2022

Modeling premalignant lung squamous carcinoma via gene expression changes associated with EP300 knockout

BOSTON UNIVERSITY

GRADUATE SCHOOL OF ARTS AND SCIENCES

Thesis

MODELING PREMALIGNANT LUNG SQUAMOUS CARCINOMA VIA GENE EXPRESSION CHANGES ASSOCIATED WITH EP300 KNOCKOUT

by

DANY FU

B.S., Pepperdine University, 2014

Submitted in partial fulfillment of the

requirements for the degree of

Master of Science

2022

Approved by

First Reader

Sarah A. Mazzilli, PhD
Assistant Professor of Medicine

Second Reader

Jennifer E. Beane-Ebel, PhD
Assistant Professor of Medicine

Third Reader

Trevor W. Siggers, PhD
Associate Professor of Biology

Acknowledgments

First and foremost, I would like to thank my advisor Dr. Sarah Mazzilli, for welcoming me into her lab and giving me the opportunity and guidance for pursuing this research. I would also like to thank Dr. Jennifer Beane for the computational directions, and Dr. Trevor Siggers for being my program advisor and reader.

All the work in this thesis was made possible by the people in the Mazzilli lab, especially Roxanna Pfefferkorn, who patiently trained me on all the experimental protocols, Emily Green for placing orders for all the sequencing and reagents even on her days off, and Shannon McDermott and Austin Potter for their cell culture help.

I'm also grateful to the Computational Biomedicine department, particularly the staff that we share a lab space with - for their advice, instruments training, and kindness.

MODELING PREMALIGNANT LUNG SQUAMOUS CARCINOMA VIA GENE EXPRESSION CHANGES ASSOCIATED WITH EP300 KNOCKOUT

DANY FU

ABSTRACT

Lung cancer is the third most common type of cancer and the leading cause of cancer death, in both men and women, and prognosis for lung carcinoma remains poor due to late diagnosis. While lung squamous cell carcinoma (LUSC) makes up 20-30% of all lung cancer cases, identification of genetic signatures and successful targeted therapies remain limited. An ongoing effort is being made to create an in vitro system for modeling the early stages of lung squamous carcinoma and premalignancy, which will ultimately serve as a model for drug discovery. A previous effort performed whole exome and targeted DNA sequencing to reveal the somatic mutations in endobronchial biopsies that harbored lung squamous premalignant histology. *EP300* was identified as a candidate gene which may act as a driver for carcinogenesis, but remains understudied when compared to prominent oncogenic driver genes such *TP53*, *NOTCH1*, or *NFE2L2* in LUSC. The p300 protein is a histone acetyltransferase that regulates gene expression by means of chromatin remodeling and has been implicated in various diseases, including cancer. My objective as part of my thesis was to first generate stable *EP300* knockout (KO) clones from the NL20 bronchial epithelial cell line utilizing the CRISPR/Cas gene editing system. Using the NL20 clones and *EP300* KO clones in the HBEC-3KT cell line generated in a previous effort, I then validated the knockouts at the DNA, RNA, and protein levels. Literature review was also conducted to identify possible cellular pathways that *EP300* participates in

and validate its role in those pathways by observing changes in downstream protein targets. Finally, I generated RNA sequencing data from the functionally validated clones to identify differentially expressed genes and cellular pathways perturbed by *EP300* knockout. Through these efforts, I developed sets of gene signatures unique to each cell line and found that *EP300* is associated with bronchial carcinogenesis progression and likely functions as an oncogene in LUSC.

Contents

List of Tables

List of Figures

List of Abbreviations

CIS	.....	Carcinoma In Situ
CPM	.....	Counts Per Million
CPTAC	.....	Clinical Proteomic Tumor Analysis Consortium
CRISPR	.....	Clustered Regularly Interspaced Short Palindromic Repeats
DSB	.....	Double Stranded Break
EMT	.....	Epithelial-to-Mesenchymal Transition
FDR	.....	False Discovery Rate
ESCC	.....	Esophageal Squamous Carcinoma
DEG	.....	Differentially Expressed Genes
FACS	.....	Fluorescence Activated Cell Sorting
GSEA	.....	Gene Set Enrichment Analysis
GSVA	.....	Gene Set Variation Analysis
H3K27ac	.....	Histone 3 Lysine 27 Acetylation
HAT	.....	Histone Acetyltransferase
ICE	.....	Inference of CRISPR Edits
KIX	.....	Kinase-Inducible Domain of Interacting
KO	.....	Knockout
LFC	.....	Log-2 Fold Change
LUAD	.....	Lung Adenocarcinoma
LUSC	.....	Lung Squamous Cell Carcinoma
MDS	.....	Multidimensional Scaling
MSigDB	.....	Molecular Signatures Database
NSCLC	.....	Non-Small Cell Lung Cancer
P300	.....	E1A Binding Protein P300
P53	.....	Tumor Protein 53
PCGA	.....	Pre-Cancer Genome Atlas
RING	.....	Really Interesting New Gene
SCLC	.....	Small Cell Lung Cancer
TCGA	.....	The Cancer Genome Atlas
TMM	.....	Trimmed mean of M-values

Chapter 1

Introduction

1.1 Lung Cancer

Despite recent advances in lung cancer therapy and increased survival rates, lung cancer still remains the leading cause of cancer death worldwide, and more than two thirds of predicted lung cancer deaths in 2022 in the United States will be related to cigarette smoking [1, 2]. The two major types of lung cancer are small (SCLC) and non-small cell (NSCLC) lung cancer, with the latter being the dominant type making up 85% of all cases. Most of NSCLC can be subcategorized as lung adenocarcinoma (LUAD) which makes up 40% of all lung cancer, or lung squamous cell carcinoma (LUSC) (20-30%) depending on the location of origination in the lungs and cellular morphology. The five-year survival rate of those diagnosed with stage I NSCLC is 68.4% but this rate drops drastically to 5.8% for stage IV patients. However, less than one third of all NSCLC cases in the US were detected at stage I [3], driving the need for improvements in early detection and intervention.

The shift towards personalized medicine, aided by the falling cost and advances in next generation sequencing technologies, has enabled genomic profiling of cancer on a large scale. The Cancer Genome Atlas (TCGA) program was established in 2005 and has since generated over 2.5 petabytes of genomics, epigenomic, transcriptomic, and proteomic data spanning 33 cancer types [4]. Data driven research within and across cancer types has both shed light on substantially more oncogenic driver genes, tumor subtyping, and identification of genomic alterations for targeted therapeutics.

Comprehensive analysis of LUAD and LUSC using TCGA data and other sources have found that their mutational profiles are largely distinct [5, 6, 7], and as such, successful targeted therapies aimed at prominent oncogenic pathways such as EGFR and KRAS remain almost exclusively to LUAD as these mutations are rarely found in LUSC [7, 8, 9]. There remains significant motivation to discover targetable biomarkers that drive carcinogenesis in LUSC.

1.2 Premalignant Lung Squamous Cell Carcinoma

Prior to becoming invasive (ie. malignant), tumors go through a series of molecular changes in a premalignant phase. The steps in this progression are recognizable histologically and also at the genomic and epigenetic levels. The nine stages of LUSC carcinogenesis (stages 0 - 8) can be divided into four distinct molecular groups based on the gene expression profiles of each of the stages. Tissues progress from normal histology to 'low grade' lesions, to 'high grade' lesions (including carcinoma in situ), before finally becoming invasive [10]. Bronchial dysplasia lesions at each stage have the ability to progress, regress, or remain the same [11]. Although progression to malignancy becomes more likely in higher-grade histology, invasion occurs only in 5-10% of the highest grades [12], and carcinoma in situ (CIS) lesions still revert to normal tissue in 30% of cases [13]. Given the invasive nature and clinical risks associated with oncological intervention, it may be inappropriate to do so at the first detection of such lesions. There is considerable clinical potential in these earliest stages of tumor formation if patients could be advised on the likelihood of progression towards malignancy.

While Teixeira et al. created a proof of concept that successfully demonstrated the possibility of making such predictions based on biopsies from 85 patients with preinvasive LUSC lesions, a larger data set involving international cooperation would in-

variably provide a more exhaustive set of predictive signatures including less-common genomic mutations [13]. Other studies have emerged in recent years to illustrate the power of studying the earliest phases of carcinogenesis [14]. For example, an FDA approved drug for advanced metastasized breast cancer was found to also decrease premalignant lesions and therefore has the potential to be used in preventing a certain type of breast cancer [15]. Motivated by the advance in research that the TCGA data has enabled, our group has made a push for the creation of the Pre-Cancer Genome Atlas (PCGA), a database of comprehensive genomic profiles of premalignant tissues [16]. This goal will require coordinated effort from dozens of medical institutions to systematically collect and annotate clinical samples across cancer types. By following the progression of tumors that become invasive, and indirectly, those that do not, we can identify molecular markers that arise during the invasion process and develop biological models to be used for therapeutic discovery.

1.3 Current Understanding of *EP300* in Lung Cancer

The E1A Binding Protein P300 (p300), encoded by the gene *EP300*, is a histone acetyltransferase that regulates gene expression by means of chromatin remodeling via the acetylation of all four core histones, which is generally associated with increased gene transcription. It is also capable of self-acetylation and functions as a coactivator of non-histone targets such as tumor protein 53 (p53) [17, 18]. It is therefore involved in numerous cellular pathways including cell proliferation and differentiation. The 31 exons of *EP300* create a full length protein composed of several granular domains including the kinase-inducible domain interacting (KIX) domain, bromodomain, really interesting new gene (RING), and histone acetyltransferase (HAT) domain. Notably implicated in Rubinstein-Taybi syndrome, there is growing evidence that *EP300* acts as a driver mutation in lung cancer and is mutated in pre-invasive CIS lesions [13, 19].

A recent review found that *EP300* mutation occurs in 7-12% of SCLC [19], and an unpublished analysis from our group found *EP300* to be mutated in 11% of samples harboring premalignant LUSC histology (Fig 1·1). A cBioPortal [20, 21] search of *EP300* found that the gene is mutated in 3.75% (3 out of 80) of LUSC samples in the Clinical Proteomic Tumor Analysis Consortium (CPTAC) data and 5.13% (25 out of 487) of LUSC samples in the TCGA Pan-Cancer Atlas database [22, 23]. A schematic of the exons and domains of *EP300* can be see in Figure 1·2, along with the locations of mutations within the gene found in patient samples. A hot spot for *EP300* mutations occur most frequently in the HAT domain. All three in the CPTAC data are missense mutations, as are the majority of the ones in the TCGA data set.

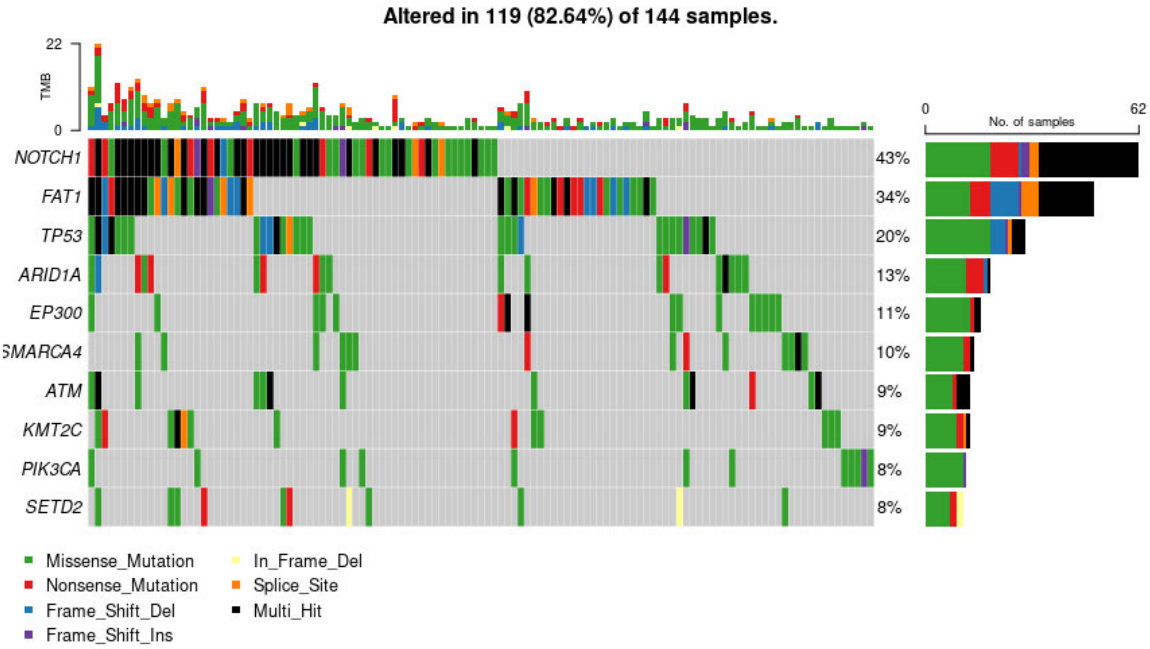


Figure 1·1: Oncoplot of top 10 altered genes from targeted exome sequencing of 144 samples from PCGA lung data. *EP300* mutations are found in 11% of samples.

