THE ROLE OF LIM DOMAIN AND LIM DOMAIN BINDING PROTEINS IN

HUMAN HEAD AND NECK CARCINOMA

By

Elizabeth A. Simonik

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Cancer Biology

December, 2016

Nashville, Tennessee

Utpal Dave, M.D. (Chair)

Stephen J. Brandt, M.D.

Alexander Zaika, Ph.D.

Bryan Venters, Ph.D.

ORIGINAL PUBLICATION

1.   Simonik, E., Cai, Y., Kimmelshue, K., Brantley-Sieders, D. M., Loomans, H.,
      Andl, C., Westlake, G., Youngblood, V. M., Chen, J., Yarbrough, W. D.,
      Brown, B. T., Nagarajan, L., Brandt, S. J., (2016). *LIM-Only Protein 4
      (LMO4) and LIM-Domain Binding Protein 1 (LDB1) Promote the Growth and
      Metastasis of Human Head and Neck Cancer.* PLOS One (Under Review).

To my beloved husband, John

and

To my parents, Tom and Denise

and

To my brother and sister-in-law, TJ and Jessie

for their endless encouragement and infinite support

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF EQUATIONS

LIST OF ABBREVIATIONS

.bai – BAM index file

.gff – General feature format, simple tab-delimited text file

AAALAC – Association for Assessment and Accreditation of Laboratory Animal Care

ARNT – Aryl Hydrocarbon Receptor Nuclear Translocator

B-ACTIN – One of six different actin isoforms

BAM – Binary Alignment/Map

BED – Tab-delimited text file

bHLH – Basic helix-loop-helix

BSA – Bovine serum albumin

BsmBI – Restriction endonuclease

BQ – Betel quid

C2H2 – Cys2-His2 zinc finger protein

CDC27 – Cell division cycle protein 27 homolog

CDC27 – Cell division cycle protein 27 homolog

CDH11 – Cadherin-11

ChIP – Chromatin immunoprecipitation

ChIP-exo – Chromatin immunoprecipitation; exonuclease degradation; sequencing

ChIP-seq – Chromatin immunoprecipitation sequencing

CKAP2L – Cytoskeleton Associated Protein 2 Like

CCI – Collective cell invasion

CCM – Collective cell migration

CRISPR/Cas9 – Gene editing system

CSCs – Cancer stem cells

cwpair.gff – file containing peak pairs and midpoint peak values

DAPI – 4',6-diamidino-2-phenylindole, fluorescent marker

DD – Dimerization domain

DMEM – Dulbecco's Modified Eagle Medium

dNTPs – Deoxynucleotide solution mix, dATP, dTTP, dCTP, dGTP

E-box – Enhancer box, sequence CANNTG

E-CADHERIN – Epithelial Cadherin

EBV – Epstein-Barr virus

ECM – Extracellular matrix

EDTA – Ethylenediaminetetraacetic acid

EGFR – Epidermal growth factor receptor

EMT – Epithelial-mesenchymal transition

ENV – Envelope protein

fasta.fa – Output sequence file

FBS – Fetal bovine serum

FOXC2 – homeobox transcription factor

FOXD1 – Forkhead Box D1

GAG/POL – Polyprotein Group Antigens/reverse transcriptase

GAPDH – Glyceraldehyde 3-phosphate dehydrogenase

gRNA – Guide RNA

H3K27ac – Histone H3 acetyl Lys27, activating mark

HEK 293t – Human embryonic kidney cell line

HEPES – 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid

HG19 – Human genome reference used in UCSC Genome Browser

HMGA2 – High Mobility Group AT-Hook 2

HNSCC – head and neck squamous cell carcinoma

HPV – Human papillomavirus

HPV-16 – Human papillomavirus type 16

IgG – Immunoglobulin G

IGV – Integrative Genomics Viewer

IHC – Immunohistochemistry

IMDM – Iscove's Modified Dulbecco's Medium

KO – Knockout

LCCD – LDB1/ChIP conserved domain

LDB1 – LIM domain-binding protein 1

LHX – LIM-homeobox

LID – LIM domain

LMO1 – LIM-only protein 1

LMO2 – LIM-only protein 2

LMO3 – LIM-only protein 3

LMO4 – LIM-only protein 4

MEF2A – Myocyte Enhancer Factor 2A

MIR3648 – MicroRNA 3648

MIR3687 – MicroRNA 3687

miRNAs – MicroRNAs

MMP-3 – Matrix metalloproteinase 3

MMP-9 – Matrix metalloproteinase 9

NCI – National Cancer Institute

NEB – New England Biolabs

NOTCH – Type 1 transmembrane protein family

OCC – Oral cavity carcinoma

OSCC – Oral squamous cell carcinoma

Opti-MEM – Reduced serum media

p16INK4A – principal member of the Ink4 family of CDK inhibitors

P53 – Tumor protein 53

PBS – Phosphate buffered saline

PCNA – Proliferating cell nuclear antigen

PCR – Polymerase chain reaction

Phi-29 – DNA polymerase

PI3K-Akt – intracellular signaling pathway important in regulating the cell cycle

pSPAX2 – Lentiviral packaging plasmid

pVSVG – Retroviral vector

Q5 LM-PCR – Ligation mediated polymerase chain reaction

RB – Retinoblastoma protein

RecJf – single-stranded DNA specific exonuclease

RIPA – Radioimmunoprecipitation assay

RLIM – RING finger LIM domain-binding protein; RNF12

RNA-seq – RNA next generation sequencing

RT-qPCR – Real time quantitative polymerase chain reaction

RUNX1 – Runt Related Transcription Factor 1

SAM – Sequence Alignment/Map

SDS – Sodium dodecyl sulfate

SEM – Standard error of the mean

SIX1 – homeobox transcription factor

SNAIL-2 – Zinc finger protein 2; SNAI2; SLUG

SSBP2 – Single-stranded DNA binding protein 2

SSBP3 –Single-stranded DNA binding protein 3

STAMP – Alignment, Similarity, & Database Matching for DNA Motifs

T-cell – Type of lymphocyte produced by the thymus

T4 PNK – T4 Polynucleotide Kinase

TGF-B – Transforming growth factor beta

TMA – Tissue microarray

TNM – Tumor, node, metastasis

TPSR – Vanderbilt Core: Translational Pathology Shared Resource

Tris – Tris(hydroxymethyl)aminomethane

TSS – Transcriptional start site

TWIST – Basic helix-loop-helix transcription factor

UCSC ENCODE – The Encyclopedia of DNA Elements

VU-1729-1:2 – CRISPR/cas9-mediated LDB1 knockdown clone 1:2

VU-1729-2:7 – CRISPR/cas9-mediated LDB1 LDB1 knockdown clone 2:7

VU-1729-P – Parental 1729 cell line, squamous cell carcinoma

vWF – Von Willebron Factor

Wnt – Proteins in the int/wingless family; 'wingless-related integration site'

XL1 – Competent E. coli strain

CHAPTER I

INTRODUCTION

Human head and neck carcinoma statistics

Worldwide, head and neck carcinoma is the sixth leading cancer by incidence and is responsible for greater than 550,000 new cases and 300,000 deaths each year (Jemal *et al.*, 2011). Within the United States, head and neck carcinoma accounts for three percent of all malignancies, and this year alone, it is estimated that approximately 60,000 new patients will develop head and neck carcinoma and an estimated 12,300 patients will die from the disease (NCI, 2016).

