UNIVERSITY OF CALIFORNIA SAN DIEGO

A Pan-Cancer Analysis of Immune-Associated Genes and Pathways Dysregulated by Tobacco Smoke

A thesis submitted in partial satisfaction of the requirements

for the degree Master of Science

in

Biology

by

Jaideep Chakladar

Committee in charge:

Professor Weg M. Ongkeko, Chair

Professor Li-Fan Lu, Co-Chair

Professor Alistair Russell

2022

The thesis of Jaideep Chakladar is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego

2022

TABLE OF CONTENTS

Thesis Approval Page……iii

Table of Contents……...…….iv

List of Abbreviations…….……..v

List of Figures……...…….....vi

List of Tables……vii

Acknowledgements……..viii

Abstract of the Thesis……..……...ix

Chapter 1: Introduction……..……..1

Chapter 2: Materials and Methods……...4

Chapter 3: Immune-Associated Gene Expression……...……...……9

3.1: Dysregulated and survival-correlated IA genes…….……...9

3.2: Contrasting effects TNF-related gene expression in LUSC and HNSCC…….….14

Chapter 4: Osteopontin and Smoking……....20

4.1: Osteopontin expression in HNSCC smoking samples…….……...20

4.2: Pan-cancer Osteopontin expression…….……...23

Chapter 5: miRNA expression……...25

Chapter 6: Discussion……27

6.1: Interpretation of results……...27

6.2: Setbacks and potential solutions……29

6.3: Future directions……30

References……..32

LIST OF ABBREVIATIONS

IA: Immune-associated

TCGA: The Cancer Genome Atlas

HNSCC: Head and Neck Squamous Cell Carcinoma

LUSC: Lung Squamous Cell Carcinoma

LUAD: Lung Adenocarcinoma

ESCA: Esophageal Carcinoma

BLCA: Urothelial Bladder Carcinoma

OSCC: Oral Squamous Cell Carcinoma

CNV: Copy Number Variation

MUT: Mutation

OPN: Osteopontin

REVEALER: Repeated Evaluation of VariablEs conditionAL Entropy and Redundancy

GSEA: GeneSet Enrichment Analysis

FDR: False Discovery Rate

miRNA: microRNA

HPV: Human Papillomavirus

CIC: Conditional Information Coefficient

TNF: Tumor Necrosis Factor alpha

HIF-1: Hypoxia-Inducible Factor-1

LIST OF FIGURES

Figure 1: Overview of the proposed analysis pipeline and patient population……6

Figure 2: Differentially expressed IA genes in 5 smoking-mediated cancers……...11

Figure 3: Survival correlations to IA gene expression……...15

Figure 4: Conflicting effects of TNF-related expression in HNSCC and LUSC smokers……19

Figure 5: OPN expression in HNSCC……...…22

Figure 6: Correlation of cancer and immune pathways to OPN expression in HNSCC……...…24

Figure 7: Pan-cancer OPN expression trends……26

Figure 8: Pan-cancer miRNA expression patterns……..……..…….28

LIST OF TABLES

Table 1: Dysregulated IA genes between smoking and nonsmoking cancer patients……...13

Table 2: Survival correlated IA genes in smoking cancer patients……..…..17

ACKNOWLEDGEMENTS

I would like to acknowledge Professor Weg M. Ongkeko for his support as the chair of my committee and for his invaluable mentorship and support throughout my time at UCSD.

I would also like to acknowledge current and past members of the Ongkeko lab for supporting me throughout my journey and keeping me motivated through tough times.

This thesis in full will be submitted for publication. The thesis author will be the primary investigator author of this paper, and Ongkeko, Weg M. will be the principal investigator.

ABSTRACT OF THE THESIS

A Pan-Cancer Analysis of Immune-Associated Genes

and Pathways Dysregulated by Tobacco Smoke

by

Jaideep Chakladar

Master of Science in Biology

University of California San Diego, 2022

Professor Weg M. Ongkeko, Chair
Professor Li-Fan Lu, Co-Chair

Although many of the mechanisms induced by cigarette smoke are highly conserved in cancers in general, the use of new genomic and transcriptomic data analysis tools suggests that some mechanisms underlying smoking induced cancers may be unique. This project aims to elucidate connections between smoking-associated cancers and novel immune-associated (IA) mechanisms underlying smoking-induced carcinogenesis, with an emphasis on IA gene

expression and microRNA (miRNA) activity. We investigated a total of 5 cancers whose incidence is known to be well-correlated with smoking. Patient whole genome sequencing, miRNA sequencing, and clinical variable data was downloaded from The Cancer Genome Atlas (TCGA) and was analyzed computationally. We found little overlap between survival-corelated immune-associated genes dysregulated in each of the 5 cancers studied. However, further downstream analysis suggested the potential importance of a select few genes. One of these genes is osteopontin (OPN), which was upregulated in HNSCC and ESCA patients alongside key oncogene upregulation, tumor suppressor downregulation, and mutation presence. Dysregulation of TNF-related genes was unique between HNSCC and LUSC samples, suggesting that smoking causes different behavior of TNF depending on the cancer type or tumor site. Analysis of microRNA expression indicated that survival-correlated IA genes were likely unaffected by miRNA expression. These findings indicate the presence of common and unique patterns of IA gene dysregulation between smoking-mediated cancers that may be used for future therapeutic strategies.

CHAPTER 1: INTRODUCTION

Smoking and cancer

Currently, smoking is one of the most prominent risk factors for cancer, with tobacco smoking being a primary cause of 12 different cancers. Cigarette smoke contains thousands of components in its gaseous phase, of which numerous are direct and indirect carcinogens[1]. Carcinogens in cigarette smoke have been shown to initiate various oncogenic mechanisms that eventually result in advanced stage tumors[2]. One such mechanism is the action of nitrosamines that causes gene mutations and DNA and protein adducts that lead to dysfunction of pivotal cell cycle regulator mechanisms[3]. Nicotine has also been shown to promote carcinogenesis by activating cell growth and migration pathways, along with cancer-specific functions such as angiogenesis and tissue invasion[4]. Often, increased smoking duration and frequency correlates to increased risk of cancer because of the additive effects of cancer-promoting mechanisms. Formation of adducts by carcinogens may be normally repaired by an otherwise healthy cell. Cell intrinsic defense mechanisms also prevent adducts from taking effect by utilizing apoptosis as a last line of defense[5]. However, with greater exposure to cigarette smoke comes a greater likelihood of mutation of genes governing normal cell function, thereby rendering defenses against mutated proteins useless.

Recent findings have implicated areas outside of simple genomics with smoking-induced carcinogenesis. For example, the ability of smoking carcinogens to alter the methylation state of DNA has been shown to contribute to promote carcinogenesis, via hypermethylation of promotors corresponding to tumor-suppressor genes[6].

Smoking and the immune system

The investigation of the relationship between smoking and cancer has indicated the importance of the immune system. Interestingly, the effects that smoking can have on the immune system can be contradictory, as smoking has been shown to cause both pro-inflammatory and immunosuppressive outcomes[7,8]. However, studies into the main targets of cigarette smoke in the immune system are not all in agreement, likely due to the variance amongst patient populations in terms of smoking duration and volume, demographics, and other health conditions like nutrition and obesity[9]. One way in which cigarette smoke affects the immune system is through compromisation of the respiratory epithelium[10-12]. This compromisation leads to a host of downstream effects, including oxidative damage that leads to single-strand DNA breaks, somatic mutations in epithelial cells, and overall hyperinflammation at the epithelium[13-15]. Such an environment can lead to tumor formation at the respiratory epithelium. Indeed, it has been suggested that activating KRAS mutations are common in lung tumors and may be key contributors to a pro-inflammatory tumor microenvironment[16].

Smoking may also inhibit the ability of the immune system to recognize and kill tumor cells. *In vivo* studies suggest that mice exposed to cigarette smoke have an inhibited immune response to transplanted tumors, and such tumors frequently metastasize[17,18]. One explanation for these findings may be that cytotoxic T cell activity is decreased, as observed in a separate study[19].

Outside of lung cancers, the effects of an altered immune system caused by smoking have not yet been well established in smoking-mediated cancers. Head and neck squamous cell carcinoma (HNSCC) patients have been shown to have a worse survival if they have a history of smoking[20,21]. HNSCC-specific research seems to suggest that the effects of smoking on the

immune system are not conflicting as it may be in other cancers, as the HNSCC tumor microenvironment is immunosuppressive[22-24]. Therefore, it is likely that the contribution of smoking to HNSCC must also be immunosuppressive. Indeed, a recent study indicated that HNSCC smokers exhibit a decrease in cytotoxic T cell activity and a significant downregulation of interferon and chemokine-related genes[25]. However, mutation burden was not associated with smoking status[25].

The effects of a smoking-altered immune system in bladder urothelial (BLCA) and esophageal squamous cell cancer (ESCA) remain unclear. One study suggests that a sub-clinical immune suppression caused by cigarette smoke can lead to bladder cancer[26]. However, there are few studies that investigate the connection between the immune system and smoking in bladder cancer, despite the fact that over 50% of BLCA cases are caused by smoking[27]. A recent study of ESCA indicated a unique immune phenotype in ESCA smokers compared to ESCA nonsmokers[28]. However, the sample size of less than 100 ESCA smokers and the lack of validation of these findings leaves the connection between ESCA, the immune system, and smoking unclear.

Applications of unique immune genotypes

Although many of the mechanisms induced by cigarette smoke are highly conserved in cancers in general, the advent of new genomic, transcriptomic, and metabolomic data analysis tools suggest that some mechanisms underlying smoking induced cancers may be unique. However, there has been little progress made into identifying such unique mechanisms. With survival rates for some of these cancers stagnating over the past decade, the development of new treatment and diagnostic modalities has been a large focus of the oncology research community. Immune therapy is an alternative for patients who do not respond well to traditional

chemotherapies. However, few immune therapies have been brought to market, and many do not pass clinical trials. A problem with immune therapy development is a lack of known targets in the immune system that can be exploited to target tumors. It is therefore important for targets in the immune system to be identified, especially amongst patient populations that may harbor genotypic alterations that result in lower efficacy of traditional chemotherapy.

This project aims to elucidate any possible connections between smoking-associated cancers and novel mechanisms underlying smoking-induced carcinogenesis. This is accomplished by analyzing patient sequencing data from The Cancer Genome Atlas (TCGA) and identifying genomic and transcriptomic characteristics that may underlie smoking-induced carcinogenesis.