There is evidence to suggest that *EP300* could be involved in either oncogenic [24, 25, 26] or tumor suppressing activities [27, 28, 29]. However, in NSCLC or related cancers, *EP300* so far has been described as a gene that drives proliferation and invasion.

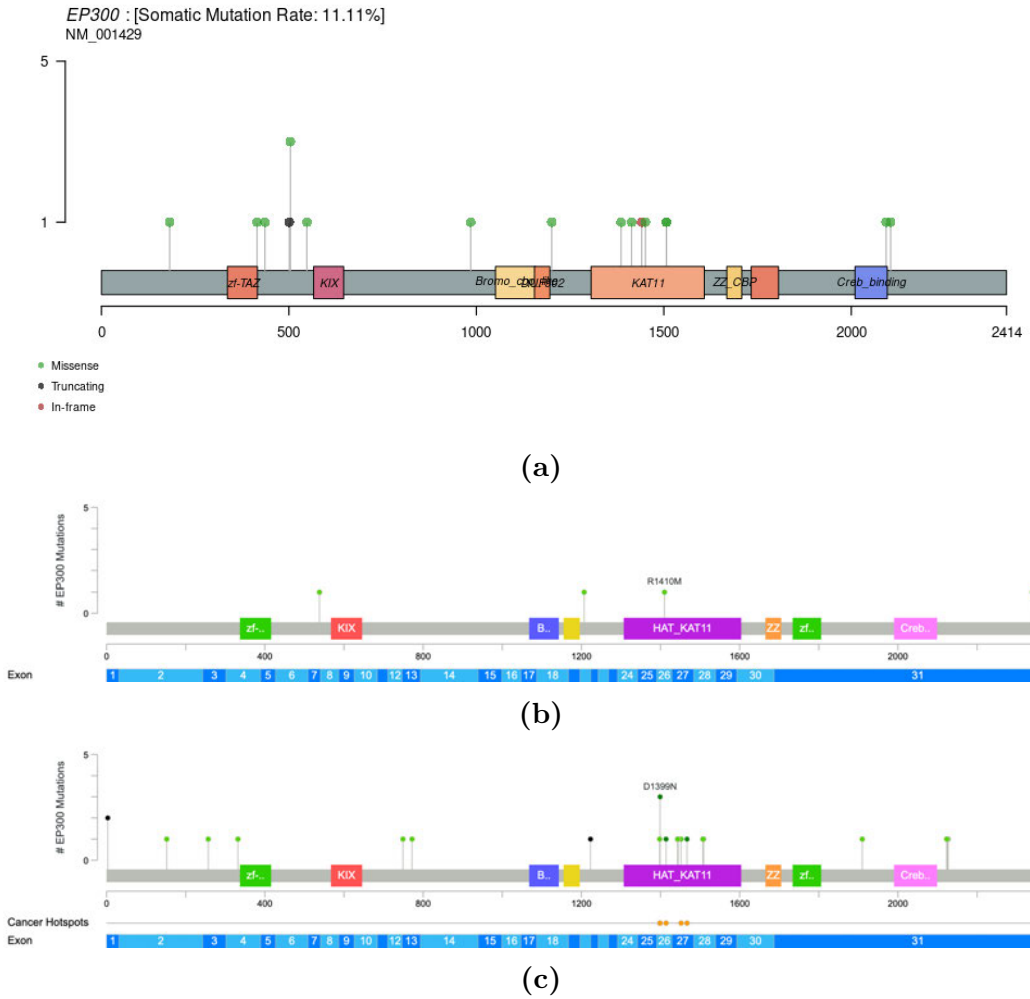


Figure 1·2: Mutation plots of *EP300* mutation types and locations in gene from a) PCGA b) CPTAC c) TCGA Pan-Cancer data sets. Most alterations are missense mutations. Figures b and c were generated by cBioPortal. Color code: missense truncating splice

Bi et al. investigated the role of *EP300* in esophageal squamous carcinoma (ESCC), another type of cancer closely associated with smoking, and found that *EP300* was not only highly mutated in ESCC patients but that a higher expression correlated with a shorter survival outcome. The authors knocked down *EP300* in two highly upregulated ESCC cell lines and found that proliferation, colony formation, migration and invasion capabilities were all negatively impacted. RNA sequencing analysis of

one of the cell lines identified differentially expressed genes in pathways involved in angiogenesis, hypoxia, and epithelial-to-mesenchymal transition (EMT) [24]. Specifically, p53 signaling pathway and nucleotide excision repair were implicated. These findings are echoed in an analogous study in NSCLC; an analysis of 169 primary patient samples (106 LUAD, 52 LUSC, 11 adenosquamous carcinoma) found that those with higher expression of *EP300* had lower survival rates [25]. The same group knocked down *EP300* in two highly expressed NSCLC cell lines, and overexpressed the gene in two cell lines that had low endogenous levels. They confirmed that the expression of *EP300* had a positive correlation with proliferation, colony formation, migration and invasion rate and that genes in the EMT pathway were also affected [26]. Although Kim et al. similarly found decreased proliferation in SCLC cell lines after *EP300* knockout, they did not see the same effect in three NSCLC cell lines [30].

Determination of the physical structure of p300 protein has increased our understanding that its mutational effects could be domain specific [17, 30, 31]. Delvecchio et al. engineered mutations into various subdomains of the full p300 protein and found that it had different effects on p300 autoacetylation and p53 acetylation: while inactivation mutations of the protein depleted autoacetylation and decreased p53 acetylation, mutations that targeted the RING domain considerably increased both [17]. Kim et al. demonstrated in vitro that knocking out at *EP300* exon 27 (start of the HAT domain) in a precancerous neuroendocrine cell-based model of SCLC development led to higher colony formation in soft agar. Using the "HAT-less" cell line, subsequent knockout of exon 16 (start of the bromodomain) showed no change in colony formation. However, knockout at both exon 9 (start of KIX domain) and exon 2 (start of the TAZ domain) significantly decreased colony formation [30].

One analysis of TCGA data linked *EP300* mutations to be associated with genome instability, which results in an increase in genomic mutations throughout the life

cycle of a cell and is one of the hallmarks of cancer. The pan-cancer analysis of 11 cancer types from TCGA, including LUAD, found significantly higher tumor mutation burden in *EP300* mutated cancers compared to *EP300* wild type. Specifically, genes involved in DNA repair such as base and nucleotide excision, mismatch repair, and damage sensor were co-mutated. The p53 signaling pathway was found to be highly enriched in *EP300* mutated pan-cancer, which the authors described as an indicator of increased DNA damage repair activity [32]. However, this upregulation could be futile given the presence of co-mutational frequency of *EP300* and DNA repair genes.

In addition to the association between p300 and p53 already described, there is substantial evidence that the two proteins interact and are coactivators of one another. The gene *TP53*, which encodes the protein p53, is of significant interest in the study of cancer as it is the most commonly mutated gene across all cancer types. It has been dubbed the "Guardian of the Genome" because of the essential role it plays in DNA repair and cell cycle regulation such as apoptosis and senescence [33]. Given the crucial role *TP53* has in either repairing mutations or eliminating unrecoverable cells, it is unsurprising that mutations in the gene itself or dysregulation of it could lead to aberrant cellular growth. Evidence of regulation of p53 by p300 was summarized over two decades ago, which suggest that p300 can stabilize p53 to prevent degradation and is necessary for transactivation [34]. Since then, the advance in 3D imaging of protein structures has further confirmed the interactions between the two proteins [35]. This naive model of the interaction between *EP300* and *TP53* seem to suggest that overexpression of EP300 could be beneficial in its role as a helper to the "Guardian", but this is in contrast to what was previously discussed in ESCC and NSCLC. Needless to say, this relationship is complex and warrants further investigation in cancer both at the genomic and epigenomic levels.

The growing evidence of *EP300*'s mutational presence across cancer types and

premalignant tumors, association with genomic instability, and relationship with tumor suppressor p53 all make the gene an attractive candidate for further investigation in tumorigenesis and targeted therapeutic development. In fact, the biotechnology company CellCentric announced in 2018 that the small molecule inhibitor the company had developed for p300/CBP was entering clinical trials for late stage prostate cancer patients. In 2019, the trial expanded to patients with hematological malignancies such as acute myeloid leukemia and certain lymphomas. However, the effects of *EP300* mutations in NSCLC and its gene signature remains unclear due to the fact that it is a less common mutation when compared to more prominent cancer genes in NSCLC such as TP53, NOTCH1, or NFE2L2 [7].

1.4 Preclinical NSCLC Models

The ultimate challenge of cancer treatment lies in efficiently targeting cancer cells: eliminate the malignant cells that proliferate uncontrollably but leave the normal healthy cells unharmed. Traditional approaches such as chemotherapy and radiation do not differentiate between the two. After identifying unique hallmarks of tumor cells in silico, the goal is to then replicate the mutational profiles in vitro and in vivo to create therapeutic approaches to specifically target those mutations. Cancer research relies on preclinical platforms such as cell lines, primary patient samples, and animal models to study disease progression and ultimately for drug discovery and validation. Cell lines are the most common NSCLC models used as they are widely available, inexpensive, can be genetically manipulated easily, and provide an unrestricted platform to study cellular behavior and toxicology in response to novel therapeutic treatments. Nevertheless, a recent review of preclinical NSCLC models pointed out several limitations on cell line models including the low number of modern LUSC models available and the uneven distribution of models across disease

progression, notably in the premalignancy space [36].

As part of a larger effort to create an in vitro system for modeling the early stages of LUSC and premalignancy, seven genes were identified in a prior effort to be commonly mutated between LUSC and premalignant lung squamous lesions including *TP53*, *NOTCH1*, *FAT1*, *FGFR1*, *NFE2L2*, *EP300*, and *PIK3CA*. Alterations in six of the seven had already been completed, so my aim was to generate at least three stable *EP300* knockout (KO) cell lines in NL20, an immortalized, nontumorigenic human bronchial epithelial cell line, utilizing the CRISPR/Cas gene editing system. *EP300* KO was also carried out in another immortalized human bronchial epithelial cell line, HBEC3-KT, in a previous effort and were simultaneously validated alongside the NL20 clones at the DNA, RNA, and protein levels in order to identify the clones with a functional deletion. My second aim was to generate RNA sequencing data and create gene signatures unique to *EP300* KO, and validate whether the same signature can be associated with existing data from clinical samples.

Chapter 2

Methods

2.1 Cell Culture

NL20 cells were cultured in F12K Medium (ATCC) and supplemented according to ATCC instructions. HBEC-3KT cells were cultured in Airway Epithelial Cell Basal Medium (ATCC) supplemented with the Bronchial Epithelial Cell Growth Kit (ATCC). All cells were grown in a humidified incubator at 37°C and 5% CO2.

2.2 CRISPR/Cas9 Transfection

Wild type NL20 cells were transfected following Synthego's "CRISPR Editing of Immortalized Cell Lines with RNPs Using Lipofection" protocol. *EP300* CRISPR gRNAs were ordered from Synthego (guide 1 sequence: CTAGAAGTCATTCCTGGTTG, guide 2 sequence: GGAACCATGCCTGCAGCATT, guide 3 sequence: AGATTTTCTCAGCTAGAAGG). Transfection conditions were optimized for NL20 cells using the TRAC gene following Synthego's "Transfection Optimization Kit (Multi-guide)" protocol. RNP was complexed in a 8:1 gRNA to Cas9 ratio for *EP300* knockout and 0:1 ratio for control, and incubated with 2 x 104 cells per well. Media was replaced on day 2 (one day after transfection). DNA was extracted on day 5 for knockout efficiency analysis. Cells were sorted into single cell colonies on day 9.

