregulated or upregulated, 6 gene are further validated by microarray analysis. For validation three microarrays are used, two microarrays are related to Merkel Cell Carcinoma (MCC) and one microarray is for Vemurafenib drug which is FDA approved drug for Merkel Cell Carcinoma. After that several related literatures is also studied for validation of critical gene for carcinoma. In conclusion, the findings of the current work shows that gene found by these tools are responsible for MCC and provide prerequisite for future studies to identify useful markers and understand underlying mechanisms related to MCC carcinogenesis.

REFERENCES

[1]. Schadendorf D, Lebbe C, Zur Hausen A, et al. (2017). Merkel cell carcinoma: epidemiology, prognosis, therapy and unmet medical needs. Eur J Cancer, 71, 53–69.

[2]. Tang CK, Toker C. (1978). Trabecular carcinoma of the skin: an ultrastructural study. Cancer, 42(5), 2311–21.

[3]. Kaae J, Hansen AV, Biggar RJ, Boyd HA, Moore PS, Wohlfahrt J, et al. (2010). Merkel cell carcinoma: incidence, mortality, and risk of other cancers. J Natl Cancer Inst, 102 (11), 793-801.

[4]. Kaufman HL, Russell J, Hamid O, Bhatia S, Terheyden P, D'Angelo SP, et al. (2016). Avelumab in patients with chemotherapy refractory metastatic Merkel cell carcinoma: A multicentre, single-group, open-label, phase 2 trial. Lancet Oncol, 17(10), 1374–85.

[5]. Nghiem PT, Bhatia S, Lipson EJ, Kudchadkar RR, Miller NJ, Annamalai L, et al. (2016). PD-1 blockade with Pembrolizumab in advanced Merkel-cell carcinoma. N Engl J Med, 374(26), 2542–52.

[6]. Nuncia-Cantarero, Miriam et al. “Functional transcriptomic annotation and protein-protein interaction network analysis identify NEK2, BIRC5, and TOP2A as potential targets in obese patients with luminal A breast cancer.” *Breast cancer research and treatment* vol. 168,3 (2018): 613-623. doi:10.1007/s10549-017-4652-3

[7]. Tolani, Bhairavi et al. “Heterogeneous Contributing Factors in MPM Disease Development and Progression: Biological Advances and Clinical Implications.” *International journal of molecular sciences* vol. 19,1 238. 13 Jan. 2018, doi:10.3390/ijms19010238

[8]. Han, Yuping et al. “Identification of key genes associated with bladder cancer using gene expression profiles.” *Oncology letters* vol. 15,1 (2018): 297-303. doi:10.3892/ol.2017.7310

[9]. Aytes, Alvaro et al. “Cross-species regulatory network analysis identifies a synergistic interaction between FOXM1 and CENPF that drives prostate cancer malignancy.” *Cancer cell* vol. 25,5 (2014): 638-651. doi:10.1016/j.ccr.2014.03.017.

[10]. Huang X, Yuan T, Liang M, Du M, Xia S, Dittmar R, Wang D, See W, Costello BA, Quevedo F, Tan W, Nandy D, Bevan GH, Longenbach S, Sun Z, Lu Y,Wang T, Thibodeau SN, Boardman L, Kohli M, Wang L. (2014). Exosomal miR-1290 and miR-375 as prognostic markers in castration-resistant prostate cancer. Eur Urol, 67(1), 3-41.

[11]. Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, Simonovic M, Doncheva NT, Morris JH, Bork P, Jensen LJ, Mering CV. (2019). STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Res., 47(D1):D607-D613.

[12]. Bader GD, Hogue CW. (2003). An automated method for finding molecular complexes in large protein interaction networks. BMC Bioinformatics, 13; 4:2.

[13]. Asioli S, Righi A, Volante M, et al. (2007). p63 expression as a new prognostic marker in Merkel cell carcinoma. Cancer, 110(3), 640–7.

[14]. Becker JC. Merkel cell carcinoma. (2010). Ann Oncol, 21(7), 81–85.

[15]. Burack J, Altschuler EL. (2003). Sustained remission of metastatic Merkel cell carcinoma with treatment of HIV infection. J R Soc Med. 9(6), 238–239. doi: 10.1258/jrsm.96.5.238.

[16]. Clapham, D. E. (2007). Calcium signaling. Cell, 131, 1047–1058.

[17]. Demetriou S K, Ona-Vu K, Sullivan EM, Dong TK, Hsu SW, & Oh DH. (2012). Defective DNA repair and cell cycle arrest in cells expressing Merkel cell polyomavirus T antigen. International journal of cancer, 131(8), 1818–1827.

[18]. Gorrini, C., Harris, IS. & Mak, TW. (2013). Modulation of oxidative stress as an anticancer strategy. Nat. Rev. Drug Discov. 12(12), 931–947.

[19]. Hafner C, Houben R, Baeurle A, Ritter C, Schrama D, Landthaler M, et al. (2012) Activation of the PI3K/AKT Pathway in Merkel Cell Carcinoma. PLoS ONE 7(2).

[20]. Hughes-Fulford M, Li CF, Boonyaratanakornkit J. & Sayyah S. (2006). Arachidonic acid activates phosphatidylinositol 3-kinase signaling and induces gene expression in prostate cancer. Cancer Res, 66(3), 1427–1433.