CHAPTER 2: MATERIALS AND METHODS

TCGA and non-TCGA RNA-sequencing datasets and clinical data

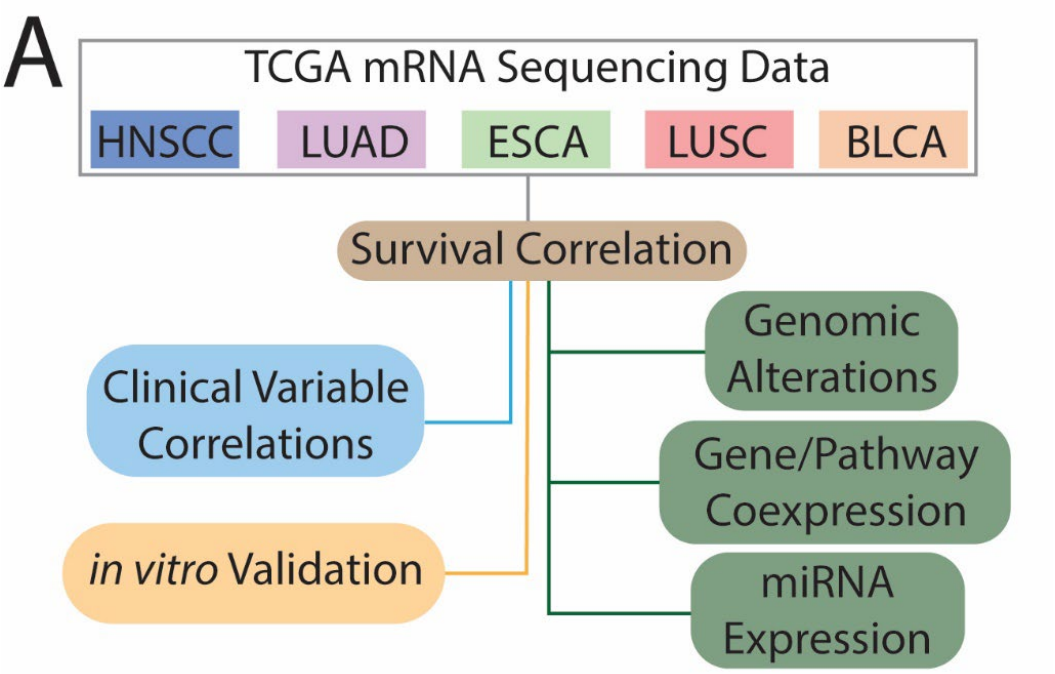
Level 3 normalized mRNA and miRNA expression read counts for tumor samples from 501 LUSC patients, 514 LUAD patients, 500 HNSCC patients, 408 BLCA patients, and 158 ESCA patients as well as the patients' clinical data were downloaded from TCGA (https://tcga-data.nci.nih.gov/tcga) (Figure 1). mRNA and miRNA read counts for adjacent solid normal tissue samples of 49 LUSC patients, 59 LUAD patients, 44 HNSCC patients, 19 BLCA patients, and 65 ESCA patients were also obtained. Additional validation datasets were obtained from the Sequencing Read Archive (SRA). Tumor samples from 167 OSCC, 17 oral dysplasia, and 45 normal oral epithelial mouse models were downloaded from the SRA along with matching tumor stage data (GSE30784).

Clinical information for all patients were downloaded from the Broad GDAC Firehose (https://gdac.broadinstitute.org/). Genomic alteration information for each patient was obtained from the most up-to-date analysis report (2016) of the Broad Institute TCGA Genome Data Analysis Center (http://gdac.broadinstitute.org/runs/analyses__latest/reports/).

Differential expression analysis for identification of dysregulated IA genes

mRNA read count tables were imported into edgeR v3.5 (http://www.bioconductor.org/packages/release/bioc/html/edgeR.html), and lowly expressed mRNAs (CPM < 1 in number of samples greater than the size of the smaller cohort of each analysis) were filtered from the analysis. Following trimmed mean of M-values (TMM) normalization, pairwise designs were applied to identify significantly differentially expressed mRNAs for each cancer. The initial comparisons are as follows: (1) cancer smoking vs. cancer nonsmoking, (2) cancer smoking vs. normal, and (3) cancer nonsmoking vs. normal.

IA genes differentially expressed were identified as dysregulated and retained as candidates. A gene is determined to be immune associated if it is involved in adaptive or innate immunological pathways or constitute tumor antigens. We sourced our list of IA genes from ImmPort (http://www.immport.org/immport-open/public/reference/genelists), listing all adaptive and innate immunity associated genes; the InnateDB (http://www.innatedb.com/), listing innate immunity associated genes; and the TANTIGEN database (http://cvc.dfci.harvard.edu/tadb/), listing genes that can form tumor antigens. Differential expression is defined as $p<0.05$ and fold change < -2 or > 2 in edgeR analysis. The *p*-value was corrected by using the FDR provided by edgeR.



B

Cancer	Patient Population	
	Cancer Smokers	Cancer Non-smokers
HNSCC	170	111
LUAD	123	74
ESCA	33	47
LUSC	133	18
BLCA	93	109

Figure 1: Overview of the proposed analysis pipeline and patient population. **A**. A flowchart of the general analyses and goals of the proposed project. **B**. Total number of smoking and nonsmoking cancer patients for each of the 5 TCGA cancers studied.

Association of gene expression with patient survival and clinical variables

Survival analyses were performed using the Kaplan-Meier Model with gene expression designated as a binary variable based on expression above or below the median expression of all samples. Univariate Cox regression analysis was used to identify candidates significantly associated with patient survival ($p < 0.05$). Survival-correlated genes were evaluated for clinical significance. The Kruskal-Wallis test ($p<0.05$) was used to correlate gene expression to clinical variables, including pathologic stage, pathologic TNM stages, perineural invasion, and neoplasm status after treatment. In pathologic T stage analysis, patients with stages T1a and T1b were grouped into stage T1, and likewise for stages T2, T3, and T4.

Functional pathway clustering of candidate IA genes and functional coexpression analysis

A list of all significantly dysregulated IA genes was inputted into the ReactomeFIViz plugin of Cytoscape to cluster genes into functional pathways by cancer. For each cancer, a matrix listing each gene's expression value for every patient was inputted for functional coexpression analysis. Only coexpression modules with 4 genes or more and FDR<0.05 were presented. Functional pathways available for analysis were sourced from Reactome, Kegg, BioCarta, Pathway Interactions Database, and PantherDB. The top pathways representing genes in each cluster were determined according to FDR, number of genes dysregulated, and relevance to the immune system.

Association of candidate genes' expressions with genomic alterations

Mutation and CNV data for the TCGA tumor samples were obtained from mutation and CNV annotation files generated by the Broad Institute GDAC Firehose. All mutation and CNV annotations were compiled into a binary input file for the information coefficient-based

algorithm REVEALER. REVEALER is designed to computationally identify a set of specific CNVs and mutations most likely responsible for the change in activity of a target profile. The target profile was defined in our study to be individual IA gene expression. In order to identify a set of most relevant genomic alterations, REVEALER runs multiple iterations of the correlation algorithm, with the genomic feature exhibiting the strongest correlation in each run serving as a seed for the successive run. We set the maximum number of iterations to three. A seed is a particular mutation or copy number gain or loss event that most likely accounts for the target activity. When given a seed, REVEALER will focus correlation on only patients with altered target activity not accounted for by the seed. Since we do not know which genomic alteration is responsible for the dysregulation of each gene, we set the seed for the first iteration to null. We set the threshold of genomic features to input to features present in less than 75% of all samples. Significant association was defined by the conditional information coefficient (CIC > 0.25).

GeneSet Enrichment Analysis (GSEA)

GSEA was used to correlate IA gene expression to the dysregulation of highly conserved signatures and pathways. Genesets were chosen from the CP, C2, and C7 sets from the Molecular Signatures Database[29]. Significantly enriched signatures were identified by a nominal enrichment score > 1 and a nominal p-value < 0.05.

Gene co-expression analysis

Spearman scatter plots were used to plot gene expression read count data of individual genes against each other. Significant gene co-expression was determined using the Spearman correlational coefficient ($|\rho| > 0.3$). Further analysis of genes that could be co-expressed with IA genes was determined by referring to the Pathway Commons[30]. Lists of genes that participate in

common cancer and immune pathways were obtained and were referenced for specific genes of interest.

miRNA expression analysis

miRNAs of interest were first determined by inputting survival-correlated IA genes into the TargetScanHuman website (https://www.targetscan.org/vert_71/), which predicts highly and poorly conserved miRNAs that target genes. Highly conserved miRNAs for each gene were chosen, and differential expression analysis was used to determine whether or not these miRNAs were differentially expressed between smoking and nonsmoking cancer patients. Further study of miRNA expression patterns was achieved using Principal Component Analysis. This analysis was performed using the prcomp library in R, and the results of the analysis were visualized using the ggplot2 and ggfortify2 libraries.

CHAPTER 3: IMMUNE-ASSOCIATED GENE EXPRESSION

Dysregulated and survival correlated IA genes

Patient gene expression data was downloaded from TCGA and computationally analyzed to determine differentially expressed genes between smokers and nonsmokers. A total of 5 cancers were selected based on lowest survival rates and availability of patient data (>100 patients) (Figure 1A-B). Differential expression was done on 3 different cohorts, comparing them in a pairwise manner. The comparisons were as follows: (1) cancer smoking vs. cancer nonsmoking, (2) cancer smoking vs. normal, and (3) cancer nonsmoking vs. normal. These specific comparisons were chosen to enable an analysis of how cancer patients that smoked differed from other cancer patients (comparison 1) as well as from normal samples (comparison 2 and 3). It was proposed that the latter analysis would be accomplished by analyzing the

common and uniquely dysregulated genes between comparisons 2 and 3 in order to establish how the cancer smoking phenotype may cause differences from the normal phenotype that are not present in the cancer nonsmoking phenotype. However, upon analysis of the differential expression results of all sequenced genes, it was discovered that the landscapes of differentially expressed genes from comparisons 2 and 3 were largely similar (Figure 2A). In hindsight, this may have been expected, given that gene expression between cancer samples is much more similar than gene expression between cancer and normal samples. Additionally, comparisons 2 and 3 did not indicate many novel dysregulations that were a significant departure from what is already known about how smoking causes cancer. Therefore, the results that differential expression analysis of comparison 1 yielded became the primary focus of this thesis.

The results of the differential expression analysis revealed a finite number of genes that were dysregulated between smoking and nonsmoking cancer patients (Figure 2A-B). The majority of significantly dysregulated genes had an absolute fold change of less than 5. The significantly dysregulated genes were filtered to select for genes that are immune associated (Figure 2C). HNSCC, BLCA, and LUAD had the most dysregulated IA genes, with HNSCC having more than double the number of dysregulated IA genes than any other cancer. The cancers that had the most dysregulated IA genes in common were HNSCC, BLCA, and LUAD. However, the direction of dysregulation of these overlapping genes was not consistent across cancers, indicating that the immune system may have been impacted in contradicting ways between cancers (Table 1).

Interestingly, the IA gene dysregulation profiles of LUAD and LUSC did not have many overlaps, and LUSC had much fewer dysregulated IA genes. Both of these findings were contrary to what was expected. It was hypothesized that LUSC would have a significant number