While ninety percent of all head and neck cancers are squamous cell carcinomas (HNSCC), the disease can be further subcategorized based on the region of origin. There are four major subcategories: 1) the oral cavity, which includes the anterior tongue, lips, roof and floor of mouth (bucosa); 2) the oropharynx, which includes the base of tongue, pharynx, surrounding area; 3) the nasopharynx, which includes all of the nasal sinuses; and 4) the larynx (see Figure 1.1). The focus of this body of work is on head and neck carcinomas originating from the oral cavity and oropharyngeal subcategories, as the majority (approximately seventy-seven percent) of all head and neck carcinomas are derived from those regions (NCI, 2016). Oral and oropharyngeal cancer diagnoses are twice as high in men as women, with an estimated 32,670 men and 13,110 women being diagnosed this year (NCI, 2016). Approximately seventy percent (6,010 men and 2,640 women) of patients that die each year from HNSCC will have disease derived from the

1

Head and Neck Cancer Regions

Paranasal sinuses

Nasal cavity

Nasopharynx

Oral cavity

Pharynx

Oropharynx

Tongue

Salivary glands

Hypopharynx

Larynx

© 2012 Terese Winslow LLC
U.S. Govt. has certain rights
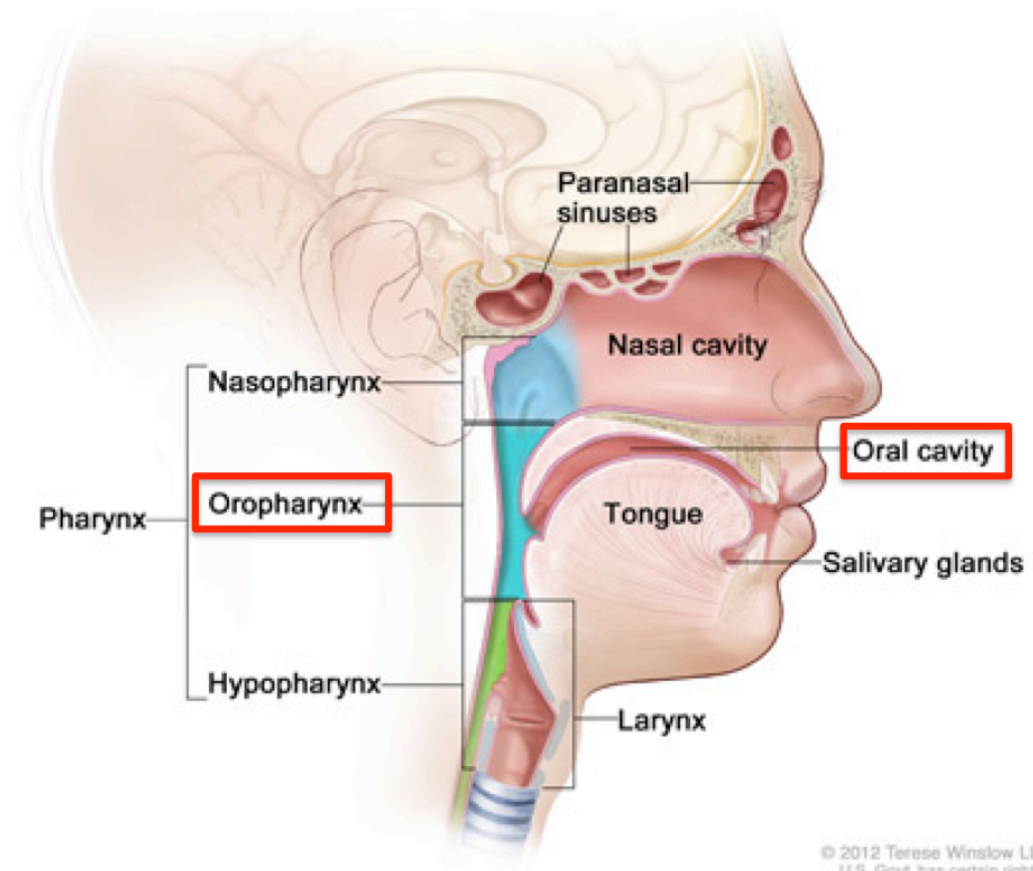
Figure 1.1 Head and neck cancer regions. Shown are the different regions of the head and neck from which HNSCC can originate. Carcinomas originating from the oral cavity and the oropharynx (red boxes) are explored in this dissertation. Image retrieved from the National Cancer Institute (http://www.cancer.gov/types/head-and-neck/)

oral cavity and oropharynx (NCI, 2016).

The highly variable survival rates of oral and oropharyngeal carcinomas depend on a variety of factors. These factors include the stage and metastatic spread of the particular tumor at the time of diagnosis and the underlying mechanism of tumor promotion. Overall, five-year and ten-year survival rates of patients with oral cavity and oropharyngeal cancer remain relatively poor at sixty-three percent, and fifty-one percent, respectively (NCI, 2016). Importantly, there are improved survival rates for patients that have been diagnosed at an early stage in their disease. For example, five-year survival rates can be as high as eighty-three percent in patients diagnosed with Stage I oral or oropharyngeal cancer, however, these survival rates decrease as the stage at time of diagnosis increases (NCI, 2016). The average survival rate of patients with cancer that has invaded local or surrounding tissues or that has metastasized to regional lymph nodes is as low as sixty-one percent survival (NCI, 2016). In the rare case that the oral or oropharyngeal cancer has metastasized to a distant organ, survival rates decrease even further to thirty-seven percent survival (NCI, 2016).

Risk Factors

Tobacco, alcohol, and betel quid

There are several different environmental risk factors that lead to the development and promotion of oral cavity and oropharyngeal carcinoma. The use of tobacco products, including cigarettes, cigars, pipes, chewing tobacco, and snuff, is the single largest risk factor for HNSCC (NCI, 2016). The use of tobacco products is so significant to the development of this disease that it is linked to eighty-five percent of all HNSCC (NCI,

2016). Chewing tobacco, in particular, is associated with a 50% increase in the risk of

developing HNSCC within the oral cavity (specifically in the cheeks, gums, inner lip)

that has come in direct contact with the tobacco product (NCI, 2016).

Other chewing products such as Betel Quid (BQ) have been shown to be risk

factors for increased development of oral cavity, oropharyngeal, and esophageal

carcinomas as a result of predominant use in southeast Asia, Taiwan and Papua New

Guinea (reviewed in Nair *et al.,* 2004). BQ, a mixture of areca nut (Areca catechu),

catechu (Acacia catechu) and slaked lime (calcium oxide and calcium hydroxide)

wrapped in a betel leaf (Piper betel), promotes cancerous transformations that lead to

HNSCC, and has also been shown to have an increased negative effect when used

together tobacco products (reviewed in Nair *et al*., 2004).

Alcohol use that is described as both frequent and heavy is another serious risk

factor that leads to the development of oral cavity and oropharyngeal carcinomas.