2.3 Fluorescence Activated Cell Sorting (FACS)

Cells were detached and resuspended in 500 µL DPBS and 1uL of 5mM Calcein Blue (Invitrogen) for a final concentration of 10 µM, followed by an incubation period of 30 minutes at room temperature. The cells were centrifuged, washed and resuspended in FACS buffer (2% FBS and 2mM EDTA in PBS). Live cells were sorted from dead and single live cells were placed in 500uL of media per well in 96-well plates.

2.4 Cell Proliferation Assay

Three replicates of 4 x 10^4 cells were seeded per well with 1mL of medium in 24-well plates. Cells were collected each day for 4 days and counted with Nexcelom Cellometer Auto 1000 Cell Viability Counter.

2.5 PCR & Gel Electrophoresis

DNA was isolated from cell pellets using QuickExtract DNA Extraction Solution (Lucigen). EP300 PCR was performed using Kapa HIFI plus dNTPs (Roche) with an annealing temperature of 65°C for 30 cycles. EP300 primers were ordered from IDT (forward sequence: AGTAGCGACTTAACTGTTGTTC; reverse sequence: CAAGCCAGTCTC AGAGAAATC). DNA was visualized in 90% agarose gel and 5% ethidium bromide.

2.6 DNA Sequencing & ICE Analysis

Unpurified PCR products were sent to Azenta for Sanger sequencing. Results were uploaded to Synthego's Inference of CRISPR Edits (ICE) [37] analysis tool (https://ice.synthego.com) in order to analyze knockout efficiency.

2.7 RNA Extraction & RT-QPCR

Total RNA was extracted from cell pellets using 1mL of Qiazol followed by 250 µL of chloroform. Upper aqueous phase was incubated with 100% ethanol at either -80°C for 1 hour over 4°C overnight. The solution was centrifuged for 45 minutes at 20,000xg at 4°C. Supernatant was discarded and the pellet washed with 100% ethanol three times. RNA was reverse transcribed to cDNA using the Transcriptor First Strand cDNA Synthesis Kit (Roche). *EP300* Hs00914204_m1 (ThermoFisher, FAM-MGB) was used to target the region between exons 10 and 11. All experiments were done in triplicates with *PPIA* primer Hs04194521_s1 (ThermoFisher,FAM-MGB) as the endogenous control gene.

2.8 Western Blot

Protein was extracted with NE-PER™ Nuclear and Cytoplasmic Extraction Reagents (ThermoFisher Scientific). Detection of the EP300 protein was performed on 3-8% tris acetate gel (ThermoFisher Scientific), 1:500 p300 Monoclonal Antibody (Invitrogen, RW128), 1:2000 anti-vinculin control (Abcam, ab91459), and 1:1000 HRP conjugated anti-mouse secondary (Cell Signaling Technologies, 7076). Detection of H3K27Ac was performed on 4-12% bis tris gel (ThermoFisher Scientific, NW04120), 1:200 Anti-Histone H3 (acetyl K27) (Abcam, ab4729), 1:1000 beta actin control (Santa Cruz Biotechnology, 47778), and 1:500 HRP conjugated anti-rabbit secondary (Cell Signaling Technologies, 7074). Membranes were incubated for 5 minutes in SuperSignal™ West Pico PLUS Chemiluminescent Substrate (ThermoFisher Scientific) for chemiluminescent imaging, followed by 10 minute incubation in Pierce™ DAB Substrate Kit (ThermoFisher Scientific) for colorimetric detection.

2.9 RNA Sequencing & Analysis

Total RNA was extracted from all cell lines with miRNeasy Kit (Qiagen). RNA quality control and sequencing was completed at Azenta. Samples were prepared using the NEBNext Ultra II RNA Library Preparation Kit (Illumina) and sequenced on the Illumina HiSeq platform to generate paired-end 150 base pair reads. Sequencing reads were aligned to reference genome hg38 using 2-pass STAR [38] alignment. Gene and transcript level counts were calculated using RSEM [39] using Ensembl v100 annotation. Trimmed mean of M-values (TMM) scaling factors were generated using EdgeR [40] and TMM-normalized log-2 counts per million (CPM) values were used for figures containing boxplots, multidimensional scaling analysis, and heatmaps. Linear modeling and differential expression analysis were performed on gene-level counts after voom-transform with limma [41] utilizing the Benjamini-Hochberg Procedure [42] for false discovery rate (FDR) adjustment. The fGSEA R package [43] and the hallmark gene set from the Molecular Signatures Database (MSigDB) [44] was used to carry out Gene Set Enrichment Analysis [45]. Datasets used for Gene Set Variation Analysis (GSVA) [46] can be found in NCBI Gene Expression Omnibus under accession codes GSE109743 and GSE109743. Three biological replicates per clone was used for all bioinformatics analysis.

Chapter 3

Results

3.1 *EP300* Knockout and Validation

CRISPR mediated targeting of *EP300* at exon 10 (KIX domain) was carried out in parental NL20 cells, and the initial transfection achieved an editing efficiency of 100% according to Inference of CRISPR Edits (ICE) analysis [37] (Fig 3·1). The knockout cell pool underwent fluorescence activated cell sorting to isolate single-cell clones; of the 23 clones that proliferated from the initial cell sorting, two were lost due to contamination (clones 6 and 16), and six were eventually lost after varying (2-4) passage numbers due to stoppage of growth (clones 3, 8, 12, 19, 21, and 22). In the latter group, although a few attached cells were still present after several weeks, they did not divide and grow into robust colonies, suggesting possible cellular senescence associated with the loss of *EP300*.

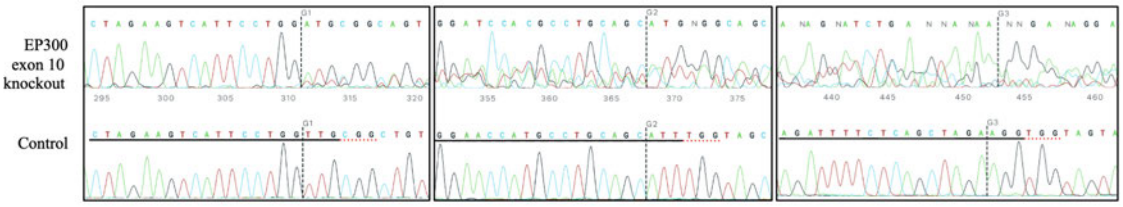
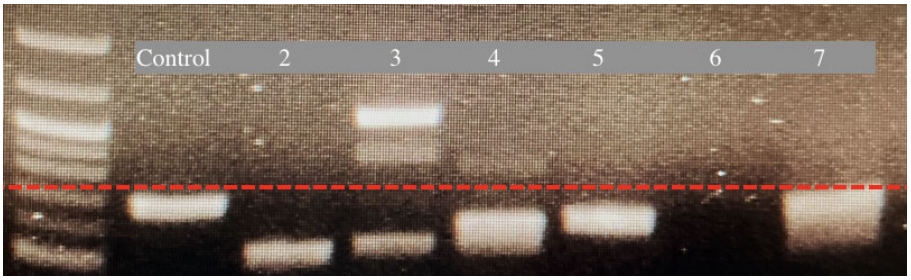
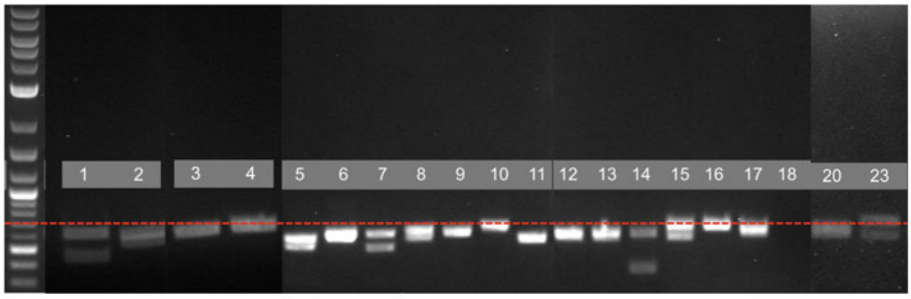


Figure 3·1: Sanger chromatograms showing the target sites of the three gRNAs of the pooled EP300 KO population versus the control. Horizontal black line indicates the sequences of the gRNA and the dotted red line indicates the PAM sequences. The vertical dotted line represents where the edit occurred. Chromatograms were generated from Synthego's ICE analysis tool.



(a)

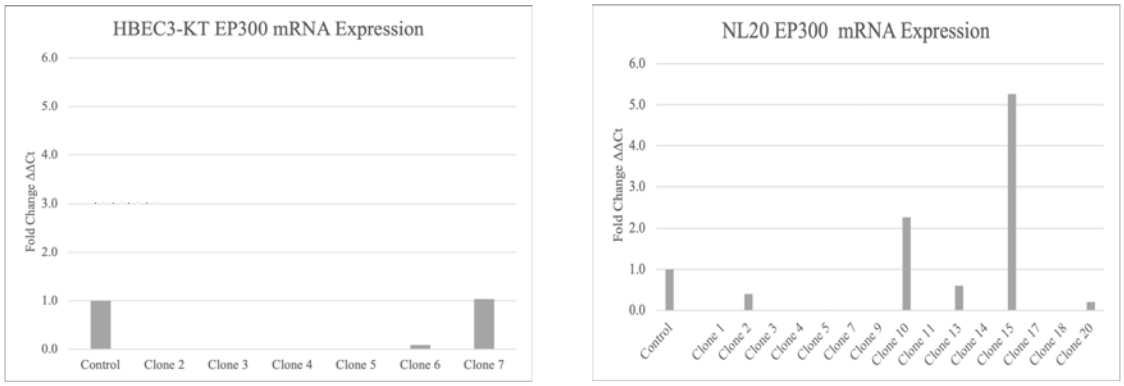


(b)

Figure 3·2: Gel electrophoresis images of a) HBEC3-KT and b) NL20 *EP300* KO cell lines . Red dotted line represents the expected length of the wild type. NL20 clone 18 and HBEC3-KT clone 6 did not amplify after multiple attempts.

Validation at the DNA level was carried out by PCR with primers flanking the deleted region (Fig 3·2). Notably, NL20 clone 18 and HBEC3-KT clone 6 did not amplify after multiple attempts. It is possible that the primer binding sites were deleted, or DNA rearrangement occurred post Cas9 cleavage. Different sets of primers, possibly further out from the CRISPR target site, could be used to confirm genomic editing. More than one band was present in some of the clones suggesting a heterozygous deletion. Sanger sequencing of the PCR products and subsequent ICE analysis gave a knockout score of greater than 90% in all except NL20 clones 2 and 10, and a knockout score of 100 in HBEC3-KT clones 2, 3, 4, and 5, signifying a high likelihood of the editing to result in a functional knockout. RNA expression was measured using RT-qPCR with a primer specifically targeting the exon boundaries between 10 and

11 and fold change was calculated with double delta Ct analysis (Fig 3·3).



(a) **(b)**

Figure 3·3: Analysis of knockdown efficiency of a) HBEC3-KT and b) NL20 *EP300* KO cell lines via mRNA expression by RT-qPCR. A primer targeting the unction between exons 10 and 11 was used.

The efficiency of the knockout was further confirmed at the protein level by western blot on both the cell lines (Fig 3·4a, 3·4b). All these results together suggest that NL20 clones 1, 11, 18, and 20, and HBEC3-KT clones 2, 5, and 6 achieved functional knockout of the *EP300* gene. The four NL20 clones underwent further validation via examination of the level of histone 3 lysine 27 acetylation (H3K27ac) via western blot. Consistent with previous findings [31], all four clones showed reduced H3K27ac compared to the parental line (Fig 3·4c), further confirming a functionally successful CRISPR mediated knockout of *EP300*. H3K27 acetylation level changes could not be confidently determined in the HBEC3-KT clones due to the low endogenous level observed in the control cells. Alternate targets such as p53 or p300 acetylation could be investigated. Proliferation rate of the four NL20 clones was also recorded as a preliminary investigation of its phenotypic profile (Fig 3·5), which agreed with some but not all previous findings [24, 26, 30] .