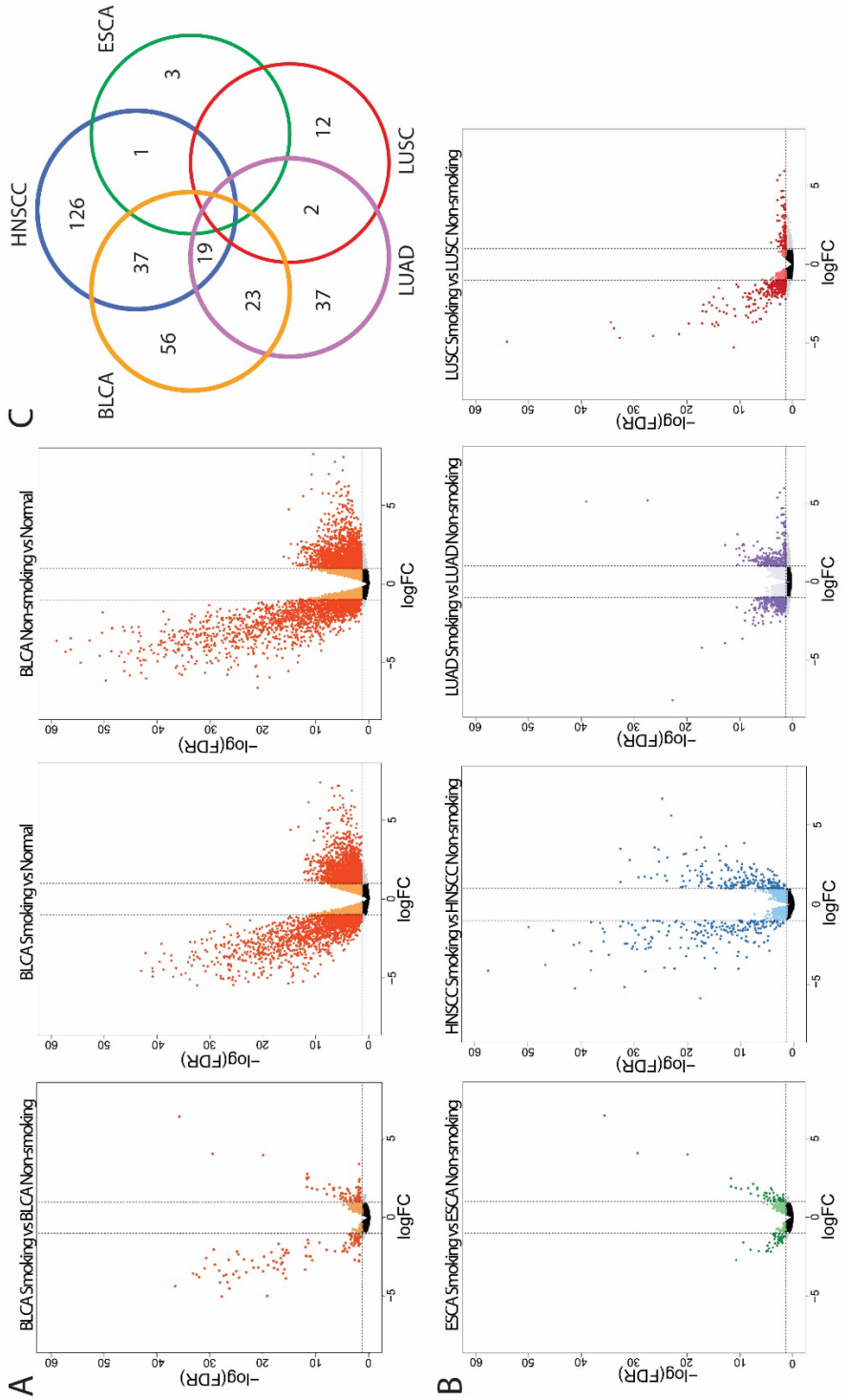


Figure 2: Differentially expressed IA genes in 5 smoking-mediated cancers. **A**. Representative volcano plots depicting the differential expression parameters for three comparisons in BLCA. Significance thresholds (|logFC|>1 and FDR<0.05) are indicated by dotted lines. Each point is one gene, and plots are representative of all genes, including IA genes. **B**. Volcano plots for differential expression comparison 1 in the remaining 4 cancers analyzed. **C**. Common and uniquely dysregulated IA genes between all 5 cancers analyzed.

of IA genes dysregulated by smoking due to the well-established connection between smoking and lung epithelial damage and hyperinflammation. Additionally, mechanisms inhibited by smoking such as immune cell activity would most likely affect the epithelial surface of the lungs rather than a deeper location. It is therefore interesting that LUSC did not have many significantly dysregulated IA genes. The differences in LUAD and LUSC were unexpected due to their shared location in the body. Though LUAD and LUSC are known to have significantly different clinical outcomes and molecular features, it was hypothesized that, because cigarette smoke exerts its effects primarily through lung epithelial barrier damage, there would be a significant amount of shared IA gene expression patterns between LUAD and LUSC. A possible explanation for the difference observed is that LUAD prognosis is heavily influenced by smoking. Specifically, even though LUAD is the most common lung cancer amongst nonsmokers, smoking greatly increases the risk of LUAD development and metastasis[31,32]. In contrast, LUSC most commonly manifests in smokers, as indicated by previous studies and by the fact that TCGA contains only 18 nonsmoking LUSC patients[33]. It is possible that, regardless of smoking status, the immune-associated genotype of LUSC cases is largely the same, while smoking greatly shifts the genotype in LUAD, which leads to the increased prevalence and metastasis. Additionally, the low number of LUSC nonsmoking samples could explain these findings.

We next used the Kaplan-Meier test to determine which of the significantly dysregulated IA genes were correlated with patient survival. Only survival of patients that smoked was used in this analysis. HNSCC had the greatest number of survival-correlated genes, and the effects of these genes on patients survival were often conflicting (Figure 3A). However, there were few, if any, survival-correlated IA genes in the other 4 cancers. These findings may be due to the

Table 1: Dysregulated IA genes between smoking and nonsmoking cancer patients

Cancers							
BLCA	ESCA	HNSCC				LUAD	LUSC
AMH	ACVR1C	ACTR3	DICER1-AS1	LDLRAD3	RAG1	ART4	APOBEC3A
ANGPTL2	CFI	ADAM17	DKK1	LRRC41	RASGEF1B	C8B	CA9
AVPR1A	ESRRG	ADAM33	DUSP3	LTBP2	RBP5	CD300LB	CEACAM5
BTC	MMP12	ADCY10P1	EFTUD2	LY96	RPS6KA5	CECR1	EPGN
CCL19	SEMA3B	ADM	EIF2AK2	MAP2K1	RSAD2	CTSG	GNAI2
CCR4	OPN	AIMP1	EPHA3	MAPK9	RUSC1-AS1	DEFB1	IL1RL2
CD19	WFDC2	ALKBH2	EPOP	MAPKAPK2	S1PR2	EREG	IL22RA1
CD200		ALKBH6	ESR1	MEF2D	SCARB1	F11	IL23A
CD28		ANGPTL7	ETV6	MIF	SELENBP1	F2RL1	PGLYRP3
CD48		AREG	F2RL2	MUM1	SELENOT	FABP6	PGLYRP4
CD79B		ARF1	FABP4	NAMPT	SEMA3G	FCN3	PRKD1
CDK1		ARHGEF18	FASTKD2	NCKAP1	SEMA4C	GDF10	RNASE7
CFP		ARID3B	FGF19	NDRG1	SEMA5A	HNF4G	SLAMF9
CLEC10A		ARL14EP	FLT4	NINJ2	SERPINE1	IGHV3-64	SLC40A1
CX3CR1		ARL2BP	FYN	NPPC	SH3BP2	IGHV3-72	USP2
CXCR4		ARL4D	GALNS	NR2C2	SHC3	KLB	XCL1
ESM1		ARL6IP4	GALNT3	NR2F1	SIGIRR	LGR5	
F2RL2		ARRDC2	GREM1	NR3C2	SIRT1	LY6K	
FCGR3A		ATG16L1	GREM2	OAS2	SLX4	MMP14	
FCN1		AVPR2	GSTP1	OAS3	STC2	MSR1	
GRAP2		BECN1	HBEGF	OASL	STRAP	MYL9	
GZMM		BMP6	HFE2	OSGIN1	SYNM	NR2F1	
IL6R		BST2	HIF1A	OTUD5	SYNPO2	OAS3	
LRRK2		C1QBP	HLA-A	PAEP	SYT1	P2RX7	
LTB		C20orf96	HMGB1P5	PARP12	TGFB1	PDGFB	
NCKAP1L		C22orf46	HMGB1P6	PDF	TMSB10	PPBP	
NLRP3		C2CD5	HSPA1L	PDGFA	TNF	PTGES	
OXTR		C3orf70	HSPA4	PGF	TNFRSF4	PTHLH	
PLA2G2A		CACTIN	IFI27	PIAS4	TNFRSF8	RARRES2	
PRKCB		CALM1	IFI6	PLAU	TNFSF15	SCG2	
RAC3		CAPZA2	IFIH1	PLXNB1	TNRC6A	SCGB3A1	
RAET1E		CAV1	IFIT1	POLR1D	TP53INP1	SIGLEC11	
SCG2		CCR7	IFIT2	PPARGC1B	TPT1	SIGLEC7	
SEMA3E		CDK17	IFIT3	PPP1CA	TSC2	SIRPB1	
SLAMF1		CELF2	IFNAR2	PRDX1	TUBB3	SLC11A1	
SOCS3		CELSR3	IGHD	PRSS36	TUFM	SMAD7	
TNFSF8		CFDP1	IGHV3-73	PSMD1	TXLNA	TERT	
TRAT1		CHPT1	IGKV2-24	PSMD11	TXN2	TNS3	
TRBV28		CLTC	IKBKB	PSMD14	TXNDC12	TRIM15	
TRIM36		CNOT8	IL11RA	PSMD2	TYK2	TRPM8	
TTK		CSF2	IL17RB	PSMD5	UBE2L6	TSLP	
UBD		CSRP1	IL32	PSMD7	UBXN11	TTK	
VGF		CTSG	IL34	PSMD8	ULBP2	TUBB3	
		DCTD	ILF3	PSME3	USB1	TYMS	
		DDX11	ISG15	PSPN	USP49		
		DDX12P	ITGA3	PSTPIP1	VEGFC		
		DDX19A	ITGB3BP	PTGS2	WNK2		
		DDX58	KITLG	PTX3	XRCC6		
		DDX60	KLF2	RABEP2	ZBTB1		
		DESI1	KLF8	RAC2			

side-effects of cigarette smoking. Invariably, a patient's life expectancy is significantly decreased when they smoke, and such a decrease is often proportional to the duration that a patient has been a smoker. Therefore, it is likely that the survival rates of cancer patients were more correlated with the comorbidities that present alongside smoking such as respiratory disorders. Additionally, it is possible that the effects of smoking on the tumor transcriptome were primarily on core tumor suppressors and oncogenes, and therefore enabled a faster progression to late-stage tumors and eventual morbidity. Our results may therefore suggest that, in cancers other than HNSCC, smoking-mediated immune dysregulations may play a smaller role in carcinogenesis than the other effects of smoking and would therefore not correlate well to survival rates.

The survival correlation results in HNSCC may be indicative of a unique smoking phenotype that we had hypothesized would present itself in all of the cancers studied. Previous studies have established HPV infection and smoking as two of the primary causes of HNSCC, with HPV being the most prevalent risk factor[34-37]. Cases of HNSCC that occur independently of HPV infection often occur later in life[38]. Together with the data from the survival analyses, these findings suggest that further investigation into HPV- smoking HNSCC patients may confirm the initial hypothesis of the presence of an immune genotype unique to smoking-mediated cancers.

Contracting effects TNF-related gene expression in LUSC and HNSCC

Overall, the landscapes of survival-correlated IA genes were largely different across cancers. However, further analysis of select IA genes and their correlations to cancer and immune pathways revealed potential connections between cancers that warrant further investigation.