Similar to the combinatorial effect of using BQ and tobacco, alcohol and tobacco use also

have a synergistic effect on the risk of developing the disease and are frequent

comorbidities within HNSCC patients.

Genetic mutations

Genetic mutations are also risk factors that predispose patients to a higher

probability of developing head and neck squamous cell carcinoma. The most common

example are mutations in the tumor suppressor TP53 that promote cancer by altering the

regulation of the cell cycle and ultimately immortalizing otherwise normal epithelial cells

(Leemans *et al*., 2011). In addition to mutations in TP53, proteins associated with tumor

survival (PIK3CA-AKT1-MTOR-PTEN, EGFR and MET pathways), tumor proliferation

(p16, RB, MET, CCND1, CDKN2A/CDKN2B), and tumor differentiation (NOTCH1)

have also been implicated as risk factors for malignant transformation (Agrawal *et al*.,

2011; Lechner *et al*., 2013; Marur and Forastiere, 2016). Other genetic predispositions

include inability to metabolize carcinogens and impaired ability to repair damaged DNA

(Scully, 2000). Genetic mutations resulting in the development and progression of

HNSCCs also reduce the sensitivity of those tumors to chemoradiation and ultimately

more difficult to treat.

Infection with oncogenic human papillomavirus

Human papillomavirus (HPV), a sexually transmitted virus, has been linked to the

development of carcinomas in cutaneous squamous epithelium located in the cervix,

anogenital region, and oropharyngeal region (reviewed in Marur and Forastiere, 2016;

NCI, 2016). If viral infections are not cleared by the body's immune response, the viral

DNA can integrate into the human genome. The resulting expression of E6 and E7 viral

proteins target and alter the normal functions of tumor suppressors TP53 or RB that result

in uncontrolled cell growth and cell transformation (Vidal *et al*., 2008). Disruption of RB

function by viral protein E7 results in the expression of p16 ^{INK4A}, and is used as the

clinical marker to identify HPV-positive tissues (Zhang *et al*., 1999). HPV-positive head

and neck cancers can be further subcategorized due to the unique genetic and phenotypic

characteristics of the resulting tumors. There are several oncogenic subtypes of HPV,

HPV-16, HPV-18, HPV-31, and HPV-33, all of which increase the risk of malignant

transformation in transformed cells (Marur and Forastiere, 2016).   Although many HPV

strains exist, genotyping has revealed HPV-16 as the highest-risk strain that has been specifically identified as a causative agent in oropharyngeal cancer (NCI, 2016). Many patients that present with oropharyngeal cancer do not have the other traditional risk factors (frequent tobacco and alcohol use), which further implicate the role of the virus in the promotion of the disease. Importantly, rates of incidence for oropharyngeal cancer are increasing, an occurrence that is attributed to the overall increase in human papillomavirus (HPV) infections (Leemas *et al.*, 2011; NCI, 2016).

Pathology, classification, and treatment

Oral cavity and oropharyngeal cancers invade primarily via local invasion of the same tissue, and if left untreated, may metastasize to regional lymph nodes. The tumors and extent of the disease have traditionally been classified using the 'TNM' method where T describes the size and location of the primary tumor, N describes the number of lymph nodes to which the primary cancer has spread, and the M describes the number of other metastases (Leemans *et al.*, 2011; NCI, 2016). Together, these TNM classifications determine the stage of the cancer, ranging from Stage 0 to Stage IV, and are then used to direct or develop a treatment plan. More recently, HPV-status of the tumor has become an important prognostic factor in directing therapeutic options (Leemans *et al.*, 2011). Standard treatment plans begin with surgical resection of the primary tumor, followed by chemo-radiation therapy. Targeted therapies such as epidermal growth factor receptor (EGFR)-specific antibody cetuximab are being used in combination with radiotherapy with some success (Leemans *et al.*, 2011).

Despite vast improvements in surgical resection and radiotherapeutic techniques, survival rates of HNSCC patients have not significantly improved over the last thirty years (Leemans *et al*., 2011). Several factors including molecular heterogeneity and field cancerization are responsible for residual disease and frequent secondary primary tumors, all which contribute to the lack of improvement in survival rates (Rothenberg, *et al*., 2012). Field cancerization, which increases the high risk of recurrence in these patients, was first described in 1953 (Slaughter *et al*., 1953) yet only recently evaluated and understood in a molecular context (Tabor *et al*., 2001; Leemans *et al*., 2011). The neighboring mucosal epithelium often appears normal at the macroscopic level; however, the epithelial cells themselves can be dysplastic as a result of a variety of incurred genetic alterations (Tabor *et al*., 2001). The resulting pre-neoplastic regions of epithelium left behind in the surgical margins are responsible for the frequent development of metastases, second primary tumors, and add to the overall progression of HNSCC (Rothenburg *et al*., 2002; Leemams *et al*., 2011). Although much is known about the different causations of oral and oropharyngeal carcinomas (Agrawal *et al*., 2011; Park *et al*., 2011; Parfenov *et al*., 2014), the molecular mechanisms behind tumor cell invasion and metastasis have yet to be fully elucidated.

Epithelial to mesenchymal transition (EMT)

The steps for HNSCC to progress from an in situ carcinoma to a locally invading cancer have been classically described by Hanahan and Weinberg (Hanahan and Weinberg, 2000) and include the pivotal cellular change referred to as the epithelial to mesenchymal transition (EMT). This transition is the result of a variety of key

morphologic and biochemical changes in an epithelial cell in which the cell loses cell-to-cell adhesions and gains characteristics of a mesenchymal cell type. In a normal context, EMT is critically important for gastrulation in development, wound healing, and fibrosis (Natarajan *et al*., 2014). In the context of HNSCC, EMT is inappropriately activated and allows tumor cells to invade local tissue, regional lymph nodes, and metastasize to distant organs.

EMT begins with the loss of epithelial marker E-CADHERIN, an epithelial glycoprotein critical for maintaining adherens junctions and cell-cell adhesion (Natarajan *et al*., 2014). The downregulation of E-CADHERIN is considered the hallmark of EMT and is critical for this process to occur. Other genes are upregulated during EMT include the mesenchymal markers N-CADHERIN and VIMENTIN, matrix metalloproteinases (MMP-3 and MMP-9), and mesenchymal integrins, all which increase cell motility and invasiveness through changes directly to the morphology of the cell and the cell's interaction with the extracellular matrix (ECM) (Natarjan *et al*., 2014). These markers have been found to be important predictors of metastasis and poor prognosis in HNSCC, highlighting the importance of EMT in the progression of the disease. Several different signaling cascades (TGF-B, Wnt, NOTCH) and downstream transcriptional regulators also influence the induction of the EMT program (Natarajan *et al*., 2014). For example, basic helix-loop-helix (bHLH) transcription factors, TWIST and SNAIL-2, are associated with the initiation of EMT programs both in normal development and in cancer (Natarajan *et al*., 2014). Other homeobox transcription factors, SIX1 and FOXC2, have also been implicated to indirectly repress E-CADHERIN (Yang *et al*., 2008). The importance of EMT as a critical step in global metastatic cascade has been well

established, but the molecular understanding remains incomplete. Further insight into these molecular mechanisms that regulate EMT in HNSCC are necessary to develop novel targeted therapies and increase survival rates in patients with this highly invasive and recurrent disease.