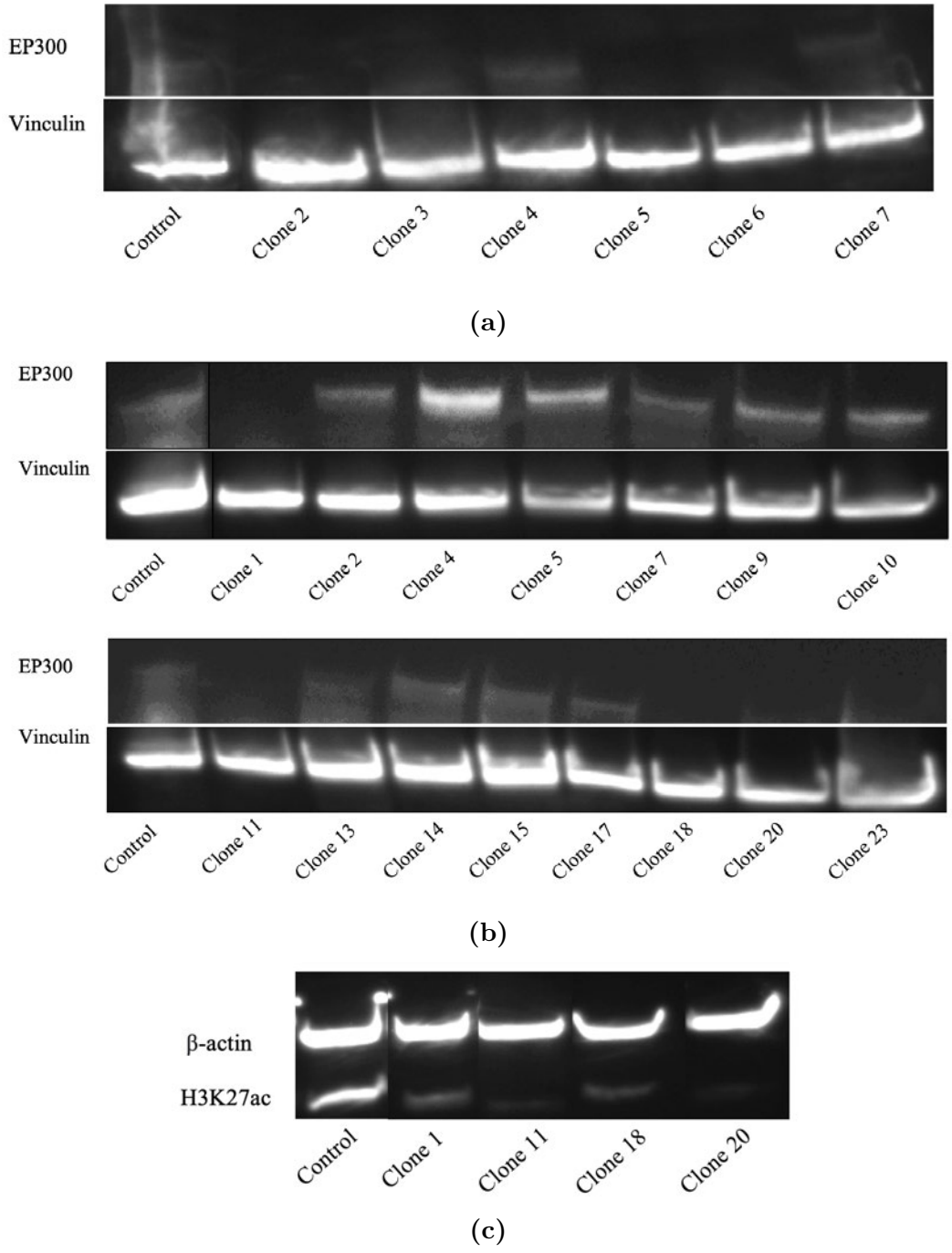


Figure 3·4: Analysis of knockdown efficiency of a) HBEC3-KT and b) NL20 *EP300* KO cell lines via protein expression by western blots. c) Western blot of H3K27 acetylation in NL20 clones that exhibited the strongest knockdown of *EP300*

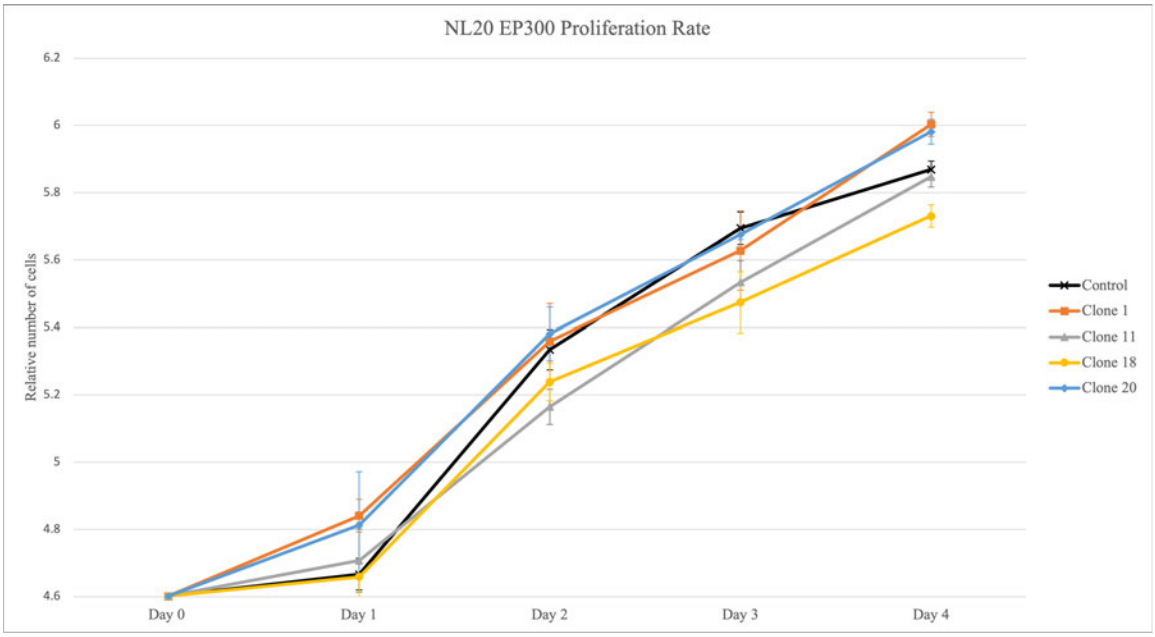


Figure 3·5: Cell growth curve of the four NL20 clones that exhibited the strongest evidence of a functional knockout compared to the control. Error bars represent standard deviation.

3.2 *EP300* RNA Sequencing Analysis

Biological triplicates of each of the NL20 clones 11, 18, and 20 and HBEC3-KT clones 4, 5, and 6 were selected for bulk next generation RNA sequencing. The perturbed clones were analyzed against control cells in their respective cell lines to identify differentially expressed genes (DEGs) in order to understand the effect that knocking out *EP300* has on genome-wide RNA expression in lung cells and to create a unique gene signature. *EP300* expression was specifically evaluated as a form of initial quality control and confirmed knockdown of the gene in all CRISPR/Cas9 edited samples compared to its respective controls (Fig 3·6a, 3·6b). Although *EP300* mRNA levels were not completely eliminated in any of the clones, all knockdowns were statistically significant and were used as a benchmark for setting false discovery rate (FDR) adjusted p-values and log2-fold change (LFC) cutoffs to determine DEGs. Given the relationship between EP300 and TP53 previously discussed, *TP53*

levels were also evaluated (Fig 3·6c, 3·6d) and while expression did decrease in the NL20 clones, *EP300* knockdown level did not have a direct relationship with *TP53* expression level. *TP53* levels were not significantly altered in HBEC3-KT clones.

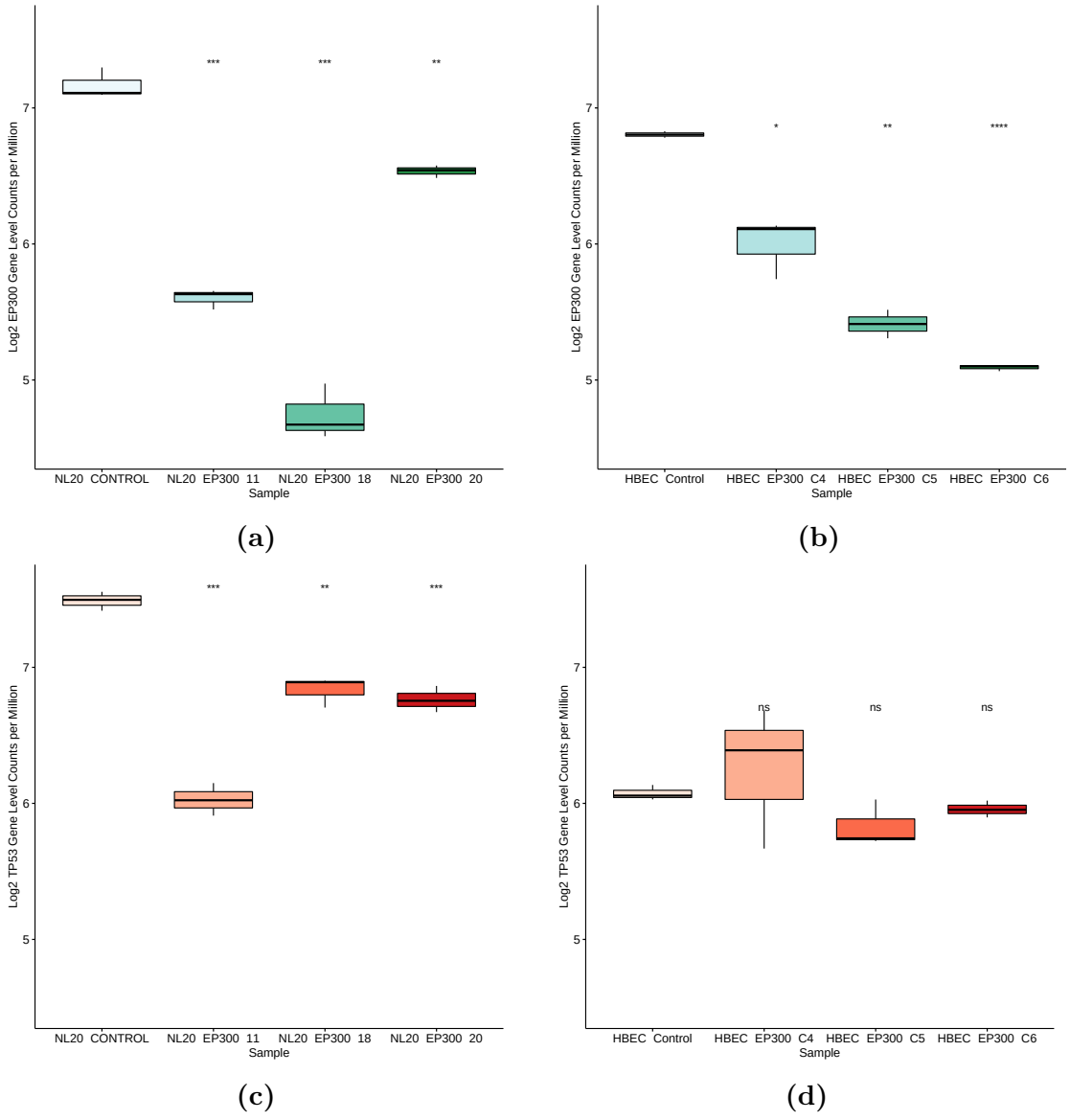
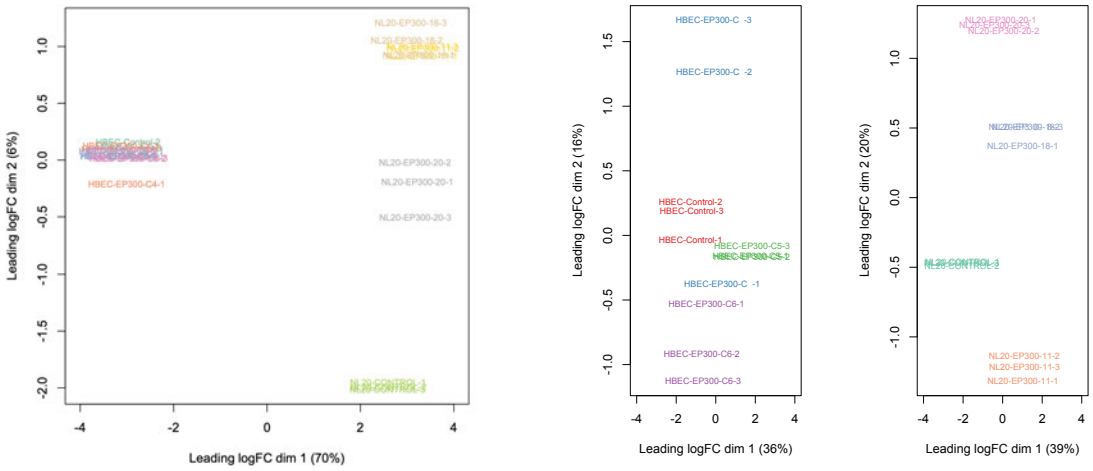


Figure 3·6: Boxplots showing NL20 RNA expression levels of a) *EP300* b) *TP53* and HBEC3-KT RNA expression levels of c) *EP300* d) *TP53*. *TP53* expression did not have a directly correlation with *EP300* expression levels. Significance was determined by Student's t-test. *** $p \leq 0.001$, ** $p \leq 0.01$, * $p \leq 0.05$, ns=not significant.