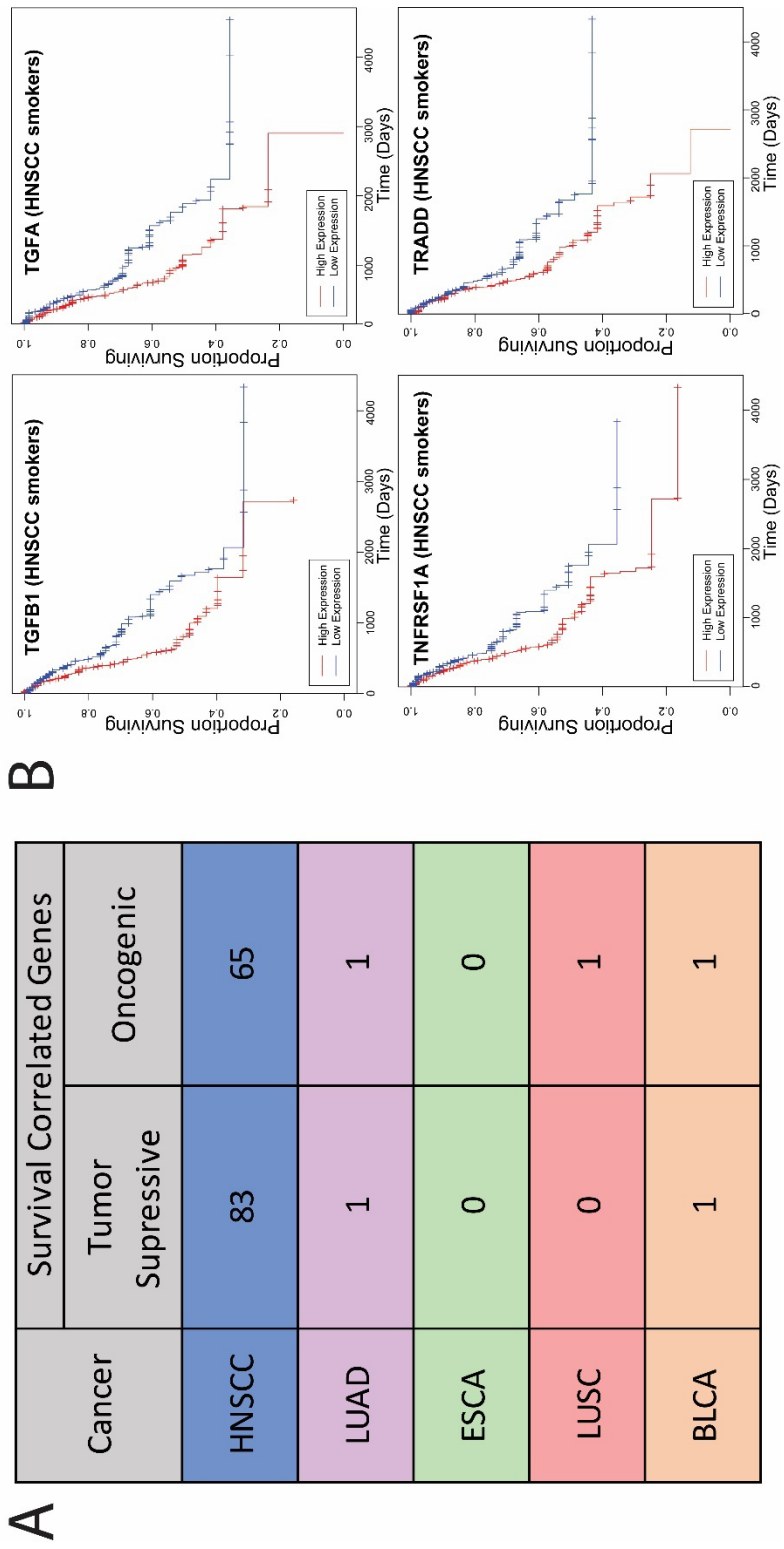


Figure 3: Survival correlations to IA gene expression. **A**. Total number of survival correlated genes per cancer. Tumor suppressive genes are those whose downregulation is correlated to poorer survival, while oncogenic genes are those whose upregulation is correlated to poorer survival. **B**. Representative Kaplan-Meier plots of oncogenic IA genes in HNSCC.

One such connection is the expression if TNF-related genes in HNSCC and LUSC. TNF is a proinflammatory cytokine that directs a host of immune mechanisms. Its role in cancer is paradoxical. As a tumor suppressor, TNF expression has been shown to trigger apoptosis of endothelial cells that are cancerous or precancerous[39,40]. As an oncogene, TNF can lead to the development of a hyperinflammatory state via indirect promotion of EMT-like and immune checkpoint gene expression patterns that lead to carcinogenesis and immune evasion[41-45]. Additionally, overproduction of TNF can lead to the production of immunosuppressive T-regs via TNFR2, which aid tumors in escaping the immune system[46].

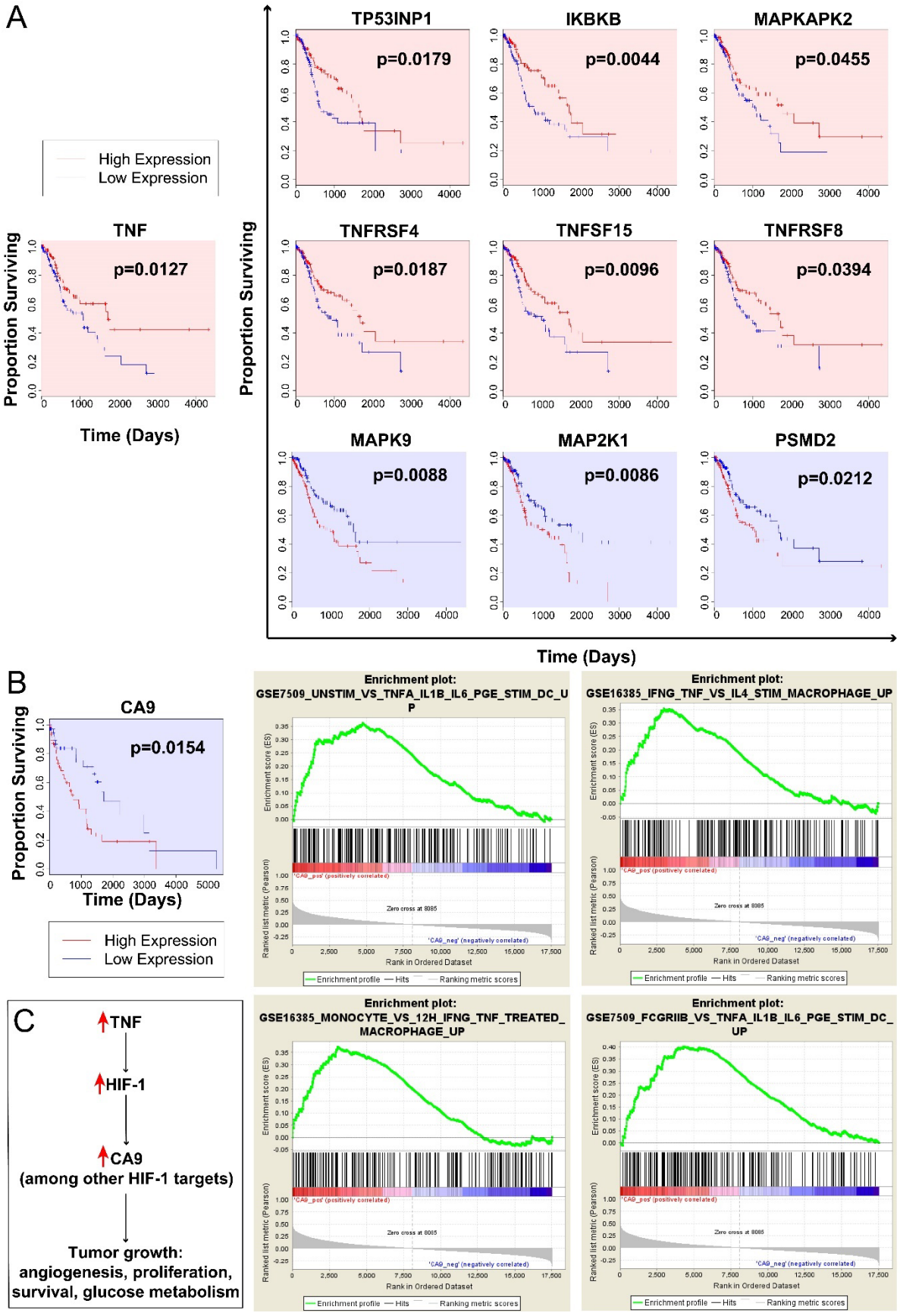
In HNSCC smoking patients, TNF itself along with a host of related genes are significantly dysregulated in smoking versus nonsmoking patients (Table 1). A group of these TNF-related genes are also correlated to survival (Figure 4A). Interestingly, their effects on survival are not consistent. While TNF, TP53INP1, IKBKB, MAPKAPK2, TNFRSF4, TNFSF15, and TNFRSF8 upregulation was correlated to increased survival rates, MAPK9, MAP2K1, and PSMD2 upregulation was correlated to poorer survival. These patterns corroborate previous findings about the potential conflicting roles of TNF expression. Expression of genes related to the MAPK pathway should work synergistically with TNF-related pathways to cause cancer. However, TNF-related genes in this cohort of HNSCC smoking patients seemingly protect against cancer-related morbidity.

TNF-related gene expression patterns in LUSC revolve around the dysregulation of CA9. CA9 is a metalloenzyme that is part of the HIF-1 transcription factor network (https://pathcards.genecards.org/card/hif-1-alpha_transcription_factor_network). HIF-1 expression has previously been identified as oncogenic, as HIF-1 is activated by the PI3K and MAPK pathways, which are hallmark cancer pathways[47]. Genes downstream of HIF-1 have been

Table 2: Survival correlated IA genes in smoking cancer patients

Cancers							
BLCA	ESCA	HNSCC				LUAD	LUSC
AVPR1A	SEMA3B	ACTR3	DDX11	LTBP2	RBP5	PTGES	CA9
UBD		ADAM17	DDX12P	LY96	RPS6KA5	TRPM8	
		ADCY10P1	DDX19A	MAP2K1	RUSC1-AS1		
		ADM	DESI1	MAPK9	S1PR2		
		AIMP1	DICER1-AS1	MAPKAPK2	SCARB1		
		ALKBH2	DKK1	MEF2D	SELENBP1		
		ALKBH6	DUSP3	MIF	SELENOT		
		AREG	EFTUD2	MUM1	SEMA4C		
		ARF1	EPHA3	NAMPT	SEMA5A		
		ARHGEF18	EPOP	NCKAP1	SH3BP2		
		ARID3B	ESR1	NDRG1	SHC3		
		ARL14EP	ETV6	NINJ2	SIGIRR		
		ARL2BP	FASTKD2	NR2C2	SIRT1		
		ARL4D	FLT4	NR2F1	SLX4		
		ARL6IP4	FYN	OSGIN1	STC2		
		ARRDC2	GALNS	OTUD5	STRAP		
		ATG16L1	GALNT3	PDF	SYNPO2		
		BECN1	GSTP1	PIAS4	SYT1		
		BMP6	HBEGF	PLXNB1	TNF		
		BST2	HIF1A	POLR1D	TNFRSF4		
		C1QBP	HMGB1P5	PPARGC1B	TNFRSF8		
		C20orf96	HMGB1P6	PPP1CA	TNFSF15		
		C22orf46	HSPA1L	PRDX1	TNRC6A		
		C2CD5	HSPA4	PRSS36	TP53INP1		
		C3orf70	IFNAR2	PSMD1	TPT1		
		CACTIN	IGHV3-73	PSMD11	TSC2		
		CALM1	IGKV2-24	PSMD14	TUFM		
		CAPZA2	IKBKB	PSMD2	TXLNA		
		CCR7	IL11RA	PSMD5	TXN2		
		CDK17	IL17RB	PSMD7	TXNDC12		
		CELF2	IL32	PSMD8	TYK2		
		CELSR3	ILF3	PSME3	UBXN11		
		CFDP1	ITGB3BP	PSPN	USB1		
		CHPT1	KITLG	PSTPIP1	USP49		
		CLTC	KLF2	PTGS2	WNK2		
		CNOT8	KLF8	PTX3	XRCC6		
		CSRP1	LDLRAD3	RABEP2	ZBTB1		
		DCTD	LRRC41	RASGEF1B			

Figure 4: Conflicting effects of TNF-related expression in HNSCC and LUSC smokers **A.** Survival plots of TNF and its related genes that are significantly dysregulated and correlated to survival in HNSCC smoking patients. **B.** CA9 correlations to survival and immunological signatures via GSEA in LUSC smoking patients. **C.** A schematic of the hypothesized relationship between CA9, TNF, and their synergistic promotion of LUSC in smokers specifically.


A
High Expression
Low Expression
TNF
p=0.0127
Proportion Surviving
Time (Days)
TP53INP1
p=0.0179
IKBKB
p=0.0044
MAPKAPK2
p=0.0455
TNFRSF4
p=0.0187
TNFSF15
p=0.0096
TNFRSF8
p=0.0394
MAPK9
p=0.0088
MAP2K1
p=0.0086
PSMD2
p=0.0212
Time (Days)
B
CA9
p=0.0154
Proportion Surviving
Time (Days)
High Expression
Low Expression
Enrichment plot:
GSE7509_UNSTIM_VS_TNFA_IL1B_IL6_PGE_STIM_DC_UP
Enrichment plot:
GSE16385_IFNG_TNF_VS_IL4_STIM_MACROPHAGE_UP
Enrichment plot:
GSE16385_MONOCYTE_VS_12H_IFNG_TNF_TREATED_MACROPHAGE_UP
Enrichment plot:
GSE7509_FCGRIIB_VS_TNFA_IL1B_IL6_PGE_STIM_DC_UP
Enrichment score (ES)
Ranked list metric (Pearson)
Rank in Ordered Dataset
Enrichment profile
Hits
Ranking metric scores
C
TNF
HIF-1
CA9
(among other HIF-1 targets)
Tumor growth:
angiogenesis, proliferation,
survival, glucose metabolism

shown to promote glucose metabolism, cell proliferation, and survival via angiogenesis in tumors[48-50]. Analysis of smoking LUSC patients indicated that CA9 expression was significantly correlated to poorer patient survival (Figure 4B). Additionally, CA9 upregulation was correlated to the upregulation of immune signatures related to TNF-mediated activation of the immune system (Figure 4B). Together, these findings suggest a role of TNF and CA9 in LUSC smokers that is opposite that of HNSCC smokers. Namely, in LUSC smokers, TNF serves as an activator of HIF-1, which, through downstream genes like CA9, promotes tumorigenesis (Figure 4C). Therefore, TNF may act as an oncogene in LUSC smokers, but may act as a tumor suppressor in HNSCC smokers.