LIM-only protein 4 (LMO4)

LIM-only protein 4 (LMO4) is a nuclear adapter protein critical for the assembly of multiprotein transcriptional complexes that regulate epithelial proliferation and differentiation in development (Yamada *et al*, 2000; Visvader *et al*., 2001; Mizunuma *et al*., 2003; Sum *et al*., 2005; and Yamada *et al*., 2008). During embryogenesis, LMO4 has been reported to modulate gene expression in neural crest cells as a cofactor in SNAIL2-mediated epithelial-mesenchymal transition (EMT) (Ferronha *et al*., 2013). This may account for the embryonic lethality and explain the phenotype of LMO4-*null* mice, manifested in failure of neural tube closure and exencephaly (Lee *et al*., 2005).

LMO4 expression, in particular overexpression, also characterizes a variety of epithelial malignancies, including oral cavity carcinomas (Mizunuma *et al*., 20003) and in carcinoma of the breast (Sum *et al*., 2005; Visvader *et al*., 2001; Montañez-Wiscovich *et al*., 2009; Tian *et al*., 2010), in which it has been associated with reduced tumor cell differentiation and increased lymph node metastasis (Mizunuma *et al*., 2003, Kwong *et al*., 2011). Additionally, LMO4 overexpression has also been reported in alveolar rhabdomyosarcomas (Armeanu-Ebinger, 2011). In contrast, knockdown of LMO4 reduced proliferation in human breast cancer cells and increased differentiation of mouse mammary epithelial cells (Sum *et al*., 2005).

The three remaining LMO proteins, LMO1, LMO2, and LMO3, have also been shown to have oncogenic properties when inappropriately expressed. LMO1 and LMO2 are dysregulated in T-cell acute lymphoblastic leukemia (reviewed in Aifantis *et al*., 2008; Rabbitts *et al*, 1998), while another study suggested that overexpression of LMO2 was associated with progression of prostate cancer (Ma *et al*., 2006). Likewise, LMO3 is upregulated in neuroblastoma (Ohira *et al*., 2005). Furthermore, enforced overexpression of LMO1, LMO2, and LMO4 in transgenic mice was either sufficient to induce tumor development (Ohira *et al*., 2005, Sum *et al*., 2005) or induced a pre-malignant phenotype. The following body of work suggests that LMO4 may also play an important regulatory role in the progression of head and neck carcinoma.

LIM domain-binding protein 1 (LDB1)

LDB1, like LMO4, is a transcriptional adapter protein critical for bridging and stabilizing multiprotein transcriptional complexes. Also known as Chip in Drosophila melanogaster, LDB1 was first reported to mediate DNA-looping and bridge enhancer and promoter regions (Morcillo *et al*., 1997), which pointed to the importance of this adaptor protein for transcriptional regulation. The protein's LIM domain (LID) enables it to bind to the diverse group of LIM domain proteins including LIM-only and LIM-homeodomain proteins (Figure 2, top). Together with these proteins, LDB1 regulates cell-fate determination and diverse developmental processes in both embryonic and adult tissues (Matthews and Visvader, 2003). Importantly, LDB1-*null* mice had severe anterior-posterior patterning defects (Mukhopadhyay *et al*., 2003), which further highlights the

importance of this co-factor for the normal function of many transcriptional regulatory programs.

Importantly, LDB1 expression in oral cavity carcinomas was found to correlate closely in abundance and location with LMO4 (Mizunuma *et al.*, 2003). LMO4 and LDB1 have also been reported to be frequently overexpressed in breast cancer, particularly in high-risk disease, and enforced expression of either gene inhibited mammary cell differentiation and promoted tumor cell invasion (Visvader *et al.*, 2001). These studies support the notion that together, LMO4 and LDB1 may be important in the progression of head and neck carcinoma.

Single-stranded binding proteins (SSBPs)

Studies of LMO2 protein turnover in erythroid cells in the Brandt Lab resulted in the identification of a novel function for a family of LDB1-interacting proteins — protection of LMO proteins and LDB1 from ubiquitylation and subsequent proteosomal degradation (Xu *et al.*, 2007). These putative single-stranded proteins, including single-stranded binding protein-2 (SSBP2) and single-stranded binding protein-3 (SSBP3), were shown to inhibit LDB1 ubiquitylation by an E3 ubiquitin ligase, RING finger LIM domain-binding protein (RLIM, also known as RNF12) (Xu *et al.*, 2007), and, in an LDB1-dependent manner, turnover of the LMO2 and LMO4 proteins. These results support the model that high SSBP expression levels within a cell can stabilize both LMO4 and LDB1, and SSBPs themselves can participate in the resulting multiprotein complexes that can then be recruited to target genes and influence transcriptional regulation (Figure 2, left). On the contrary, when SSBP expression levels are low, LMO4

and LDB1 are degraded and stabilization of the multiprotein transcriptional complexes and recruitment to target genes is reduced (Figure 1.2, right). In the context of a HNSCC cell, the expression levels of LMO4, LDB1, and SSBPs may alter direct gene targets involved in initiating or promoting EMT.
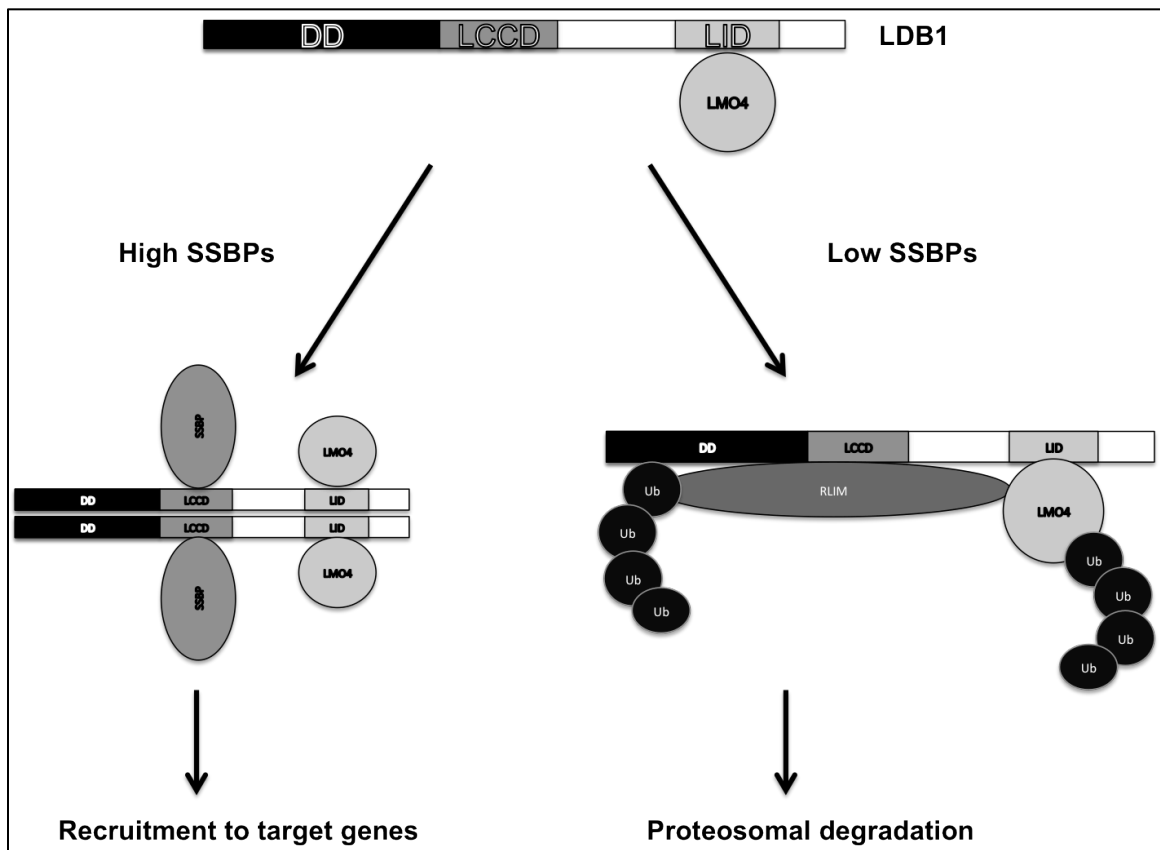
Figure 1.2 SSBPs protect LDB1 and LMO proteins from RLIM-mediated ubiquitylation and stabilize multiprotein transcriptional complexes. A schematic of the primary structure of LDB1 is shown at the top of figure, with three protein interaction domains demarcated, the dimerization domain (DD), LDB1/ChIP conserved domain (LCCD), and LIM-binding domain (LID). High SSBP expression prevents the E3 ubiquitin-ligase RLIM from binding LDB1 and promotes LDB1 homodimerization with stabilization of the trimolecular protein complex. Low SSBP levels facilitate ubiquitylation of LDB1 and LMO4, resulting in accelerated proteosomal degradation. The specific residues at which ubiquitin chains attach are unknown.