(a) **(b)** **(c)**

Figure 3·7: Multidimensional scaling plots of a) All *EP300* KO samples, b) HBEC3-KT only, c) NL20 only. Samples clustered closer together indicate more expression similarity than those further apart. Expression variance is best described by the first dimension (x axis) than by the second (y axis).

Exploratory examination of gene expression was carried out by multidimensional scaling (MDS) and revealed that the greatest expression variance came from inherent biological differences between the two cell lines (Fig 3·7a). Unexpectedly, the HBEC3-KT clones clustered closely to its controls while the NL20 samples did not. This signaled that the NL20 clones will have stronger differential gene expression when compared to its control than the HBEC3-KT clones. Nevertheless, MDS analysis of the cell lines separately indicated that expression differences can still be discerned in the HBEC3-KT line if analyzed on its own (Fig 3·7b). Therefore, all downstream analysis was carried out separately between the two cell lines. Interestingly, the HBEC3-KT clones showed greater internal variance amongst biological replicates than NL20 cells, particularly clone 4 (Fig 3·7b, 3·7c). ICE analysis of clone 4's PCR product showed four separate fragments that contributed to more than 10% of the overall sequence pool (23%, 21%, 15%, 10%). Under perfect PCR and sequencing

conditions, only one fragment contributing to 100% of the sequencing would be expected for a homozygous deletion, and two fragments (50%/ 50%) for a heterozygous deletion. Given the mixture of fragments detected and their percentage of contribution, it is possible that clone 4 is not a homogenous population. This may have occurred due to reasons such as failure to isolate a single clone during initial sorting or spontaneous mutation during initial proliferation.

Table 3.1: Number of differentially expressed genes of individual *EP300* KO clones. The two clones with the lowest knockdown rate (NL20 clone 20 and HBEC3-KT clone 4) has the least number of DEGs, but the clones with the highest knockdown rate (NL20 clone 18 and HBEC3-KT clone 6) did not always have the most number of DEGs.

FDR <0.01, LFC>1.5	
NL20 Clone	# of DEGs
11	1860
18	1869
20	861

FDR <0.05, LFC>0.6	
HBEC3-KT Clone	# of DEGs
4	860
5	3701
6	1906

As suggested by preliminary MDS analysis (Fig 3·7a), NL20 *EP300* KO clones yielded more DEGs than its HBEC3-KT counterparts even with more stringent FDR and LFC cutoffs (Fig 3·8, 3·9). Using FDR <0.01 and LFC>1.5, 560 DEGs were revealed in the NL20 clones: 479 genes were downregulated and 81 genes were upregulated. In HBEC3-KT clones, only 82 DEGs were found using FDR <0.05 and LFC>0.6. Again, more genes were found to have reduced expression than increased, 59 and 23, respectively. *TP53* was not found in either of the gene signature sets. Each clone was also analyzed individually as a quantitative measure of variance from the control (Table 3.1). Expectedly, NL20 clone 20 and HBEC3-KT clone 4, which had the lowest knockdown rate in its respective cell lines, demonstrated the least deviation of gene expression from the controls. NL20 clones 11 and 18 had no significant differences in number of DEGs, but more surprisingly, HBEC3-KT clone 5 produced

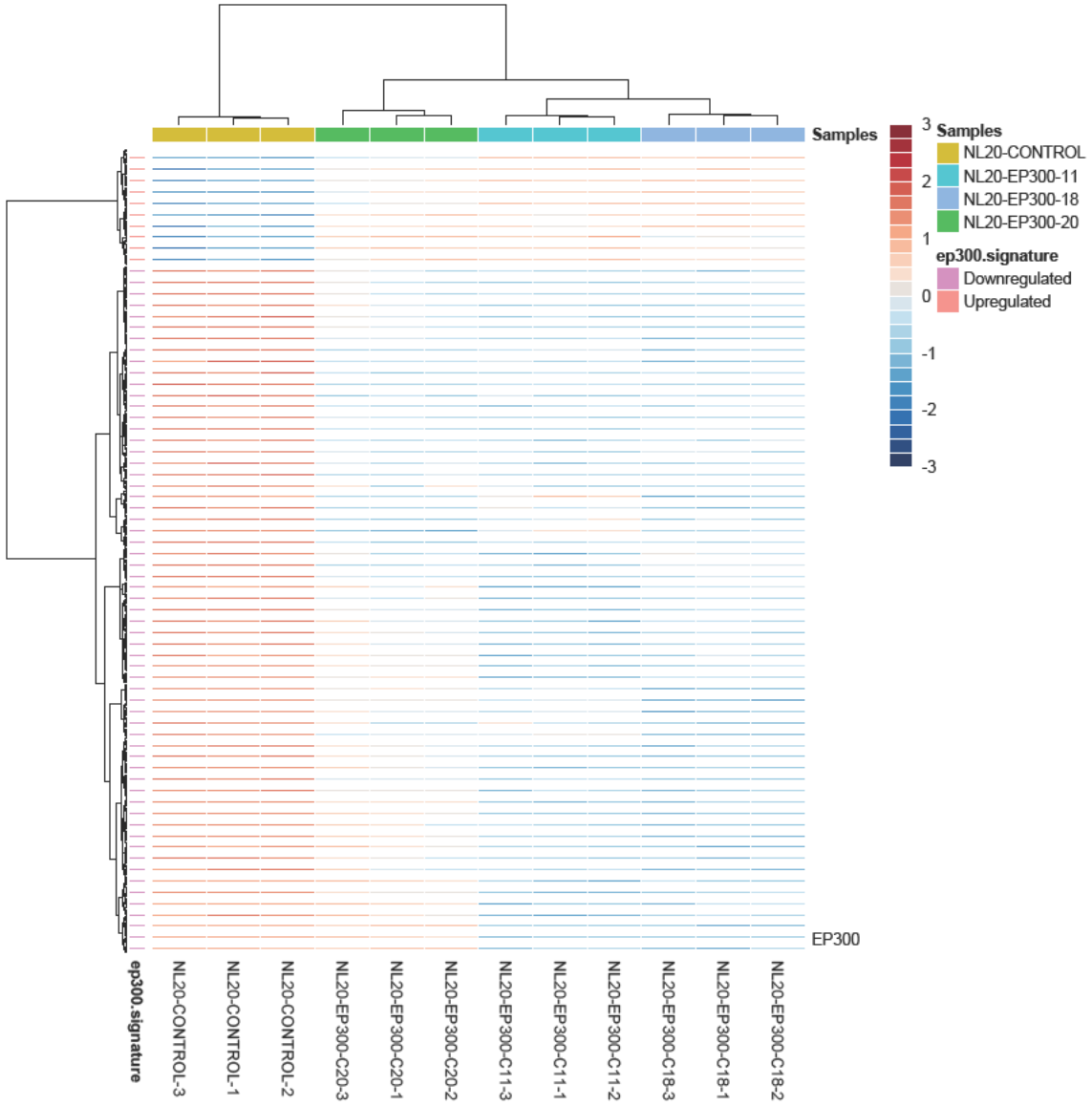


Figure 3·8: Heatmap of the 560 highly differentially expressed genes in NL20 EP300 KO clones compared to its control (FDR<0.01, LFC>1.5). 479 genes are downregulated, and 81 genes are upregulated. Each column represents one sample and each row one differential gene. Ward's minimum variance method was used for clustering.

more DEGs than clone 6 despite the latter having a higher *EP300* knockdown rate. It is possible that *CREBBP*, a functional paralog, is upregulated in order to compensate for the loss of function of *EP300*. Therefore, *CREBBP* was specifically investigated

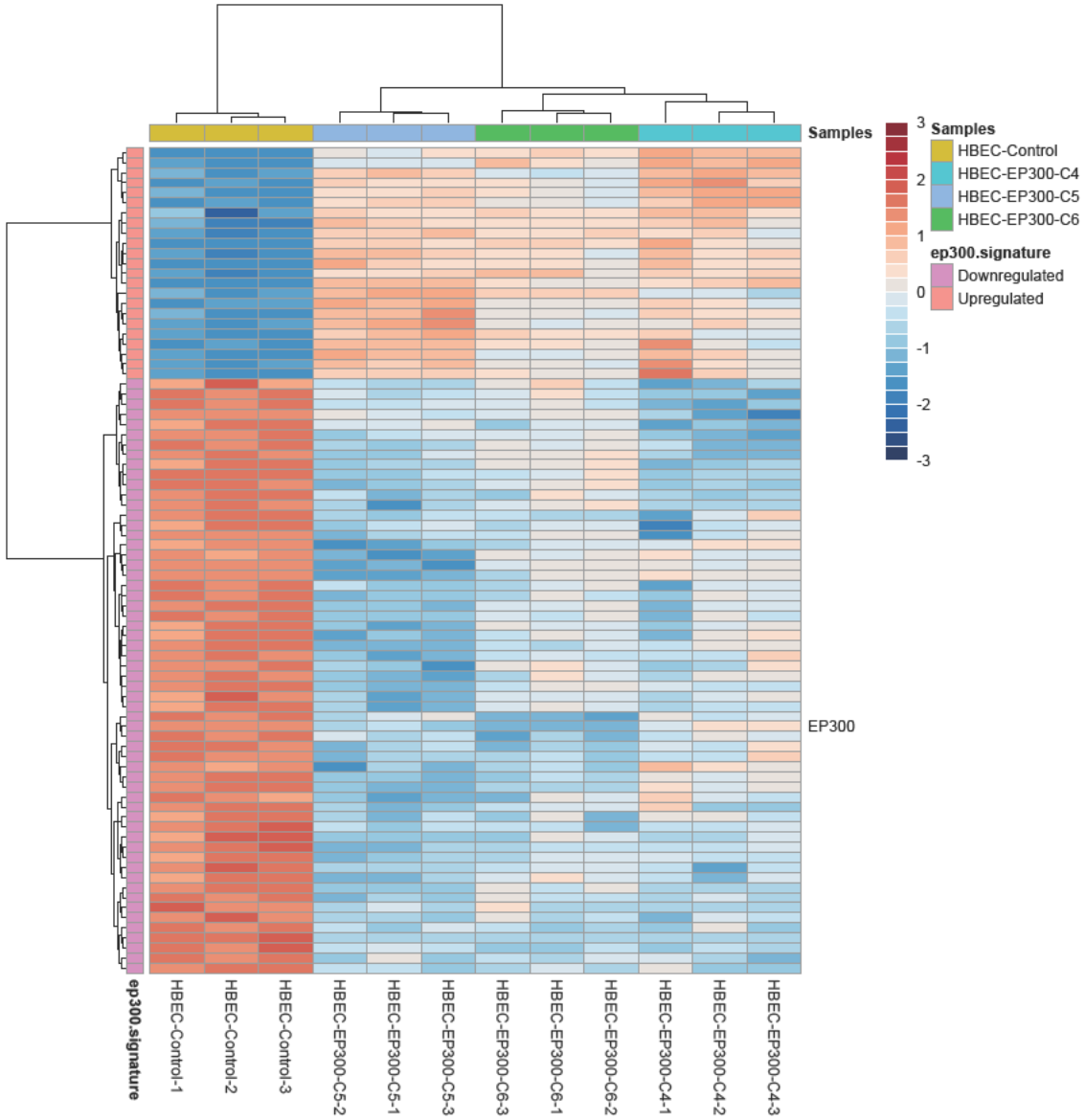


Figure 3·9: Heatmap of the 82 differentially expressed genes in HBEC3-KT EP300 KO clones (FDR<0.05, LFC>0.6). 59 genes are downregulated, and 23 genes are upregulated. Each column represents one sample and each row one differential gene. Ward's minimum variance method was used for clustering.