CHAPTER 4: OSTEOPONTIN AND SMOKING

Osteopontin expression in HNSCC smoking samples

In a separate project conducted by the Ongkeko Lab, a panel of genes was found to be correlated to cancer development in a non-TCGA set of oral squamous cell carcinoma samples from mice (Figure 5A)[51]. One of these genes was Osteopontin (OPN, aka SPP1), a gene that was included in the list of IA genes used for this thesis. OPN is a glycoprotein that was first discovered in osteoblasts but is expressed by many cell types, including epithelial cells, endothelial cells, and immune cells[52]. It is a critical regulator of cellular pathways implicated in survival, proliferation, and cell-to-cell signaling, and its reported downstream targets include other master regulators of cellular functions. A major way through which OPN executes its functions is binding to integrin proteins, which transduce extracellular signals to activate a signaling cascade within the cell. The integrins α¬5β1, α¬5β3, α¬5β5, α¬4β1, α¬8β1, and α¬9β1 are known targets of OPN, and binding to each of the above integrins would result in different

cellular responses[53]. Another significant function of OPN is the recruitment of immune cells and promotion of an inflammatory phenotype. It can also regulate specific cellular pathways inside macrophages, including downregulation of IL-10 and iNOS production[54,55].

Analyzing OPN expression in HNSCC smokers indicated that its upregulation was correlated to advanced clinical stage in TCGA samples (Figure 5B). In the initial analysis for this thesis, OPN was differentially expressed between smoking and nonsmoking HNSCC patients, but its expression was not correlated to survival (Table 2). As discussed previously, the survival and differential expression analyses indicated that an analysis of HPV- smoking HNSCC patients may suggest an immune genotype that is different from HPV+ patients. It was therefore planned that, to determine whether OPN may indeed play a role in the IA landscape of smoking-mediated HNSCC, HNSCC samples would be reanalyzed using differential expression to compare HPV- smokers to HPV- nonsmokers. However, there were fewer than 10 HNSCC samples from HPV- nonsmokers, which did not allow a statistically significant analysis. Therefore, HPV- smoking HNSCC patients were compared to solid tissue normal samples. Our results indicated that OPN was significantly upregulated in the HPV- smoking HNSCC samples (Figure 5C). To provide additional validation of OPN's upregulation, HNSCC cell lines (UM-SCC-10b and UM-SCC-22b) were exposed to cigarette smoke *in vitro*. PCR analysis of OPN expression confirmed its upregulation in smoking-exposed versus control HNSCC cell lines (Figure 5D).

To expand on the effects of OPN upregulation, OPN expression was compared to the expression of other immune and cancer genes and pathways. Gene co-expression analyses using Spearman scatterplots indicated that OPN upregulation was correlated to the upregulation of ITGA5, ITGB3, PRKCA, and MMP9 and the downregulation of ESRRA (Figure 6A). GSEA using OPN expression and MSigDB C2 pathways indicated that OPN expression is significantly

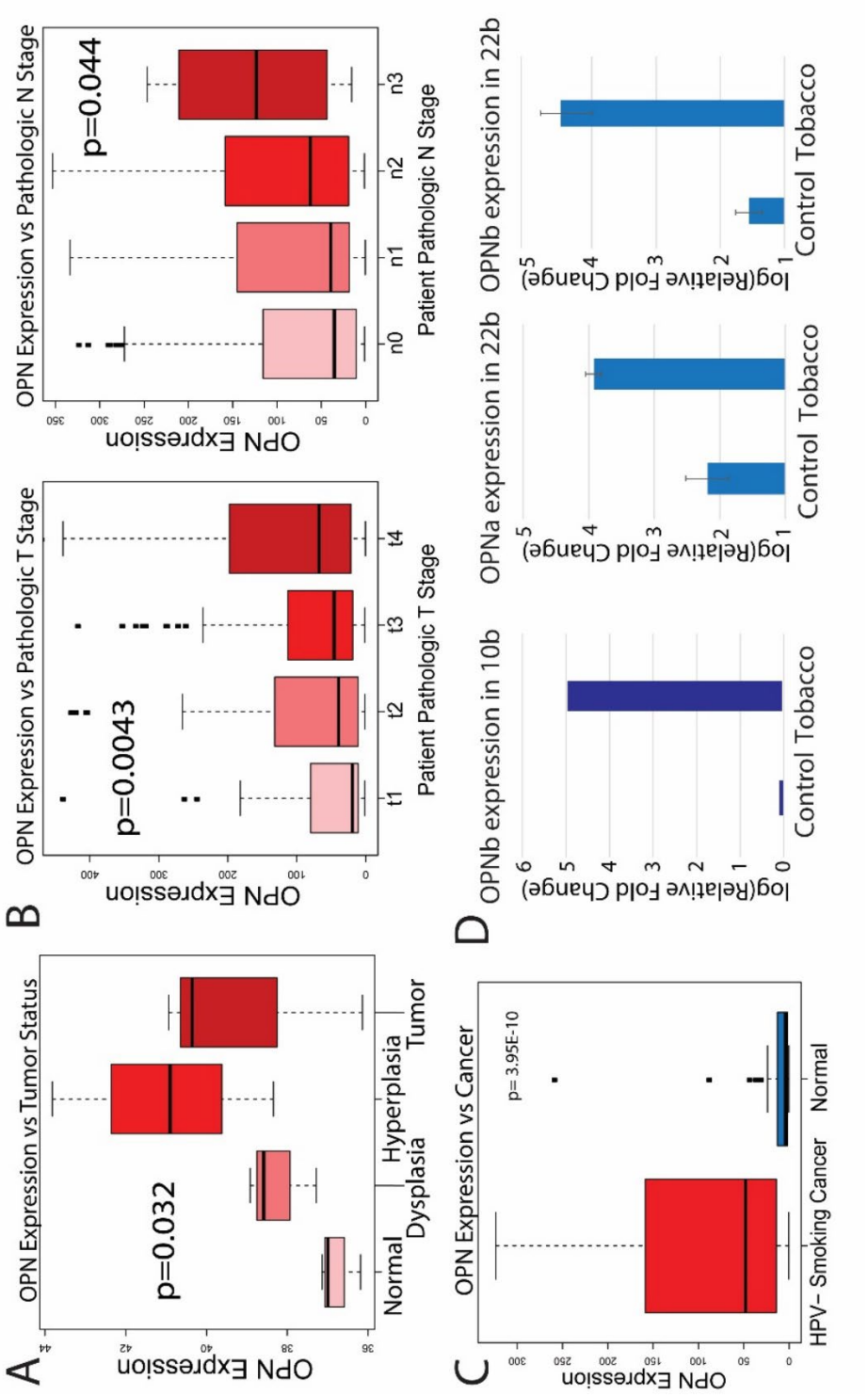


Figure 5: Correlation of cancer and immune pathways to OPN expression in HNSCC. **A.** OPN expression versus tumor status in a non-TCGA validation cohort of OSCC samples. **B.** Correlation of OPN expression to Pathologic T and N stage in smoking HNSCC patients from TCGA. **C.** OPN dysregulation in TCGA HPV- smoking vs normal samples. **D.** *In vitro* validation of OPN upregulation caused by exposure to cigarette smoke.

correlated to the dysregulation of multiple tumor and immune pathways (Figure 6B). These results along with the interactome of OPN suggest that OPN may be a contributor to a unique, smoking-mediated phenotype of HNSCC (Figure 6C).

Pan-cancer Osteopontin expression

Due to its potential importance in HNSCC smoking patients, OPN was further analyzed in the other smoking-mediated cancers in TCGA. The Kruskal Wallis box plots of OPN expression indicated some potential dysregulation patterns (Figure 7A). Further analysis of OPN expression in ESCA indicated that it is upregulated in all cancer samples versus normal samples (Figure 7B). Interestingly, OPN expression seems to be slightly higher in smoking ESCA cases than nonsmoking ESCA cases (Figure 7B). According to the original differential expression analysis, this OPN dysregulation pattern is statistically significant (Table 1). Unlike in HNSCC, however, in ESCA, OPN upregulation was significantly correlated to the presence of mutations, determined using the REVEALER algorithm. Specifically, OPN upregulation is correlated to the increased presence of IP6K2, GRID2, and C1orf125 mutations (Figure 7C). IP6K2 codes for an isoform of inositol hexakisphosphate kinase (IP6K). It generates IP7, a well-established cell growth and homeostasis regulator[56,57]. Loss of function of IP6K2 has been correlated to an increased incidence of aero-digestive tract cancer in mice[58]. Its expression induces apoptosis and can be turned off using cellular machinery shared by the heat shock protein 90 pathway[59,60]. GRID2 has been shown to be regulated by TNF, and its induction by TNF leads to apoptosis via caspase-3 activation[61]. C1orf125 (aka AXDND1) is not well-studied in the context of cancer, but has been linked to sperm motility[62]. The mutation of IP6K2 and GRID2, depending on whether mutations cause a loss or gain of function, may indicate a link between OPN expression and