Conclusion

The following body of work reports several important findings regarding the role of LMO4, LDB1, and SSBPs in head and neck carcinoma. First, LMO4, LDB1, and SSBP2 and SSBP3 protein abundance were reciprocally correlated in a panel of human head and neck carcinoma cell lines. Second, these proteins colocalized to sites of tissue invasion and/or metastases in human carcinoma samples. Third, LDB1 gene inactivation profoundly inhibited tumor cell proliferation and invasiveness *in vitro* and *in vivo* and significantly decreased tumor vascularization. Fourth, LDB1 gene inactivation alters RNA expression pathways involved in promoting invasiveness. Lastly, global binding patterns of the LDB1 and SSBP2 were evaluated and common DNA motifs for global and specific target genes were determined. Together, these findings (Chapters III-VI) and the discussion that follows (Chapter VII) support the following two-part hypothesis: 1) the concordance between LMO4 and LDB1 protein abundance in HNSCC is likely the result of stabilization of these proteins by SSBPs; 2) the resulting multiprotein complex containing LMO4, LDB1, and SSBPs can alter specific transcriptional programs controlling invasiveness, lymph node metastasis, angiogenesis, proliferation/apoptosis, and likely cancer stem cells (CSCs) within HNSCC (Figure 1.3).
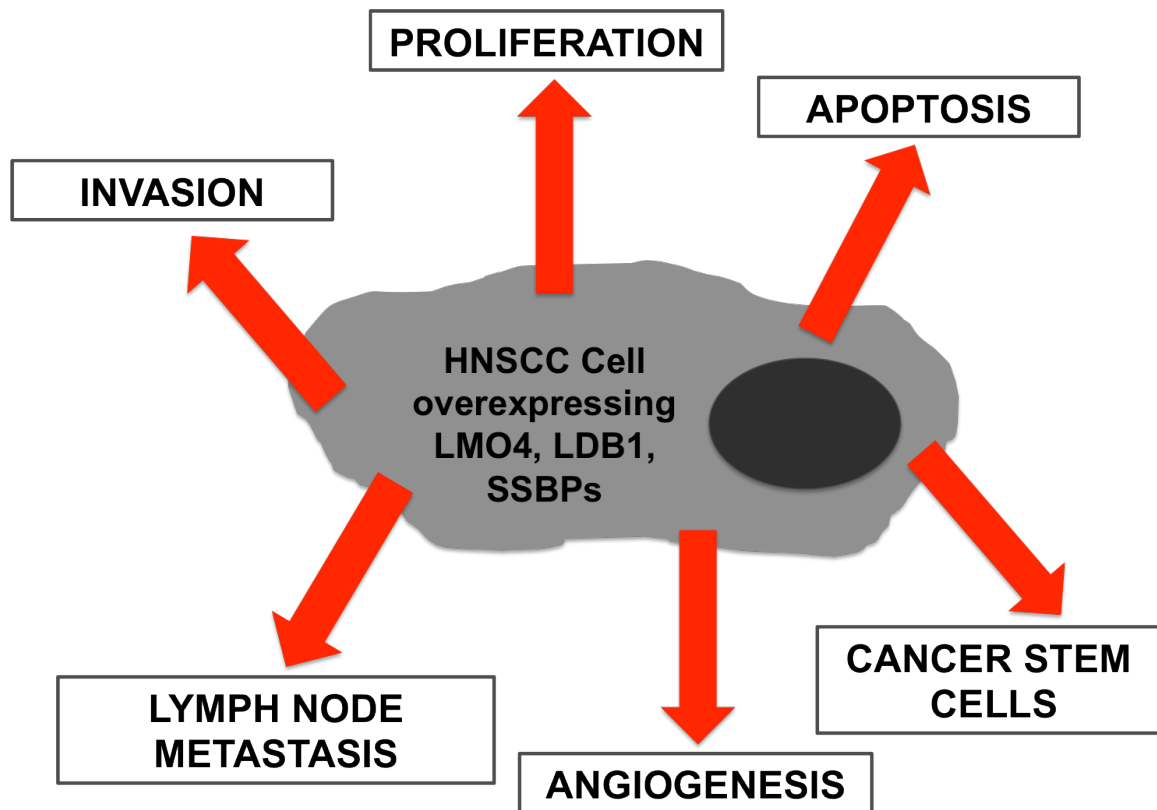
Figure 1.3 Possible mechanisms of progression in a head and neck squamous cell carcinoma cell as a result of overexpression of LMO4, LDB1, and SSBPs. A schematic of the different mechanisms of cancer progression which may be influenced or directly regulated by a complex containing LMO4, LDB1, and SSBPs in a head and neck squamous cell carcinoma cell. Increased invasion, increased proliferation, decreased apoptosis, increased lymph node metastasis, increased angiogenesis, and potentially maintenance of cancer stem cells are all potentially influenced at the level of transcriptional regulation by a prooncogenic complex containing LMO4, LDB1, and SSBPs and are explored or discussed in this dissertation.