and revealed that its RNA expression level is upregulated in NL20 clone 11, even more so than clone 18, which could explain why their variance profile are similar despite clone 18 having a stronger *EP300* knockdown (Fig 3·10a). However, this

explanation did not extend to HBEC3-KT clones, as none of the *CREBBP* expression levels changed significantly from the control (Fig 3·10b). Theoretically, a complete knockout should have taken place given that the CRISPR target site is near the front of the gene at exon 10, but considering the stochastic nature of DSB repair and the complex functionality of the different subunits of the p300 protein, it is difficult to speculate without more insight to the genomic or proteomic alteration that took place. Sanger sequencing and ICE analysis revealed a 58 base pair deletion in clone 5, which could have rendered the protein to be still active but carry out a reprogrammed function. Similar to a dominant negative effect, the mutated p300 protein could lead to more aberrant behavior in a cell than a complete knockout. The possibility of CRISPR off-target effects also cannot be completely ruled out in clone 5, which could exaggerate the mutational profile independent of *EP300* levels. DNA perturbation in clone 6 could not be determined due to unsuccessful PCR amplification (Fig 3·2a).

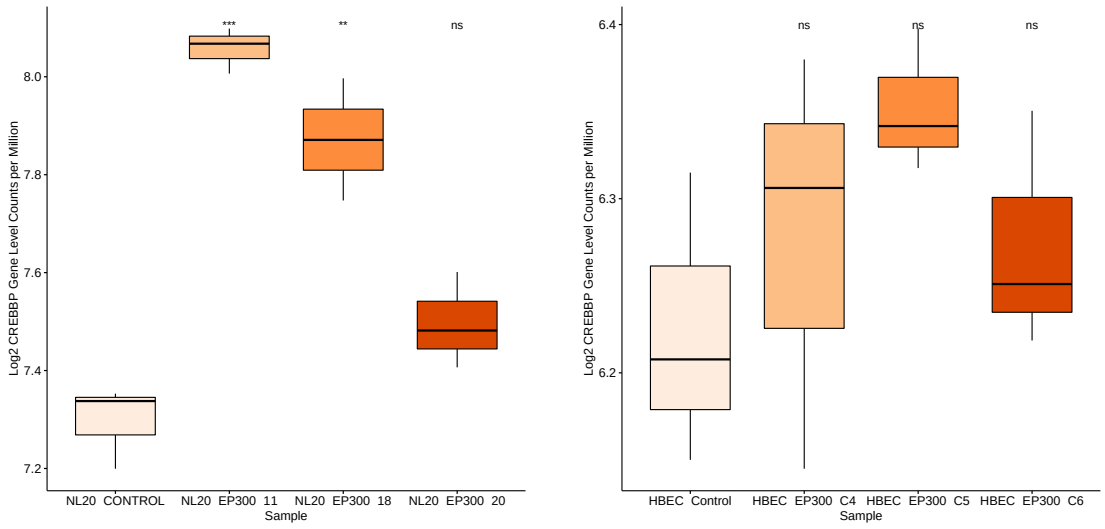


Figure 3·10: Boxplots showing *CREBBP*, a functional paralog of *EP300*, RNA expression levels in a) NL20 b) HBEC3-KT. Significance was determined by Student's t-test. *** $p \leq 0.001$, ** $p \leq 0.01$, * $p \leq 0.05$, ns=not significant.

Pathway enrichment results with FDR<0.5 using Gene Set Enrichment Analysis (GSEA) [45] and the Hallmark gene set from the Molecular Signatures Database (MSigDB) [44] are shown in Figure 3·11. Only two negatively enriched pathways

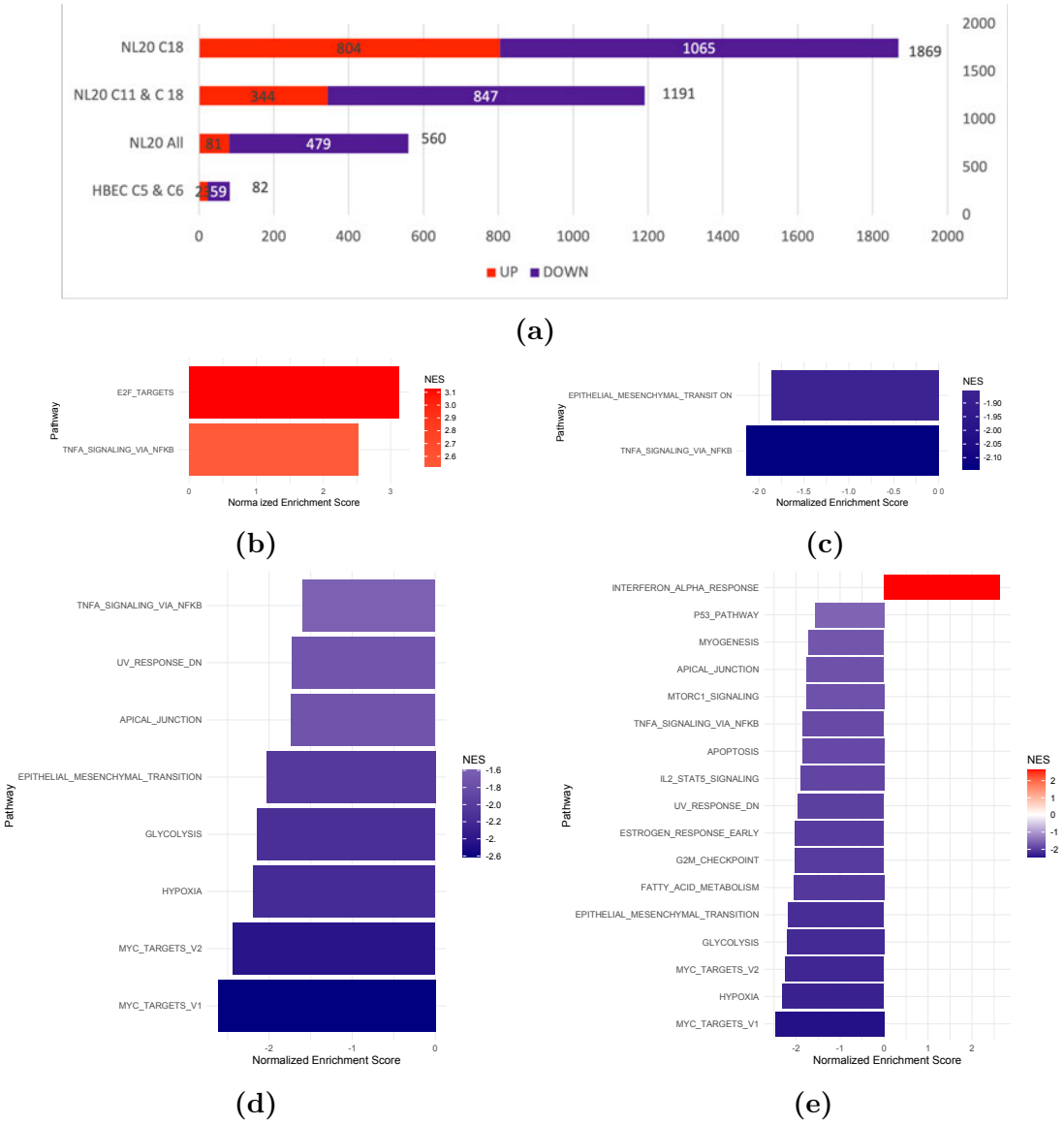


Figure 3·11: Gene Set Enrichment Analysis. a) Number of DEGs used for pathway analysis. b) HBEC3-KT clones 5 and 6 using DEGs with FDR<0.05, LFC>0.6 only revealed upregulated pathways. NL20 pathways revealed using DEGs with FDR <0.01, LFC>1.5 c) All NL20 clones d) NL20 clones 11 and 18 and e) clone 18 only. Gene sets are from the MSigDB Hallmark collection. All pathways have FDR<0.5.

emerged when using the 560 NL20 DEGs from all three clones, including the EMT pathway which echoes results from previous publications [24, 26]. To increase the number of potential pathways involved, the weakest *EP300* KO clone (ie. clone 20) was removed and DEGs were re-identified using only clones 11 and 18. Using the same FDR and LFC cutoffs as before, eight negatively enriched pathways emerged from 1191 DEGs when NL20 clone 20 was excluded. Negatively enriched pathways increased to 16 when clone 18 was analyzed on its own with 1869 DEGs, including hypoxia and the p53 signaling pathway, further confirming previous research efforts [24, 26, 32]. Additionally, the prominent oncogene *MYC* and the set of genes that it regulates are shown to be strongly negatively enriched when clone 20 was excluded from GSEA. The downregulation of these cancer pathways suggest that *EP300* likely operates as an oncogene in lung cancer. The same analysis using HBEC3-KT DEGs in Figure 3·9 did not yield any enriched pathways, but two positively enriched pathways emerged when the clone with the weakest knockdown (clone 4) was excluded (Fig 3·11b), including the TNFA Signaling Via NFKB pathway which is prominently negatively enriched in NL20. The other pathway, E2F Targets, is also notable for its role in cancer and did not manifest in the NL20 clones even when FDR and LFC cutoffs were lowered to the same as HBEC3-KTs. The contrasting GSEA results indicates that *EP300* perturbation manifests differently between the two cell lines.

To investigate the clinical significance of the *EP300* KO gene signature, DEGs used in GSEA in Figures 3·11e and 3·11b were further examined using Gene Set Variation Analysis (GSVA) [46] against human endobronchial lesion biopsies from two separate studies [12, 47]. Guided by the results of the pathway analysis, a subset of the genes (847 downregulated genes from the NL20 signature, 173 upregulated genes from HBEC3-KT signature) were evaluated against PCGA mRNA (GSE109743) and Merrick et al.'s microarray (GSE109743) data sets. Genes that were not common to

both the gene signature and the data set were removed prior to GSVA score calculation and are shown grouped by histology in Figure 3·12. Results from the PCGA data set show that the set of downregulated genes found in the NL20 *EP300* KO have a positive association with increased histology (P=$2.55e^{-}9$). In the HBEC3-KT KO clones, the set of upregulated genes also increase in expression with higher histology (P=$4.47e^{-}11$). Although the results were not always significant in the Merrick data set, the trend supports that seen in the PCGA data set (NL20 P=0.717, HBEC3-KT P=0.0185). These trends suggest that the *EP300* KO gene signature is associated with premalignant lesion progression in LUSC.