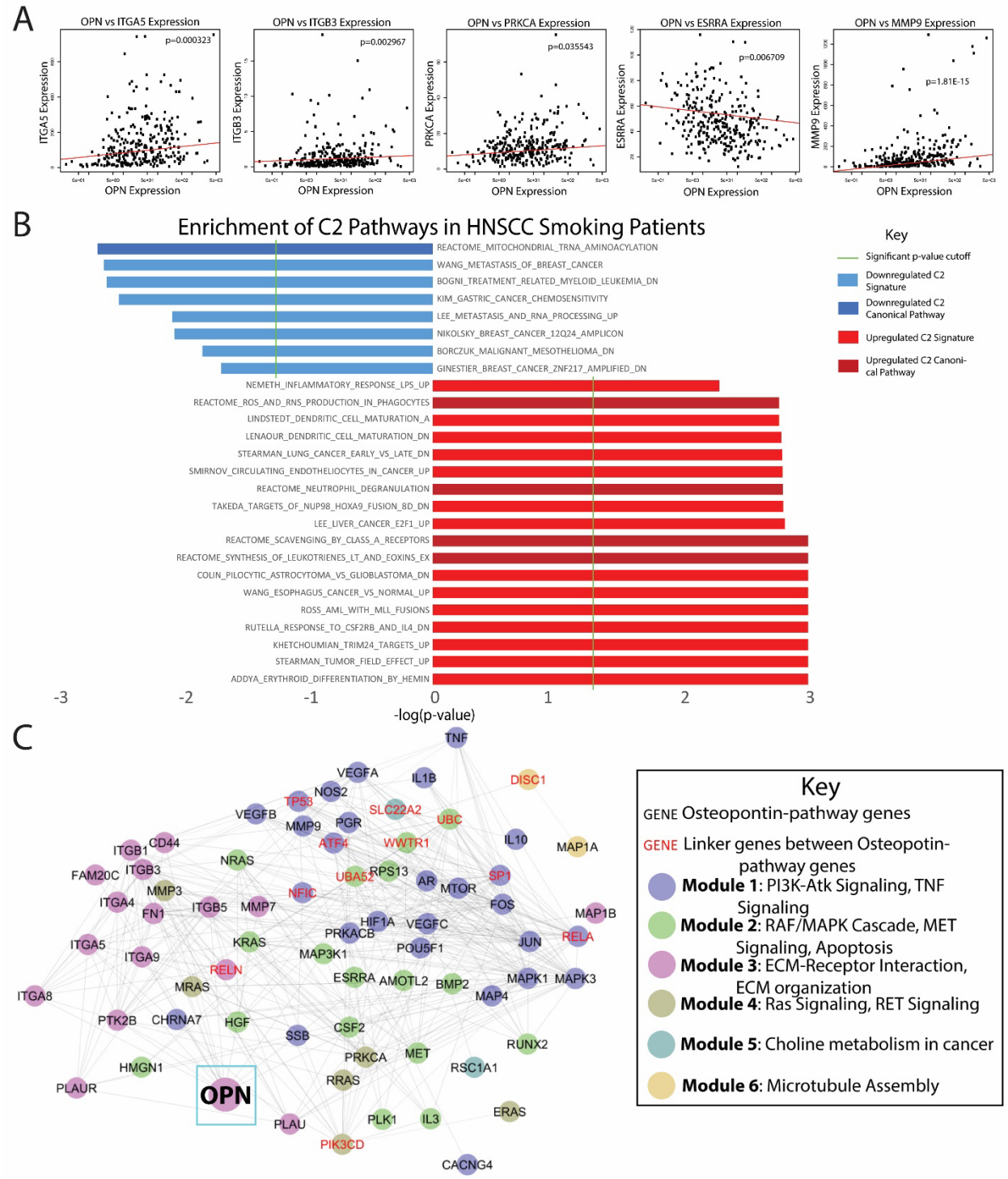


Figure 6: Correlation of cancer and immune pathways to OPN expression in HNSCC. **A.** Spearman plots of gene co-expression analysis between OPN and other cancer and immune-associated genes. **B.** Enrichment of C2 pathways correlated to OPN expression using GSEA. **C.** The OPN interactome. Red node labels indicate genes that are significantly dysregulated in HPV- smoking HNSCC samples.

Inhibition of apoptosis. The likelihood of this interaction occurring *in situ* is further justified by OPN's direct and indirect interactions with key genes involved in apoptosis (Figure 6C). Further study is needed to clarify the connections to OPN.

CHAPTER 5: PAN-CANCER miRNA EXPRESSION PATTERNS

Overall, miRNA expression does not seem to be a significant contributor to the trends observed in IA gene expression. After discovering the most highly conserved miRNA targets for each of the survival correlated IA genes of interest, these miRNA were analyzed for their dysregulation between smoking and nonsmoking cancer samples. Surprisingly, there were no significantly dysregulated miRNAs. In previous projects, the Ongkeko Lab has identified miRNAs that are correlated to specific immune phenotypes in cancer[63]. However, these analyses were conducted by comparing miRNA expression in cancer versus normal samples primarily. It may therefore be the case that miRNA expression patterns are not uniquely dysregulated by smoking. Indeed, Principal Component Analysis of the landscape of miRNA expression per cancer indicated that miRNA expression alone cannot be used to reliably distinguish between smoking cancer and nonsmoking cancer samples (Figure 8). In fact, for some cancers like LUSC, BLCA, and LUSC, miRNA expression patterns seemingly caused smokers and nonsmokers to be more similar to each other (indicated by tighter clustering patterns).

These results are in stark contrast to what was expected. Previous studies have strongly suggested that smoking is connected to miRNA expression. Willinger et al. discovered a 6-miRNA panel that could be used as a signature for smoking and was correlated to inflammatory signals[64]. Schembri et al. were also able to identify a finite set of miRNAs at the lung epithelium that were induced by smoking[65]. Numerous other findings corroborate the fact that smoking does indeed affect miRNA expression, and that there may be alterations by smoking that cause

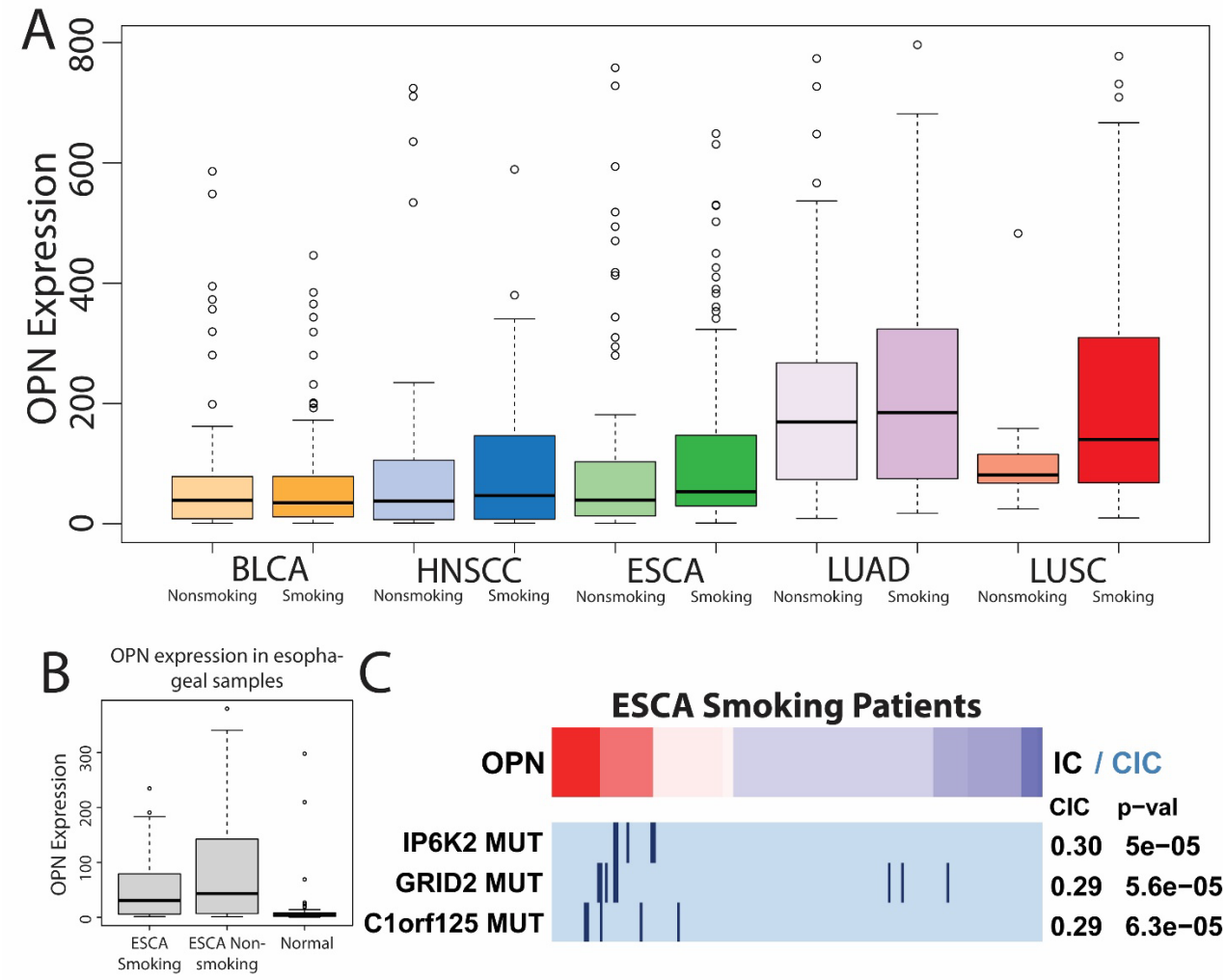


Figure 7: Pan-cancer OPN expression trends. **A**. A comparison of OPN expression between smoking and nonsmoking cancer patients on a pan-cancer scale. **B**. OPN expression in smoking ESCA, nonsmoking ESCA, and normal samples. **C**. REVEALER analysis correlating OPN expression to the frequency of mutations in smoking ESCA patients.

unique miRNA dysregulation[65]. However, from these studies, a common trend is the finite number of miRNAs identified as significant contributors to disease phenotypes. MiRNA expression is most commonly associated with fine-tuning of gene expression rather than with larger changes[66,67]. This is in stark contrast to other targets of smoking like DNA damage, which are known to cause significant changes to cell genotype and phenotype. Therefore, the observations in this thesis related to miRNA expression may be due to the much larger contributions of other machinery affected by smoking, which overshadow the more modest effects of miRNA-mediated silencing, especially when comparing between cancer samples.

CHAPTER 6: DISCUSSION

Interpretation of results

Though it was initially hypothesized that smoking would cause IA gene dysregulation patterns that would be shared amongst multiple smoking-mediated cancers, the findings of this thesis do not exactly corroborate that point. Firstly, IA genes that were uniquely dysregulated in smoking cancer patients and contributed significantly to survival did not overlap well between cancers. HNSCC may be the only cancer with a smoking-specific IA genotype. This may be due to the nature of HNSCC to occur most commonly due to HPV infection. This hypothesis was tested by analyzing HPV- HNSCC patients in particular, which indicated the potential role of OPN as a smoking-specific oncogene. OPN was shown to be correlated to a more advanced clinical stage and to the expression of several key cancer and immune signatures. Additionally, OPN was significantly dysregulated in ESCA samples, and was correlated to mutations of apoptosis inhibitor genes. These findings suggest an importance of OPN that should be studied further in other cancers. Though its expression was not significantly different between smoking

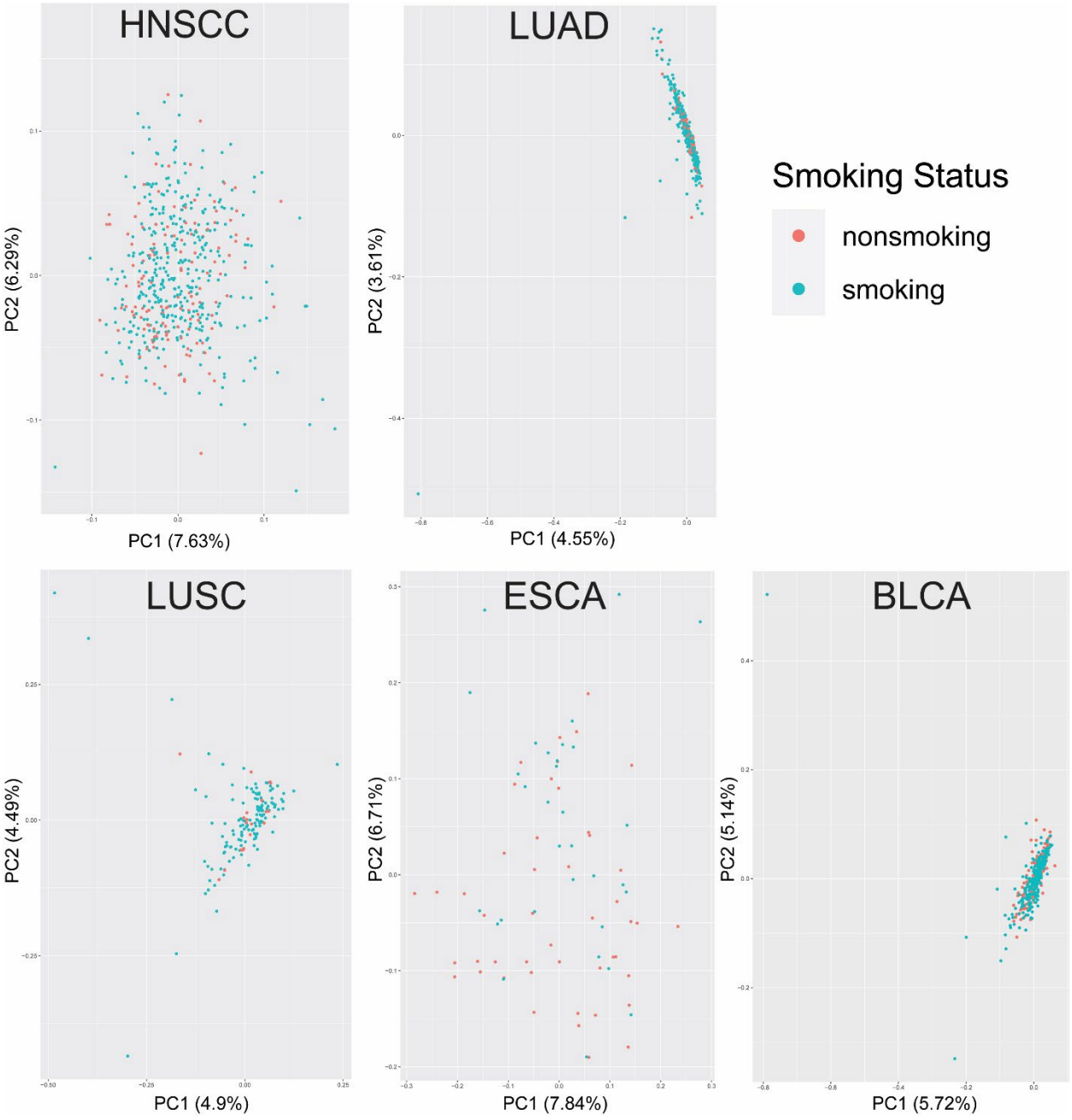


Figure 8: Pan-cancer miRNA expression patterns. PCA plots of aggregate miRNA expression read counts in smoking and nonsmoking cancer patients. The analysis determines how the variation in the data can be explained by miRNA read count patterns. Percentages on the axis labels correspond to the percentage of variation each principal component is able to account for.