CHAPTER II

MATERIALS AND METHODS

Cell culture and reagents

Human oral cavity carcinoma cell lines SCC-4 (catalog number CRL-1624), SCC-25 (catalog number CRL-1628), SCC-9 (catalog number CRL-1629), SCC-15 (catalog number CRL-1623), and Cal-27 (catalog number CRL-2095) were obtained from ATCC, SCC-61 and UM-SCC-47 were contributed by Dr. Wendell Yarbrough (Vanderbilt University), HN-SCC-131 was provided by Dr. Susanne Gollin (University of Pittsburgh), and VU-SCC-1352 and VU-SCC-1729 came from the Barry Baker Laboratory for Head and Neck Oncology at Vanderbilt University, where they were derived. All oral cavity carcinoma cell lines were grown in Dulbecco's modified Eagle medium (DMEM; Life Technologies-Thermo Fisher Scientific, Grand Island, NY) supplemented with 10% fetal bovine serum (FBS, Atlas Biologicals, Fort Collins, CO) and 1% penicillin/streptomycin and incubated in 50% CO_2 at 37°C. All oral cavity carcinoma cell lines were grown on T-175 cm^2 flasks (CLS431080, Sigma-Aldrich, St. Louis, MO) to 60-85% confluency. Fetal esophageal fibroblasts were grown in DMEM supplemented with 10% FBS and 1% penicillin/streptomycin and cultured in 50% CO_2 at 37°C.

Antibodies

To detect LDB1, LMO4, SSBP2, and SSBP3 in immunoblot and immunohistochemistry, the following antibodies were obtained: a polyclonal goat antibody for LDB1 (SC 11198) ordered from Santa Cruz Biotechnologies (Dallas, TX); a monoclonal rat antibody for LMO4 generously provided by Dr. Jane Visvader (Walter and Eliza Hall Institute, Melbourne, Australia) and described here (Sum *et al*, 2002); a custom polyclonal rabbit antibody for SSBP2 was generated by SDIX (Newark, DE) using a synthetic polypeptide antigen sequence to amino acids 163-243; and a polyclonal rabbit antibody for SSBP3 was generously provided by Dr. Lalitha Nagarajan and is described here (Liang *et al*., 2005). Control antibodies β-actin (ab44990) and glyceraldehyde phosphate dehydrogenase (GAPDH) (ab9485) were purchased from Abcam (Cambridge, MA).

Immunoblot

Immunoblots were carried out as previously described (Xu *et al*., 1995). Blots were imaged using Image J (Image J Software) and levels of LDB1, LMO4, SSBP2, SSBP3 were normalized to B-ACTIN and then normalized to the highest expresser. LDB1 knockout clones and their effects on LMO4 were evaluated using a GADPH loading control.

Human Tissue Preparation

Individual tumor samples

Oral cavity carcinomas, oropharyngeal carcinomas, and regional lymph nodes were collected from consented patients (IRB # 030062) during operative procedures at Vanderbilt University hospitals. Within thirty minutes of removal from each patient, the tissue was placed in a biopsy cassette and submerged in 10% formalin for no longer than 24 to 48 hours. The cassettes containing the tissue were then moved to 70% ethanol for transportation to Translational Pathology Shared Resource (TPSR) for paraffin embedding and subsequent sectioning.

TMA preparation

Tumor and adjacent normal tissues were collected from consented patients (IRB # 030062) during operative procedures. Within thirty minutes of removal from each patient, the tissue was placed in a biopsy cassette and submerged in 10% Formalin for no longer than 24 to 48 hours. The cassettes containing the tissue were then moved to 70% ethanol for transportation to TPSR for paraffin embedding. Slides and blocks were reviewed by Dr. Kim Ely in the department of surgical pathology (Vanderbilt University) to determine if the specimen was suitable for use. TMAs were generated from blocks pre-selected by Brandee Brown (manager of the Head and Neck Cancer Tissue Biorepository and Barry Baker Head and Laboratory, Vanderbilt University). Samples were separated by 1.5mm from one another and were punched in triplicates for statistical purposes.

Immunohistochemistry: LMO4, LDB1, SSBP2, SSBP3

Immunohistochemistry performed by the TPSR at Vanderbilt University, Nashville, TN as described. Slides were placed on the Leica Bond Max IHC stainer. All steps besides dehydration, clearing and cover-slipping were performed on the Bond Max. Slides were de-paraffinized. Heat induced antigen retrieval was performed on the Bond Max using their Epitope Retrieval 1 solution for 10 minutes. Slides were incubated with antibody (anti-LDB1 1:500 dilution; anti-LMO4 1:700 dilution; anti-SSBP2 1:1200 dilution; anti-SSBP3 1:500 dilution) for one hour and followed by the appropriate biotinylated secondary antibody for 30 minutes at 1:200 dilutions. The Bond Polymer Refine detection system was used for visualization. Slides were then dehydrated, cleared and cover-slipped.

CRISPR/Cas9-mediated deletion of *LDB1*

CRISPR/Cas9 system was used to target *LDB1* (Figure 2.1) as previously described (Wang *et al.*, 2014) using the following guide RNAs 5'-3' sequences (Integrated DNA Technologies, Coralville Iowa) as follows:

LDB1 gRNA1 top (CACCGACCATGCTGGATAGGGATGT),

LDB1 gRNA1 bottom (AAACACATCCCTATCCAGCATGGT),

LDB1 gRNA2 top (CACCGGTAGGCGGATACATGGGAGT),

LDB1 gRNA2 bottom (AAACACTCCCATGTATCCGCCTAC).
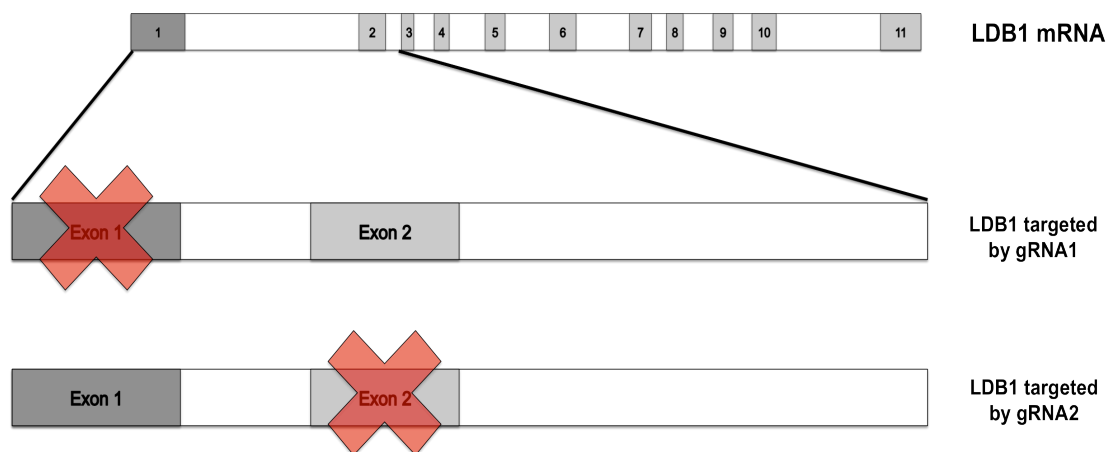
Figure 2.1 Guide RNAs target *LDB1* exons to generate CRISPR/Cas9 knockout. (Top) Full length LDB1 mRNA is depicted with 11 exons (grey boxes). (Middle) Exons 1 and 2 enlarged to show target of exon 1 by guide RNA 1 (gRNA1). (Bottom) Exons 1 and 2 enlarged to show target of exon 2 by guide RNA 2 (gRNA2).