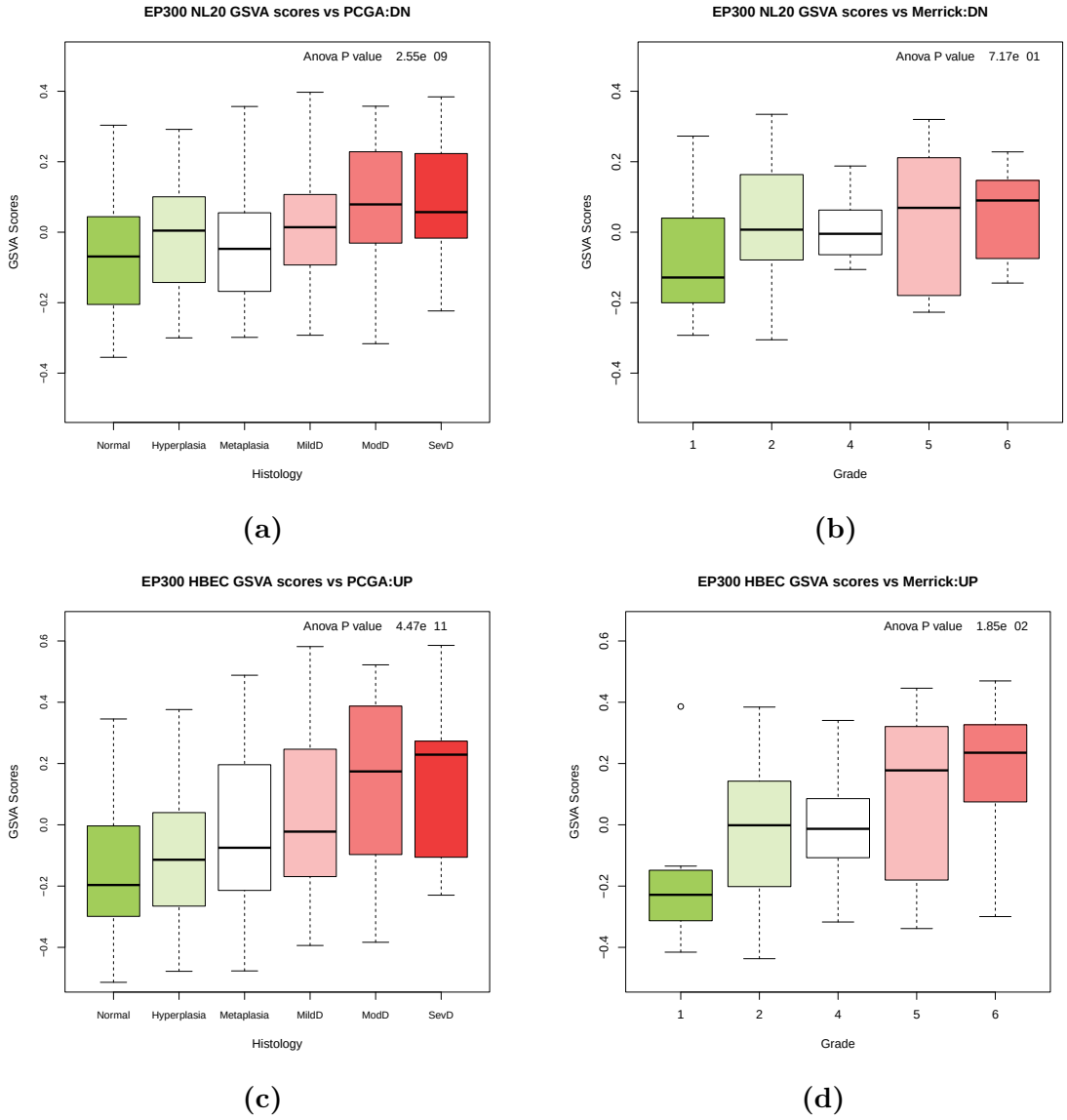


Figure 3·12: Boxplots showing GSVA scores of selective EP300 gene signature during premalignant LUSC disease progression. a-b) Down-regulated genes found in NL20 clones 11 and 18 show increase in expression as disease progresses towards malignancy in PCGA and Merrick et al. data sets c-d)Upregulated genes found in HBEC3-KT clones 5 and 6 show increase in expression with higher histology in PCGA and Merrick et al. data set.

Chapter 4

Discussion

Motivated by the need for early detection of lung cancer and lack of premalignant in vitro models in the LUSC space, our group has generated seven cell lines with single gene perturbations identified to be significantly altered in clinical LUSC and premalignant samples, including EP300 knockout clones as demonstrated in this thesis. The knockout was functionally validated at the DNA, RNA, and protein levels, and in the case of the NL20 clones, also by observing variations in downstream acetylation levels. RNA sequencing data generated from selected clones identified differentially expressed genes in both NL20 and HBEC3-KT cell lines. Pathway enrichment analysis was also carried out using DEGs from clones that exhibited the strongest EP300 knockdown and confirmed pathways found in previous studies as well as notable pathways implicated in cancer. Finally, gene signatures were evaluated against bronchial dysplasia samples from two separate cohorts and revealed a positive association between *EP300* perturbed genes and tumor histological progression.

MDS analysis, DEG discovery, and pathway analysis all indicate that the HBEC3-KT clones had weaker overall expression variance from its controls than the NL20 clones, despite *EP300* knockdown levels being similar and significant. Inherent biological variability such as method of immortalization could account for some of the differences. The NL20 cell line was established via transfection of a viral oncoprotein, SV40 large T plasmid, which could compromise p53 functions. In fact, this cell line did spontaneously become tumorigenic at one time [48]. HBEC3-KT, on the other

hand, was immortalized while retaining wild-type p53 functions [49], which could influence it to be more resistant to oncogenic mutations in comparison to NL20. Another question this raises is whether the *TP53* gene expression level and p53 pathway alteration seen in Figures 3·6c and 3·11e, respectively, were an artifact of the cell line or a true effect of *EP300* KO. Creating *EP300* perturbations in other human bronchial epithelial cell lines could be carried out to further investigate this finding.

As previously mentioned, several studies have suggested that depletion of *EP300* results in decreased cell proliferation. Although the four functional NL20 KO clones did not show a significant change in proliferation as compared to the parental line, the six clones that were eventually lost suggest a tentative link between the loss of *EP300* and cellular senescence. Cellular senescence possibly evolved as a mechanism for tumor suppression, as the definition of cancer is unchecked cellular growth. Stressors including DNA damage (eg. double stranded break (DSB)), oncogene activation, or chromatin perturbation could induce a senescent response. Moreover, *TP53* is also a well known regulator of senescence [50], and given the evidence that p300 is a coactivator of p53, it is plausible that mutations in *EP300* could trigger p53 induced senescence. Phenotypic changes of a senescent cell includes arrested growth and resistance to apoptosis, both of which were observed in clones 3, 8, 12, 19, 21, and 22. Gene expression changes or emblematic biomarkers such as senescence-associated-β-galactosidase were not investigated as part of this thesis. Furthermore, additional investigation would be necessary to determine whether CRISPR-induced editing and subsequent DSB may have been the cause of senescence rather than the mutation in *EP300*.

Bibliography

[1] Rebecca L. Siegel, Kimberly D. Miller, Hannah E. Fuchs, and Ahmedin Jemal. Cancer statistics, 2022. *CA: A Cancer Journal for Clinicians*, 72(1):7–33, January 2022.

[2] Hyuna Sung, Jacques Ferlay, Rebecca L. Siegel, Mathieu Laversanne, Isabelle Soerjomataram, Ahmedin Jemal, and Freddie Bray. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians*, 71(3):209–249, February 2021.

[3] Apar Kishor Ganti, Alyssa B. Klein, Ion Cotarla, Brian Seal, and Engels Chou. Update of incidence, prevalence, survival, and initial treatment in patients with non–small cell lung cancer in the US. *JAMA Oncology*, 7(12):1824, December 2021.

[4] The Cancer Genome Atlas Program. `https://www.cancer.gov/about-nci/organization/ccg/research/structural-genomics/tcga`. Accessed: 2022-06-24.

[5] Joshua D Campbell, , Anton Alexandrov, Jaegil Kim, Jeremiah Wala, Alice H Berger, Chandra Sekhar Pedamallu, Sachet A Shukla, Guangwu Guo, Angela N Brooks, Bradley A Murray, Marcin Imielinski, Xin Hu, Shiyun Ling, Rehan Akbani, Mara Rosenberg, Carrie Cibulskis, Aruna Ramachandran, Eric A Collisson, David J Kwiatkowski, Michael S Lawrence, John N Weinstein, Roel G W Verhaak, Catherine J Wu, Peter S Hammerman, Andrew D Cherniack, Gad Getz, Maxim N Artyomov, Robert Schreiber, Ramaswamy Govindan, and Matthew Meyerson. Distinct patterns of somatic genome alterations in lung adenocarcinomas and squamous cell carcinomas. *Nature Genetics*, 48(6):607–616, May 2016.

[6] The Cancer Genome Atlas Research Network. Comprehensive molecular profiling of lung adenocarcinoma. *Nature*, 511(7511):543–550, July 2014.

[7] The Cancer Genome Atlas Research Network. Comprehensive genomic characterization of squamous cell lung cancers. *Nature*, 489(7417):519–525, September 2012.

[8] Min Yuan, Li-Li Huang, Jian-Hua Chen, Jie Wu, and Qing Xu. The emerging treatment landscape of targeted therapy in non-small-cell lung cancer. *Signal Transduction and Targeted Therapy*, 4(1), December 2019.

[9] Ruqin Chen, Rami Manochakian, Lauren James, Abdel-Ghani Azzouqa, Huashan Shi, Yan Zhang, Yujie Zhao, Kexun Zhou, and Yanyan Lou. Emerging therapeutic agents for advanced non-small cell lung cancer. *Journal of Hematology & Oncology*, 13(1), May 2020.

[10] Céline Mascaux, Mihaela Angelova, Angela Vasaturo, Jennifer Beane, Kahkeshan Hijazi, Geraldine Anthoine, Bénédicte Buttard, Françoise Rothe, Karen Willard-Gallo, Annick Haller, Vincent Ninane, Arsène Burny, Jean-Paul Sculier, Avi Spira, and Jérôme Galon. Immune evasion before tumour invasion in early lung squamous carcinogenesis. *Nature*, 571(7766):570–575, June 2019.

[11] Evgeny V. Denisov, Anastasia A. Schegoleva, Polina A. Gervas, Anastasia A. Ponomaryova, Lubov A. Tashireva, Valentina V. Boyarko, Ekaterina B. Bukreeva, Olga V. Pankova, and Vladimir M. Perelmuter. Premalignant lesions of squamous cell carcinoma of the lung: The molecular make-up and factors affecting their progression. *Lung Cancer*, 135:21–28, September 2019.

[12] Daniel T. Merrick, Michael G. Edwards, Wilbur A. Franklin, Michio Sugita, Robert L. Keith, York E. Miller, Micah B. Friedman, Lori D. Dwyer-Nield, Meredith A. Tennis, Mary C. O'Keefe, Elizabeth J. Donald, Jessica M. Malloy, Adrie van Bokhoven, Storey Wilson, Peter J. Koch, Charlene O'Shea, Christopher Coldren, David J. Orlicky, Xian Lu, Anna E. Baron, Greg Hickey, Timothy C. Kennedy, Roger Powell, Lynn Heasley, Paul A. Bunn, Mark Geraci, and Raphael A. Nemenoff. Altered cell-cycle control, inflammation, and adhesion in high-risk persistent bronchial dysplasia. *Cancer Research*, 78(17):4971–4983, September 2018.

[13] Vitor H. Teixeira, Christodoulos P. Pipinikas, Adam Pennycuick, Henry Lee-Six, Deepak Chandrasekharan, Jennifer Beane, Tiffany J. Morris, Anna Karpathakis, Andrew Feber, Charles E. Breeze, Paschalis Ntolios, Robert E. Hynds, Mary Falzon, Arrigo Capitanio, Bernadette Carroll, Pascal F. Durrenberger, Georgia Hardavella, James M. Brown, Andy G. Lynch, Henry Farmery, Dirk S. Paul, Rachel C. Chambers, Nicholas McGranahan, Neal Navani, Ricky M. Thakrar, Charles Swanton, Stephan Beck, Phillip Jeremy George, Avrum Spira, Peter J. Campbell, Christina Thirlwell, and Sam M. Janes. Deciphering the genomic, epigenomic, and transcriptomic landscapes of pre-invasive lung cancer lesions. *Nature Medicine*, 25(3):517–525, January 2019.

[14] Bríd M. Ryan and Jessica M. Faupel-Badger. The hallmarks of premalignant conditions: a molecular basis for cancer prevention. *Seminars in Oncology*, 43(1):22–35, February 2016.