and nonsmoking cancer patients for other cancers, conducting a similar re-analysis as was done for HNSCC when HPV+ patients were removed may indicate that OPN is indeed upregulated in smoking cancer patients.

Another novel relationship revealed by this thesis is the conflicting roles of TNF and its downstream genes in carcinogenesis. In HNSCC samples, TNF seems to abate cancer progression, acting as a tumor suppressor that is able to counteract aberrant MAPK pathway gene overexpression. In LUSC samples, TNF mechanisms related to immune evasion by tumors and the HIF-1 pathway are overexpressed, leading to poorer patient survival. These conflicting activities, if proven valid, may indicate that smoking can affect the same IA genes in different ways depending on the cancer type and location of a tumor.

It was originally hypothesized that miRNA expression would play a role in a smoking-specific IA genotype. However, this was not the case, as miRNA expression was largely similar between smoking and nonsmoking cancer patients. This may be due to the fact that miRNAs exert more modest and targeted effects on the transcriptome than smoking does. Therefore, miRNA expression may be overshadowed by the large, systemic effects of smoking that have already significantly changed the genomic and transcriptomic landscape of a cell. If there were any significant contributions by miRNAs to the difference between smoking and nonsmoking cancers, we hypothesize that they would manifest in HNSCC or ESCA, due to the less-dense clustering of smokers and nonsmokers (Figure 8).

Setbacks and potential solutions

The initial proposal for this thesis included significant *in vitro* work, including validation of the targets discovered through computational analysis. However, due to technical difficulties, this *in vitro* work was not accomplished. It was initially planned that genes of interest such as the

TNF-related genes would have their smoking-specific dysregulation patterns validated *in vitro* using PCR. Additionally, various functional assays were to be used to determine how knockdown or overexpression of these genes could contribute to carcinogenesis. Due to the COVID-19 pandemic, the Ongkeko Lab's wet lab was completely closed for almost a full academic year, which coincided with half of the time taken to draft this thesis. In the second year of the master's program, the *in vitro* experiments were set to take place. But, unfortunately, a freezer malfunction over the COVID-19-related shut down led to the loss of all of the Ongkeko Lab's cell lines. It was therefore impossible for *in vitro* validation and functional assays to be completed. In the future, the lab hopes to rebuild its repertoire of cell lines in order to achieve the validation of the computational data in this thesis. Additionally, the effects of these genes on the immune system will be analyzed *in vivo* using mouse models that enable study of the immune system in a living model.

Future Directions

Overall, the results of this thesis partially confirmed our hypothesis of the existence of a set of IA gene dysregulations that is uniquely caused by smoking in cancer patients. Aside from validating the computational findings, we hope to explore each individual relationship further. The conflicting patterns of TNF expression have not yet been fully understood, and elaborations on the findings of this paper may contribute to our understanding of TNF as both an oncogene and a tumor suppressor. This may be vital for future therapies that utilize TNF, as targeting it or its related genes may not have the same effect in patients depending on whether or not they smoke and on what type of cancer they have. OPN also may be an important target that may be used for future therapeutics. Since its expression patterns seem relatively similar across cancers, the study of OPN in different, larger datasets of smoking cancer patients may be useful in

validating the trends observed in this thesis. It may be important to account and correct for as many patient demographic factors and comorbidities as possible in order to study OPN, or any other gene form this thesis for that matter, in the future in smoking cancer patients. Future studies may identify the knockdown of OPN as a potential immune therapy that is specific to smokers.

Though the results of the miRNA analysis were not as expected, miRNAs may still be important in smoking-specific cancers, especially considering the heavy emphasis that previous findings have put on the smoking-miRNA axis in both inflammation and cancer. Some future directions may entail understanding how miRNA expression can be more optimally analyzed in order to clarify their role in smoking-mediated cancers. Even if said role is not as significant of a contributor to cancer progression as other genomic and transcriptomic changes, miRNAs may still be taken advantage of for emerging therapies.

The analysis detailed in this thesis is in no way exhaustive due to the nature of analyzing a dataset as large as TCGA. There may be many correlations that remain unexplored and many aspects of smoking-mediated cancer that can still be discovered. We hope to completely analyze the significantly dysregulated IA genes in all cancers in the near future. Though the survival correlations were important in helping us narrow our scope of study, it may have also prevented us from seeing important connections. Therefore, we hope to analyze all dysregulated IA genes individually, independent of survival correlations. We are also working on projects that correlate smoking to targets outside of the genome and transcriptome, such as CpG methylation and the intratumor microbiome.

This thesis in full will be submitted for publication. The thesis author will be the primary investigator author of this paper, and Ongkeko, Weg M. will be the principal investigator.

REFERENCES

1. Smith C, Hansch C. The relative toxicity of compounds in mainstream cigarette smoke condensate. *Food and Chemical Toxicology.* 2000;38(7):637-646.
2. Rubinstein ML, Delucchi K, Benowitz NL, Ramo DE. Adolescent Exposure to Toxic Volatile Organic Chemicals From E-Cigarettes. *Pediatrics.* 2018;141(4):e20173557.
3. Wiencke JK. DNA adduct burden and tobacco carcinogenesis. *Oncogene.* 2002;21(48):7376-7391.
4. Sanner T, Grimsrud TK. Nicotine: Carcinogenicity and Effects on Response to Cancer Treatment - A Review. *Front Oncol.* 2015;5:196-196.
5. Roos WP, Kaina B. DNA damage-induced cell death by apoptosis. *Trends Mol Med.* 2006;12(9):440-450.
6. Barrow TM, Klett H, Toth R, Böhm J, Gigic B, Habermann N, Scherer D, Schrotz-King P, Skender S, Abbenhardt-Martin C, Zielske L, Schneider M, Ulrich A, Schirmacher P, Herpel E, Brenner H, Busch H, Boerries M, Ulrich CM, Michels KB. Smoking is associated with hypermethylation of the APC 1A promoter in colorectal cancer: the ColoCare Study. *J Pathol.* 2017;243(3):366-375.
7. Barnes PJ. Alveolar macrophages as orchestrators of COPD. *COPD: Journal of Chronic Obstructive Pulmonary Disease.* 2004;1(1):59-70.
8. Sopori M. Effects of cigarette smoke on the immune system. *Nature Reviews Immunology.* 2002;2(5):372-377.
9. Stämpfli MR, Anderson GP. How cigarette smoke skews immune responses to promote infection, lung disease and cancer. *Nature Reviews Immunology.* 2009;9(5):377-384.

10. Burns AR, Hosford SP, Dunn LA, Walker DC, Hogg JC. Respiratory epithelial permeability after cigarette smoke exposure in guinea pigs. *Journal of Applied Physiology.* 1989;66(5):2109-2116.

11. Jones J, Lawler P, Crawley J, Minty B, Hulands G, Veall N. Increased alveolar epithelial permeability in cigarette smokers. *The Lancet.* 1980;315(8159):66-68.

12. Dye JA, Adler KB. Effects of cigarette smoke on epithelial cells of the respiratory tract. *Thorax.* 1994;49(8):825.

13. Kim H, Liu X, Kobayashi T, Conner H, Kohyama T, Wen F-Q, Fang Q, Abe S, Bitterman P, Rennard SI. Reversible cigarette smoke extract–induced DNA damage in human lung fibroblasts. *American journal of respiratory cell and molecular biology.* 2004;31(5):483-490.

14. Kamp DW, Greenberger MJ, Sbalchierro JS, Preusen SE, Weitzman SA. Cigarette smoke augments asbestos-induced alveolar epithelial cell injury: role of free radicals. *Free Radical Biology and Medicine.* 1998;25(6):728-739.

15. Mio T, Romberger DJ, Thompson AB, Robbins RA, Heires A, Rennard SI. Cigarette smoke induces interleukin-8 release from human bronchial epithelial cells. *American journal of respiratory and critical care medicine.* 1997;155(5):1770-1776.

16. Ji H, Houghton A, Mariani T, Perera S, Kim C, Padera R, Tonon G, McNamara K, Marconcini L, Hezel A. K-ras activation generates an inflammatory response in lung tumors. *Oncogene.* 2006;25(14):2105-2112.

17. Chalmer J, Holt P, Keast D. Cell-mediated immune responses to transplanted tumors in mice chronically exposed to cigarette smoke. *Journal of the National Cancer Institute.* 1975;55(5):1129-1134.

18. Lu L-M, Zavitz CC, Chen B, Kianpour S, Wan Y, Stämpfli MR. Cigarette smoke impairs NK cell-dependent tumor immune surveillance. *The Journal of Immunology.* 2007;178(2):936-943.

19. Ng SP, Silverstone AE, Lai Z-W, Zelikoff JT. Effects of prenatal exposure to cigarette smoke on offspring tumor susceptibility and associated immune mechanisms. *Toxicological sciences.* 2006;89(1):135-144.

20. Fortin A, Wang CS, Vigneault É. Influence of smoking and alcohol drinking behaviors on treatment outcomes of patients with squamous cell carcinomas of the head and neck. *International Journal of Radiation Oncology* Biology* Physics.* 2009;74(4):1062-1069.

21. Rischin D, Young RJ, Fisher R, Fox SB, Le Q-T, Peters LJ, Solomon B, Choi J, O'Sullivan B, Kenny LM. Prognostic significance of p16INK4A and human papillomavirus in patients with oropharyngeal cancer treated on TROG 02.02 phase III trial. *Journal of clinical oncology.* 2010;28(27):4142.

22. Bhardwaj N. Harnessing the immune system to treat cancer. *The Journal of clinical investigation.* 2007;117(5):1130-1136.

23. Spanos WC, Nowicki P, Lee DW, Hoover A, Hostager B, Gupta A, Anderson ME, Lee JH. Immune response during therapy with cisplatin or radiation for human papillomavirus–related head and neck cancer. *Archives of otolaryngology–head & neck surgery.* 2009;135(11):1137-1146.

24. Desrichard A, Kuo F, Chowell D, Lee K-W, Riaz N, Wong RJ, Chan TA, Morris LG. Tobacco smoking-associated alterations in the immune microenvironment of squamous cell carcinomas. *JNCI: Journal of the National Cancer Institute.* 2018;110(12):1386-1392.