VU-SCC-1729 cells were transduced with retrovirus expressing the guide RNAs targeting *LDB1* exon 1 (gRNA1), exon 2 (gRNA2), and an empty vector.

Viral production and transduction

100 pmoles of top and bottom strand oligonucleotides were mixed and phosphorylated overnight with 2 units of T4 Polynucleotide Kinase (NEB) in T4 ligase buffer. Phosphorylation reactions were then adjusted to 200 mM NaCl and placed in a 95°C heat block for five minutes. After five minutes, the heat block setting was adjusted to 23°C, and reactions were left in the heat block until the lower temperature was reached. 33 fmoles of each duplex were then ligated to 10 fmoles of plentiCRISPR (Shalem *et al*., 2014) that had been digested with BsmBI (NEB), dephosphorylated with alkaline phosphatase, and gel purified. Ligation reactions were transformed into

20

chemically competent XL1 blue *E. coli*. A single colony of each construct was picked and expanded; a portion was used to prepare DNA for analytical restriction digest and sequencing with U6 promoter primer (ACTATCATATGCTTACCGTAAC) to verify inserts. Maxiprep DNA was prepared from 500 mL saturated culture using a modified alkaline lysis/LiCl and PEG 8000 precipitation protocol.

1 pmole of each recombinant plentiCRISPR was mixed with 2 pmoles of both pSPAX2 (GAG/POL, Addgene) and pVSVG (ENV, Addgene) and transfected in triplicate to log phase HEK 293t cells using HEPES buffered saline/Calcium Phosphate. Prior to transfection each replicate of cells was grown in a 10 centimeter dish to 50-70% confluence (5-7 x 10^6 cells per dish) in 10 mL IMDM media at 37°C and 5% CO_2. 12-18 hours after transfection, media was aspirated and exchanged with 6 mL fresh IMDM. Conditioned media containing virus was removed 24 hours later and again replaced with fresh IMDM, and collected once more after an additional 24 hours. Individual viral supernatants from replicates of both time points were pooled and centrifuged briefly at 800 x g to remove cellular debris, then subjected to ultracentrifugation in a Beckman Ti60 rotor for 3.5 hours at 100,000 x g. The viral pellet was resuspended in 1/10 the starting volume of IMDM and stored at -80°C.

VU-1729-LDB1 generation

Viruses were serially titered on VU-SCC-1729 cells. Growth and viability were estimated by vital dye staining and cell counting, with comparison to mock-transduced cells. Cells were selected with Puromycin (7 µg/mL), which was added to the medium 36 hours after transfection. Transduced cells were then cultured for 48 hours in selective

medium. Single colonies were transferred to a 96-well plate and grown in selective

medium. Monoclonal lines were screened for LDB1 and LMO4 via immunoblot. Final

monoclonal cell lines were used in matrigel invasion, co-culture, xenograft, and RNA-seq

assays.

In vitro proliferation assay

VU-SCC-1729 and VU-1729-2:7 cells were incubated in 6-well plates containing

DMEM (Gibco, #11330-032) supplemented with 10% fetal bovine serum (FBS, Atlas

Biologicals, #F-0500-D), 1% Penicillin/streptomycin, incubated at 50% CO_2, 37°C.

Initially (Day 0), a total of 46,000 cells/well were plated into four separate wells. The

contents of one well was harvested daily for viable cell count by trypan blue exclusion.

Assay was performed in biological triplicate. Standard statistical analysis is reported as

mean +/- SEM and unpaired student t-tests were used to determine statistical significance

of differences in mean cell number. $P < 0.05$ was considered statistically significant

(GraphPad Software, La Jolla, CA)

Matrigel invasion assay

Invasion was measured using a modified Boyden Chamber Matrigel method

(Albini *et al.*, 1987). Transwell 8.0 μm permeable membrane chambers (Costar Corning,

Corning, NY) were prepared as follows: the superior surface of membranes were coated

with 1:5 mixture of Matrigel (BD Biosciences, Franklin Lakes, NJ, 354230) in Opti-

MEM (Gibco, 31985-070) and the inferior surface of membranes were coated with 1:20

mixture of Matrigel in Opti-MEM. After letting the Matrigel solidify at 37°C for 1 hour,

Matrigel was blocked with 1% bovine serum albumin for 30 minutes at 37°C. VU-SCC-1729 and cell line derivatives were serum starved for 24 hours in Opti-MEM reduced serum media prior to being trypsonized, counted, and plated at a density of 100,000 cells/chamber (Figure 2.2).
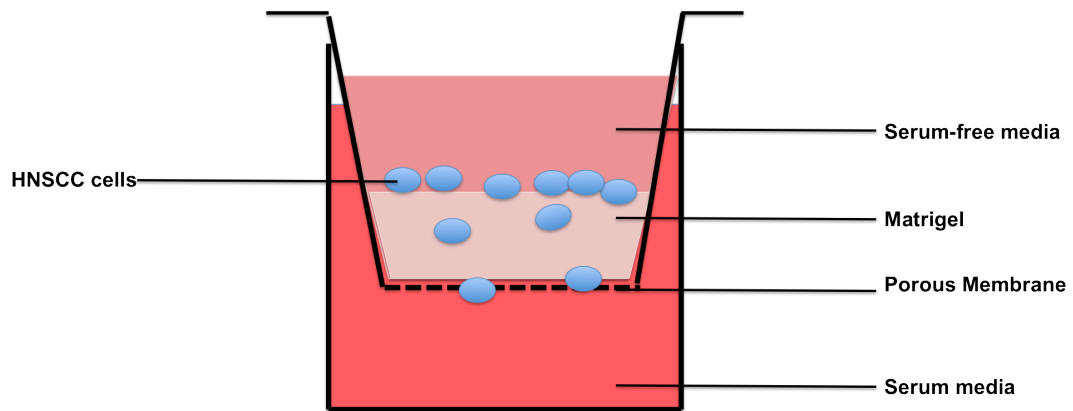
Figure 2.2 Matrigel invasion schematic. Serum-starved HNSCC cells (blue circles) are plated on top of matrigel in serum-free conditions. HNSCC cells begin to invade down into the matrigel in response to serum-containing media below. Cells are trapped in the porous membrane and fixed for subsequent quantification.