[21]. Hyde CA & Missailidis S. (2009). Inhibition of arachidonic acid metabolism and its implication on cell proliferation and tumour-angiogenesis. Int. Immunopharmacol, 9(6), 701–715.

[22]. Imamura T, Komatsu S, Ichikawa D, Miyamae M, Okajima W, Ohashi T, Kiuchi J, Nishibeppu K, Kosuga T, Konishi H, Shiozaki A, Okamoto K, Fujiwara H, Otsuji E.(2017). Low plasma levels of miR-101 are associated with tumor progression in gastric cancer. Oncotarget, 8(63), 106538-106550.

[23]. Kuo PT, Zeng Z, Salim N, Mattarollo S, Wells JW, Leggatt GR. (2018). The Role of CXCR3 and Its Chemokine Ligands in Skin Disease and Cancer. Front Med (Lausanne). doi: 10.3389/fmed.2018.00271.

[24]. Mantel A, Carpenter-Mendini A, VanBuskirk J, Pentland AP. (2014). Aldo-keto reductase 1C3 is overexpressed in skin squamous cell carcinoma (SCC) and affects SCC growth via prostaglandin metabolism. Exp Dermatol, 23(8), 573-8.

[25]. Mariot P, Vanoverberghe K, Lalevee N, Rossier MF & Prevarskaya N. (2002). Overexpression of an alpha 1H (CaV3.2) T-type calcium channel during neuroendocrine differentiation of human prostate cancer cells. J. Biol. Chem, 277(13), 10824–10833.

[26]. Marks F, Fürstenberger G & Müller-Decker K. (2007). Tumor promotion as a target of cancer prevention. Recent Results Cancer Res. 174, 37–47.

[27]. Monteith G R, Davis F M & Roberts-Thomson S J. (2012). Calcium channels and pumps in cancer: changes and consequences. J. Biol. Chem. 287(38), 31666–31673.

[28]. Panelos J, Batistatou A, Paglierani M, Zioga A, Maio V, Santi R, Pimpinelli N, De Giorgi V, Santucci M, Massi D. (2009). Expression of Notch-1 and alteration of the E-cadherin/beta-catenin cell adhesion complex are observed in primary cutaneous neuroendocrine carcinoma (Merkel cell carcinoma). Mod Pathol, 22(7), 959-68

[29]. Pfeffer SR, Grossmann KF, Cassidy PB, Yang CH, Fan M, Kopelovich L, Leachman SA, Pfeffer LM.(2015). Detection of Exosomal miRNAs in the Plasma of Melanoma Patients. J Clin Med, 4(12), 2012-27.

[30]. Scuderi MR, Anfuso CD, Lupo G, Motta C, Romeo L, Guerra L, Cappellani A,Ragusa N, Cantarella G, Alberghina M. Expression of Ca(2+)-independent and Ca(2+)-dependent phospholipases A(2) and cyclooxygenases in human melanocytes and malignant melanoma cell lines. (2008). Biochim Biophys Acta, 1781(10), 635-42.

[31]. Toker C. (1972). Trabecular carcinoma of the skin. Arch Dermatol, 105(1), 107–10.

[32]. Wang D Z & Dubois RN. (2010). Eicosanoids and cancer. Nat. Rev. Cancer, 10(3), 181–193.

[33]. Wheat R, Roberts C, Waterboer T, Steele J, Marsden J, Steven N M & Blackbourn D J. (2014). Inflammatory cell distribution in primary merkel cell carcinoma. Cancers, 6(2), 1047–1064. doi:10.3390/cancers6021047.

[34]. Wooff JC, Trites JR, Walsh NM, Bullock MJ. (2010). Complete spontaneous regression of metastatic merkel cell carcinoma: A case report and review of the literature. Am J Dermatopathol, 32(3), 614–617.

[35]. Xie H, Lee L, Caramuta S, Hoog A, Browaldh N, Bjornhagen V, Larsson C, Lui WO. (2014). MicroRNA expression patterns related to merkel cell polyomavirus infection in human merkel cell carcinoma. J. Investig. Dermatol, 134(2), 507–517. doi: 10.1038/jid.2013.355.

[36]. Zhao Q, Deng S, Wang G, Liu C, Meng L, Qiao S, Shen L, Zhang Y, Lü J, Li W,Zhang Y, Wang M, Pestell RG, Liang C, Yu Z.(2016). A direct quantification method for measuring plasma MicroRNAs identified potential biomarkers for detecting metastatic breast cancer. Oncotarget, 7(16), 21865-74. doi:10.18632/oncotarget.7990.

[37]. Zhu H. et al. (2014). Elevated Orai1 expression mediates tumor-promoting intracellular Ca2+ oscillations in human esophageal squamous cell carcinoma. Oncotarget, 5(11), 3455–3471.

LIST OF PUBLICATIONS

- **Himanshu Kumar** and Yasha Hasija. **Machine Learning In Medical Image Processing.** Information and Communication Technology for Intelligent Systems (ICTIS 2020), May 15-16th, 2020 (Springer) (Accepted).

- **Himanshu Kumar** and Yasha Hasija. **Bioinformatics Analysis of Merkel Cell Carcinoma to identify critical genes and their validation.** International Conference for Innovation in Technology (INOCON 2020), November 6-8th, 2020 (IEEE) (Accepted).