25. de la Iglesia JV, Slebos RJC, Martin-Gomez L, Wang X, Teer JK, Tan AC, Gerke TA, Aden-Buie G, van Veen T, Masannat J, Chaudhary R, Song F, Fournier M, Siegel EM, Schabath MB, Wadsworth JT, Caudell J, Harrison L, Wenig BM, Conejo-Garcia J, Hernandez-Prera JC, Chung CH. Effects of Tobacco Smoking on the Tumor Immune Microenvironment in Head and Neck Squamous Cell Carcinoma. *Clin Cancer Res.* 2020;26(6):1474-1485.

26. M Clement J, Duan F, Srivastava PK. Smoking-induced immune deviation contributes to progression of bladder and other cancers. *Oncoimmunology.* 2015;4(9):e1019199-e1019199.

27. Zeegers MP, Tan FE, Dorant E, van den Brandt PA. The impact of characteristics of cigarette smoking on urinary tract cancer risk: a meta-analysis of epidemiologic studies. *Cancer.* 2000;89(3):630-639.

28. Wang G, Pan C, Cao K, Zhang J, Geng H, Wu K, Wen J, Liu C. Impacts of Cigarette Smoking on the Tumor Immune Microenvironment in Esophageal Squamous Cell Carcinoma. *J Cancer.* 2022;13(2):413-425.

29. Liberzon A, Subramanian A, Pinchback R, Thorvaldsdóttir H, Tamayo P, Mesirov JP. Molecular signatures database (MSigDB) 3.0. *Bioinformatics.* 2011;27(12):1739-1740.

30. Cerami EG, Gross BE, Demir E, Rodchenkov I, Babur Ö, Anwar N, Schultz N, Bader GD, Sander C. Pathway Commons, a web resource for biological pathway data. *Nucleic acids research.* 2010;39(suppl_1):D685-D690.

31. Subramanian J, Govindan R. Lung cancer in never smokers: a review. *Journal of clinical oncology.* 2007;25(5):561-570.

32. Couraud S, Zalcman G, Milleron B, Morin F, Souquet P-J. Lung cancer in never smokers–a review. *European journal of cancer.* 2012;48(9):1299-1311.

33. Kenfield SA, Wei EK, Stampfer MJ, Rosner BA, Colditz GA. Comparison of aspects of smoking among the four histological types of lung cancer. *Tobacco control.* 2008;17(3):198-204.

34. Stein AP, Saha S, Kraninger JL, Swick AD, Yu M, Lambertg PF, Kimple R. Prevalence of human papillomavirus in oropharyngeal cancer: a systematic review. *Cancer journal (Sudbury, Mass).* 2015;21(3):138.

35. Isayeva T, Li Y, Maswahu D, Brandwein-Gensler M. Human papillomavirus in non-oropharyngeal head and neck cancers: a systematic literature review. *Head and neck pathology.* 2012;6(1):104-120.

36. Michaud DS, Langevin SM, Eliot M, Nelson HH, Pawlita M, McClean MD, Kelsey KT. High-risk HPV types and head and neck cancer. *International journal of cancer.* 2014;135(7):1653-1661.

37. Blot WJ, McLaughlin JK, Winn DM, Austin DF, Greenberg RS, Preston-Martin S, Bernstein L, Schoenberg JB, Stemhagen A, Fraumeni JF. Smoking and drinking in relation to oral and pharyngeal cancer. *Cancer research.* 1988;48(11):3282-3287.

38. Windon MJ, D'Souza G, Rettig EM, Westra WH, van Zante A, Wang SJ, Ryan WR, Mydlarz WK, Ha PK, Miles BA. Increasing prevalence of human papillomavirus–positive oropharyngeal cancers among older adults. *Cancer.* 2018;124(14):2993-2999.

39. Pennica D, Nedwin GE, Hayflick JS, Seeburg PH, Derynck R, Palladino MA, Kohr WJ, Aggarwal BB, Goeddel DV. Human tumour necrosis factor: precursor structure, expression and homology to lymphotoxin. *Nature.* 1984;312(5996):724-729.

40. Carswell E, Old LJ, Kassel R, Green S, Fiore N, Williamson B. An endotoxin-induced serum factor that causes necrosis of tumors. *Proceedings of the National Academy of Sciences.* 1975;72(9):3666-3670.

41. Ho M-Y, Tang S-J, Chuang M-J, Cha T-L, Li J-Y, Sun G-H, Sun K-H. TNF-α induces epithelial–mesenchymal transition of renal cell carcinoma cells via a GSK3β-dependent mechanism. *Molecular cancer research.* 2012;10(8):1109-1119.

42. Li C-W, Xia W, Huo L, Lim S-O, Wu Y, Hsu JL, Chao C-H, Yamaguchi H, Yang N-K, Ding Q. Epithelial–mesenchymal transition induced by TNF-α requires NF-κB–mediated transcriptional upregulation of Twist1. *Cancer research.* 2012;72(5):1290-1300.

43. Liao S-J, Luo J, Li D, Zhou Y-H, Yan B, Wei J-J, Tu J-C, Li Y-R, Zhang G-M, Feng Z-H. TGF-β1 and TNF-α synergistically induce epithelial to mesenchymal transition of breast cancer cells by enhancing TAK1 activation. *J Cell Commun Signal.* 2019;13(3):369-380.

44. Reinhardt J, Landsberg J, Schmid-Burgk JL, Ramis BB, Bald T, Glodde N, Lopez-Ramos D, Young A, Ngiow SF, Nettersheim D. MAPK signaling and inflammation link melanoma phenotype switching to induction of CD73 during immunotherapy. *Cancer research.* 2017;77(17):4697-4709.

45. Zingg D, Arenas-Ramirez N, Sahin D, Rosalia RA, Antunes AT, Haeusel J, Sommer L, Boyman O. The histone methyltransferase Ezh2 controls mechanisms of adaptive resistance to tumor immunotherapy. *Cell reports.* 2017;20(4):854-867.

46. Salomon BL, Leclerc M, Tosello J, Ronin E, Piaggio E, Cohen JL. Tumor Necrosis Factor α and Regulatory T Cells in Oncoimmunology. *Frontiers in Immunology.* 2018;9.

47. Semenza GL. Targeting HIF-1 for cancer therapy. *Nature Reviews Cancer.* 2003;3(10):721-732.

48. Harris AL. Hypoxia—a key regulatory factor in tumour growth. *Nature reviews cancer.* 2002;2(1):38-47.

49. Carmeliet P, Dor Y, Herbert J-M, Fukumura D, Brusselmans K, Dewerchin M, Neeman M, Bono F, Abramovitch R, Maxwell P. Role of HIF-1α in hypoxia-mediated apoptosis, cell proliferation and tumour angiogenesis. *Nature.* 1998;394(6692):485-490.

50. Ryan HE, Lo J, Johnson RS. HIF-1α is required for solid tumor formation and embryonic vascularization. *The EMBO journal.* 1998;17(11):3005-3015.

51. Foy JP, Tortereau A, Caulin C, Le Texier V, Lavergne E, Thomas E, Chabaud S, Perol D, Lachuer J, Lang W, Hong WK, Goudot P, Lippman SM, Bertolus C, Saintigny P. The dynamics of gene expression changes in a mouse model of oral tumorigenesis may help refine prevention and treatment strategies in patients with oral cancer. *Oncotarget.* 2016;7(24):35932-35945.

52. Lund SA, Giachelli CM, Scatena M. The role of osteopontin in inflammatory processes. *J Cell Commun Signal.* 2009;3(3-4):311-322.

53. Chen Q, Shou P, Zhang L, Xu C, Zheng C, Han Y, Li W, Huang Y, Zhang X, Shao C, Roberts AI, Rabson AB, Ren G, Zhang Y, Wang Y, Denhardt DT, Shi Y. An osteopontin-integrin interaction plays a critical role in directing adipogenesis and osteogenesis by mesenchymal stem cells. *Stem Cells.* 2014;32(2):327-337.

54. Shirakawa K, Endo J, Kataoka M, Katsumata Y, Yoshida N, Yamamoto T, Isobe S, Moriyama H, Goto S, Kitakata H, Hiraide T, Fukuda K, Sano M. IL (Interleukin)-10-

STAT3-Galectin-3 Axis Is Essential for Osteopontin-Producing Reparative Macrophage Polarization After Myocardial Infarction. *Circulation.* 2018;138(18):2021-2035.

55. Guo H, Wai PY, Mi Z, Gao C, Zhang J, Kuo PC. Osteopontin mediates Stat1 degradation to inhibit iNOS transcription in a cecal ligation and puncture model of sepsis. *Surgery.* 2008;144(2):182-188.

56. Shears SB. Diphosphoinositol polyphosphates: metabolic messengers? *Molecular pharmacology.* 2009;76(2):236-252.

57. Barker CJ, Illies C, Gaboardi GC, Berggren P-O. Inositol pyrophosphates: structure, enzymology and function. *Cellular and molecular life sciences.* 2009;66(24):3851-3871.

58. Morrison B, Haney R, Lamarre E, Drazba J, Prestwich G, Lindner D. Gene deletion of inositol hexakisphosphate kinase 2 predisposes to aerodigestive tract carcinoma. *Oncogene.* 2009;28(25):2383-2392.

59. Chakraborty A, Koldobskiy MA, Sixt KM, Juluri KR, Mustafa AK, Snowman AM, Van Rossum DB, Patterson RL, Snyder SH. HSP90 regulates cell survival via inositol hexakisphosphate kinase-2. *Proceedings of the National Academy of Sciences.* 2008;105(4):1134-1139.

60. Koldobskiy MA, Chakraborty A, Werner JK, Snowman AM, Juluri KR, Vandiver MS, Kim S, Heletz S, Snyder SH. p53-mediated apoptosis requires inositol hexakisphosphate kinase-2. *Proceedings of the National Academy of Sciences.* 2010;107(49):20947-20951.

61. Kalkan Z, Durasi İM, Sezerman U, Atasever-Arslan B. Potential of GRID2 receptor gene for preventing TNF-induced neurodegeneration in autism. *Neuroscience Letters.* 2016;620:62-69.

62. Ma Q, Cao C, Zhuang C, Luo X, Li X, Wan H, Ye J, Chen F, Cui L, Zhang Y, Wen Y, Yuan S, Gui Y. AXDND1, a novel testis-enriched gene, is required for spermiogenesis and male fertility. *Cell Death Discovery.* 2021;7(1):348.

63. Chakladar J, Li WT, Bouvet M, Chang EY, Wang-Rodriguez J, Ongkeko WM. Papillary Thyroid Carcinoma Variants are Characterized by Co-dysregulation of Immune and Cancer Associated Genes. *Cancers (Basel).* 2019;11(8).

64. Willinger CM, Rong J, Tanriverdi K, Courchesne PL, Huan T, Wasserman GA, Lin H, Dupuis J, Joehanes R, Jones MR, Chen G, Benjamin EJ, O'Connor GT, Mizgerd JP, Freedman JE, Larson MG, Levy D. MicroRNA Signature of Cigarette Smoking and Evidence for a Putative Causal Role of MicroRNAs in Smoking-Related Inflammation and Target Organ Damage. *Circ Cardiovasc Genet.* 2017;10(5).

65. Schembri F, Sridhar S, Perdomo C, Gustafson AM, Zhang X, Ergun A, Lu J, Liu G, Zhang X, Bowers J, Vaziri C, Ott K, Sensinger K, Collins JJ, Brody JS, Getts R, Lenburg ME, Spira A. MicroRNAs as modulators of smoking-induced gene expression changes in human airway epithelium. *Proceedings of the National Academy of Sciences.* 2009;106(7):2319-2324.

66. Lim LP, Lau NC, Garrett-Engele P, Grimson A, Schelter JM, Castle J, Bartel DP, Linsley PS, Johnson JM. Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. *Nature.* 2005;433(7027):769-773.

67. Mendell JT, Olson EN. MicroRNAs in stress signaling and human disease. *Cell.* 2012;148(6):1172-1187.