[15] Tracy E. Strecker, Qiang Shen, Yun Zhang, Jamal L. Hill, Yuxin Li, Chunyu Wang, Hee-Tae Kim, Tona M. Gilmer, Krystal R. Sexton, Susan G. Hilsenbeck, C. Kent Osborne, and Powel H. Brown. Effect of lapatinib on the development of estrogen receptor–negative mammary tumors in mice. *JNCI: Journal of the National Cancer Institute*, 101(2):107–113, January 2009.

[16] Joshua D. Campbell, Sarah A. Mazzilli, Mary E. Reid, Samjot S. Dhillon, Suso Platero, Jennifer Beane, and Avrum E. Spira. The case for a pre-cancer genome atlas (PCGA). *Cancer Prevention Research*, 9(2):119–124, February 2016.

[17] Manuela Delvecchio, Jonathan Gaucher, Carmen Aguilar-Gurrieri, Esther Ortega, and Daniel Panne. Structure of the p300 catalytic core and implications for chromatin targeting and HAT regulation. *Nature Structural & Molecular Biology*, 20(9):1040–1046, August 2013.

[18] Stephanie Kaypee, Smitha Asoka Sahadevan, Shilpa Patil, Piya Ghosh, Neeladri Sekhar Roy, Siddhartha Roy, and Tapas K. Kundu. Mutant and wild-type tumor suppressor p53 induces p300 autoacetylation. *iScience*, 4:260–272, June 2018.

[19] Kee-Beom Kim, Colin T. Dunn, and Kwon-Sik Park. Recent progress in mapping the emerging landscape of the small-cell lung cancer genome. *Experimental & Molecular Medicine*, 51(12):1–13, December 2019.

[20] Ethan Cerami, Jianjiong Gao, Ugur Dogrusoz, Benjamin E. Gross, Selcuk Onur Sumer, Bülent Arman Aksoy, Anders Jacobsen, Caitlin J. Byrne, Michael L. Heuer, Erik Larsson, Yevgeniy Antipin, Boris Reva, Arthur P. Goldberg, Chris Sander, and Nikolaus Schultz. The cBio cancer genomics portal: An open platform for exploring multidimensional cancer genomics data. *Cancer Discovery*, 2(5):401–404, May 2012.

[21] Jianjiong Gao, Bülent Arman Aksoy, Ugur Dogrusoz, Gideon Dresdner, Benjamin Gross, S. Onur Sumer, Yichao Sun, Anders Jacobsen, Rileen Sinha, Erik Larsson, Ethan Cerami, Chris Sander, and Nikolaus Schultz. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Science Signaling*, 6(269), April 2013.

[22] Clinical Proteomic Tumor Analysis Consortium. A proteogenomic portrait of lung squamous cell carcinoma. *Cell*, 184(16):4348–4371.e40, August 2021.

[23] John N Weinstein, Eric A Collisson, Gordon B Mills, Kenna R Mills Shaw, Brad A Ozenberger, Kyle Ellrott, Ilya Shmulevich, Chris Sander, Joshua M Stuart, and Cancer Genome Atlas Research Network. The cancer genome atlas pan-cancer analysis project. *Nature Genetics*, 45(10):1113–1120, September 2013.

[24] Yanghui Bi, Pengzhou Kong, Ling Zhang, Heyang Cui, Xiaoqin Xu, Feiyun Chang, Ting Yan, Jiayi Li, Caixia Cheng, Bin Song, Xia Niu, Xiangchen Liu, Xue Liu, Enwei Xu, Xiaoling Hu, Yu Qian, Fang Wang, Hongyi Li, Yanchun Ma, Jian Yang, Yiqian Liu, Yuanfang Zhai, Yi Wang, Yingchun Zhang, Haiyan Liu, Jing Liu, Jintao Wang, Yongping Cui, and Xiaolong Cheng. EP300 as an oncogene correlates with poor prognosis in esophageal squamous carcinoma. *Journal of Cancer*, 10(22):5413–5426, 2019.

[25] X. Hou, Y. Li, R.-Z. Luo, J.-H. Fu, J.-H. He, L.-J. Zhang, and H.-X. Yang. High expression of the transcriptional co-activator p300 predicts poor survival in resectable non-small cell lung cancers. *European Journal of Surgical Oncology (EJSO)*, 38(6):523–530, June 2012.

[26] Xue Hou, Run Gong, Jianhua Zhan, Ting Zhou, Yuxiang Ma, Yuanyuan Zhao, Yaxiong Zhang, Gang Chen, Zhonghan Zhang, Shuxiang Ma, Xi Chen, Fangfang Gao, Shaodong Hong, Fan Luo, Wenfeng Fang, Yunpeng Yang, Yan Huang, Likun Chen, Haoxian Yang, and Li Zhang. p300 promotes proliferation, migration, and invasion via inducing epithelial-mesenchymal transition in non-small cell lung cancer cells. *BMC Cancer*, 18(1), June 2018.

[27] Simon A. Gayther, Sarah J. Batley, Lori Linger, Andy Bannister, Karen Thorpe, Suet-Feung Chin, Yataro Daigo, Paul Russell, Annie Wilson, Heidi M. Sowter, Joy D.A. Delhanty, Bruce A.J. Ponder, Tony Kouzarides, and Carlos Caldas. Mutations truncating the EP300 acetylase in human cancers. *Nature Genetics*, 24(3):300–303, March 2000.

[28] Naoki Koshiishi, Ja-Mun Chong, Tomoki Fukasawa, Rie Ikeno, Yukiko Hayashi, Nobuaki Funata, Hideo Nagai, Michiko Miyaki, Yoshiro Matsumoto, and Masashi Fukayama. p300 gene alterations in intestinal and diffuse types of gastric carcinoma. *Gastric Cancer*, 7(2), June 2004.

[29] Muhammad Asaduzzaman, Stephanie Constantinou, Haoxiang Min, John Gallon, Meng-Lay Lin, Poonam Singh, Selina Raguz, Simak Ali, Sami Shousha, R. Charles Coombes, Eric W.-F. Lam, Yunhui Hu, and Ernesto Yagüe. Tumour suppressor EP300, a modulator of paclitaxel resistance and stemness, is downregulated in metaplastic breast cancer. *Breast Cancer Research and Treatment*, 163(3):461–474, March 2017.

[30] Kee-Beom Kim, Ashish Kabra, Dong-Wook Kim, Yongming Xue, Yuanjian Huang, Pei-Chi Hou, Yunpeng Zhou, Leilani J. Miranda, Jae-Il Park, Xiaobing Shi, Timothy P. Bender, John H. Bushweller, and Kwon-Sik Park. KIX domain determines a selective tumor-promoting role for EP300 and its vulnerability in small cell lung cancer. *Science Advances*, 8(7), February 2022.

[31] Yao-Hui Huang, Kun Cai, Peng-Peng Xu, Li Wang, Chuan-Xin Huang, Ying Fang, Shu Cheng, Xiao-Jian Sun, Feng Liu, Jin-Yan Huang, Meng-Meng Ji, and Wei-Li Zhao. CREBBP/EP300 mutations promoted tumor progression in diffuse large b-cell lymphoma through altering tumor-associated macrophage polarization via FBXW7-NOTCH-CCL2/CSF1 axis. *Signal Transduction and Targeted Therapy*, 6(1), January 2021.

[32] Zuobing Chen, Canping Chen, Lin Li, Tianfang Zhang, and Xiaosheng Wang. Pan-cancer analysis reveals that e1a binding protein p300 mutations increase genome instability and antitumor immunity. *Frontiers in Cell and Developmental Biology*, 9, September 2021.

[33] Wasim Feroz and Arwah Mohammad Ali Sheikh. Exploring the multiple roles of guardian of the genome: P53. *Egyptian Journal of Medical Human Genetics*, 21(1), November 2020.

[34] Steven R. Grossman. p300/CBP/p53 interaction and regulation of the p53 response. *European Journal of Biochemistry*, 268(10):2773–2778, May 2001.

[35] Raka Ghosh, Stephanie Kaypee, Manidip Shasmal, Tapas K. Kundu, Siddhartha Roy, and Jayati Sengupta. Tumor suppressor p53-mediated structural reorganization of the transcriptional coactivator p300. *Biochemistry*, 58(32):3434–3443, July 2019.

[36] Robert E. Hynds, Kristopher K. Frese, David R. Pearce, Eva Grönroos, Caroline Dive, and Charles Swanton. Progress towards non-small-cell lung cancer models that represent clinical evolutionary trajectories. *Open Biology*, 11(1):200247, January 2021.

[37] David Conant, Tim Hsiau, Nicholas Rossi, Jennifer Oki, Travis Maures, Kelsey Waite, Joyce Yang, Sahil Joshi, Reed Kelso, Kevin Holden, Brittany L. Enzmann, and Rich Stoner. Inference of CRISPR edits from sanger trace data. *The CRISPR Journal*, 5(1):123–130, February 2022.

[38] Alexander Dobin, Carrie A. Davis, Felix Schlesinger, Jorg Drenkow, Chris Zaleski, Sonali Jha, Philippe Batut, Mark Chaisson, and Thomas R. Gingeras. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, 29(1):15–21, October 2012.

[39] Bo Li and Colin N Dewey. RSEM: accurate transcript quantification from RNA-seq data with or without a reference genome. *BMC Bioinformatics*, 12(1), August 2011.

[40] M. D. Robinson, D. J. McCarthy, and G. K. Smyth. edgeR: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, 26(1):139–140, November 2009.

[41] Matthew E. Ritchie, Belinda Phipson, Di Wu, Yifang Hu, Charity W. Law, Wei Shi, and Gordon K. Smyth. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research*, 43(7):e47–e47, January 2015.

[42] Yoav Benjamini and Yosef Hochberg. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, 57(1):289–300, January 1995.

[43] Gennady Korotkevich, Vladimir Sukhov, Nikolay Budin, Boris Shpak, Maxim N. Artyomov, and Alexey Sergushichev. Fast gene set enrichment analysis. June 2016.

[44] Arthur Liberzon, Chet Birger, Helga Thorvaldsdóttir, Mahmoud Ghandi, Jill P. Mesirov, and Pablo Tamayo. The molecular signatures database hallmark gene set collection. *Cell Systems*, 1(6):417–425, December 2015.

[45] Aravind Subramanian, Pablo Tamayo, Vamsi K. Mootha, Sayan Mukherjee, Benjamin L. Ebert, Michael A. Gillette, Amanda Paulovich, Scott L. Pomeroy, Todd R. Golub, Eric S. Lander, and Jill P. Mesirov. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences*, 102(43):15545–15550, September 2005.

[46] Sonja Hänzelmann, Robert Castelo, and Justin Guinney. GSVA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinformatics*, 14(1), January 2013.

[47] Jennifer E. Beane, Sarah A. Mazzilli, Joshua D. Campbell, Grant Duclos, Kostyantyn Krysan, Christopher Moy, Catalina Perdomo, Michael Schaffer, Gang Liu, Sherry Zhang, Hanqiao Liu, Jessica Vick, Samjot S. Dhillon, Suso J. Platero, Steven M. Dubinett, Christopher Stevenson, Mary E. Reid, Marc E. Lenburg, and Avrum E. Spira. Molecular subtyping reveals immune alterations associated with progression of bronchial premalignant lesions. *Nature Communications*, 10(1), April 2019.

[48] Joan H. Schiller, Gerard Bittner, Shi-Qi Wu, and Lorraine Meisner. Karyotypic changes associated with spontaneous acquisition and loss of tumorigenicity in a human transformed bronchial epithelial cell line: Evidence for in vivo selection of transformed clones. *In Vitro Cellular & Developmental Biology - Animal*, 34(4):283–289, April 1998.

[49] Mitsuo Sato, Melville B. Vaughan, Luc Girard, Michael Peyton, Woochang Lee, David S. Shames, Ruben D. Ramirez, Noriaki Sunaga, Adi F. Gazdar, Jerry W. Shay, and John D. Minna. Multiple oncogenic changes (k-rasv12, p53 knockdown,

mutant egfrs, p16 bypass, telomerase) are not sufficient to confer a full malignant phenotype on human bronchial epithelial cells. *Cancer Research*, 66(4):2116–2128, February 2006.

[50] Judith Campisi and Fabrizio d'Adda di Fagagna. Cellular senescence: when bad things happen to good cells. *Nature Reviews Molecular Cell Biology*, 8(9):729–740, September 2007.

CURRICULUM VITAE