Chambers were submerged in DMEM (Gibco, 11330-032) supplemented with 10% FBS, 1% Penicillin/Streptomycin and allowed to invade at 37°C. After 6 hours, cells were fixed with 10% formalin and stained with crystal violet. Multiple sections from each transwell image were randomly selected and counted from at least two transwell chambers per biological replicate. Data is from at least 2 biological replicates. Standard statistical analysis is reported as mean +/- SEM and unpaired student t-tests were used

determine statistical significance of differences in mean cell number. P < 0.05 was

considered statistically significant (GraphPad Software, La Jolla, CA)

Organotypic Reconstruct Assay

VU-SCC-1729-P (parental) and VU-1729-2:7 (LDB1 knockout) lines were grown

in 17 day organotypic cultures (OTC) as previously described (Andl *et al*., 2010) with

specific or changed details below (Figure 2.3).
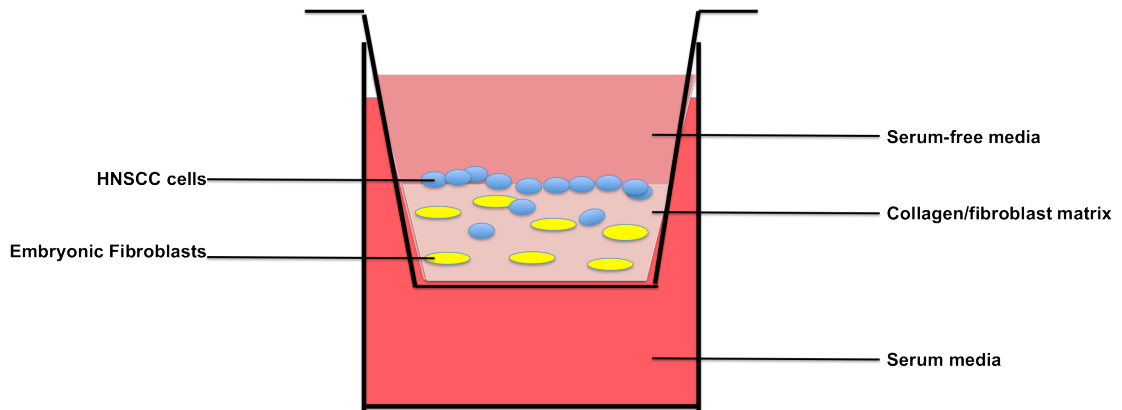
Figure 2.3 Organotypic co-culture schematic. HNSCC cells (blue circles) are plated
on top of a mixed matrix of collagen and fibroblasts (yellow ovals) in serum-free
conditions. HNSCC cells (blue circles) begin to invade down into the mixed matrix in
response to serum-containing media and signals from fibroblasts (yellow ovals). The
result of this co-culture is a 3-dimentiontial tissue that can be fixed, sectioned, and
stained to observe level of invasion.

Day 1. 16 mL/6-well plate high concentration rat-tail collagen (Corning, 354249)

was prepared on ice per BD instructions. A final concentration of 2.2 mg/mL collagen

was obtained by diluting high concentration collagen with 1 N NaOH and sterile water.

Transwell inserts were added to a deep well plate. The bottom layer consisted of the following layers, which were mixed in order on ice (table below):

	1 well (μl)	3 wells (μl)	1 plate (μl)
10xEMEM	98.3	295	590
FBS (Hyclone)	100	300	600
L-glutamine	8.3	25	50
Sodium bicarbonate	20	60	120
Collagen (2.2mg/mL)	800	2400	4800

The bottom collagen layer was incubated at 37°C for at least 20 minutes to let the matrix solidify. During this time, fetal esophageal fibroblasts were counted and adjusted to the cell concentration to 300,000 cells/mL, and used in the top layer, which was made of up the following layers, mixed in order on ice (table below):

	1 well (μl)	3 wells (μl)	1 plate (μl)
10xEMEM	275	825	1650
FBS (Hyclone)	308	925	1850
L-glutamine	25	75	150
Sodium bicarbonate	58	175	350
Collagen (2.2mg/mL)	1750	5250	10.5mL
Matrigel	583	1750	3500
Fibroblasts	250	750	1500

The top layer was incubated at 37°C for 1 hour to allow the matrix to harden. After the top and bottom layers were hardened, 15mL of fetal esophageal fibroblast media was added to the bottom of the wells and 2 mL of media to the top.

Day 7. The media was changed to DMEM+F-12 (DMEM from Sigma, F-12 from Gibco) mixed in a 3:1 ratio. 15 mL media was added to the bottom, 2 mL was added to the top and the culture was incubated for 1 hour at 37°C. Human head and neck

squamous cell carcinoma cells (VU-SCC-1729-P and VU-1729-LDB1 KO clones) were counted and adjusted to 1×10^7 cells/mL in media containing serum. DMEM+F12 media was aspirated from the culture (both top and bottom chambers) and 50 µl of cells were added to each well, forming a triangle pattern of droplets (schematic below), and incubated at 37°C for 2 hours.
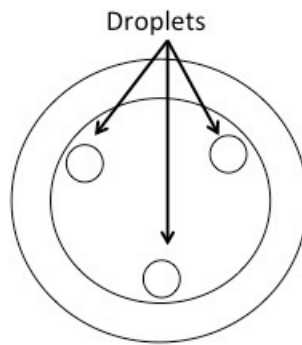
After 2 hours, Epidermalization I media (table below) was added to the plate (15 mL to the bottom, 2 mL to the top).

	Epidermalization I (17mL/well, 102mL per plate)	Epidermalization II (17mL/well, 102mL per plate)	Epidermalization III (13mL/well, 156mL per plate)	Epidermalization III serum-free (13mL/well, 78mL per plate)
DMEM (+CaCl$_2$)	36.25	36.25	23.7	23.7
F-12	12	12	23.7	23.7
L-glutamine	1	1	1	1
Hydrocortisone	0.1	0.1	0.1	0.1
ITES (500x)	0.1	0.1	0.1	0.1
O-phosphory-ethanolamine	0.1	0.1	0.1	0.1
Adenine	0.1	0.1	0.1	0.1
Progesterone	0.1	0.1	-	-
Triiodothyronine	0.1	0.1	0.1	0.1
Newborn calf serum	0.05 (chelated)	0.05	1	-
Gentamycin	0.05	0.05	0.05	0.05

Day 9. Epidermalization II media (table above) was added to the plate (15mL to the bottom, 2mL to the top).

Day 11. Epidermalization III media (table above) was added to the plate (15mL to the bottom, 2mL to the top).

Day 13. Epidermalization III media (table above) was added to the plate (13mL to the bottom of the plate only).

Day 15. Epidermalization III serum-free media (table above) was added to the plate (13mL to the bottom of the plate only).

Day 17. Cultures were harvested on Day 17 and fixed in 10% formalin overnight. Cultures were paraffin embedded, sectioned and stained with hematoxylin and eosin (H&E) by the TPSR Core (Vanderbilt University). Cell line OTCs were performed in duplicate. Area of invasion was measured from multiple views of OTC sections and calculated using Image J (Image J software). Standard statistical analysis is reported as mean +/- SEM and unpaired student t-tests were used to determine statistical significance of differences in mean invasion area. $P < 0.05$ was considered statistically significant (GraphPad Software, La Jolla, CA)

Xenografts

Athymic nude male animals were obtained from The Jackson Laboratory and housed under pathogen-free conditions. All experiments were performed in accordance with AAALAC guidelines and with Vanderbilt University Institutional Animal Care and Use Committee approval. For tumor progression studies *in vivo*, 4-5 week old nude male mice were injected subcutaneously into the dorsal flank with (500,000 cells + matrigel at

a final volume of 100 µL per mouse) as described previously (Brantley-Sieders, 2011). Implantation of tumor cells was carried out under isoflurane anesthesia and every effort was made to minimize suffering. VU-SCC-1729-P (parental) cells and VU-SCC-1729-2:7 (CRISPR/cas9-mediated deletion of LDB1 subline) cells were injected into the left and right contralateral flanks, respectively (Figure 2.4).

PARENTAL
VU-1729-P cells
(Left flank)

LDB1 KO
VU-1729-2:7 cells
(Right flank)
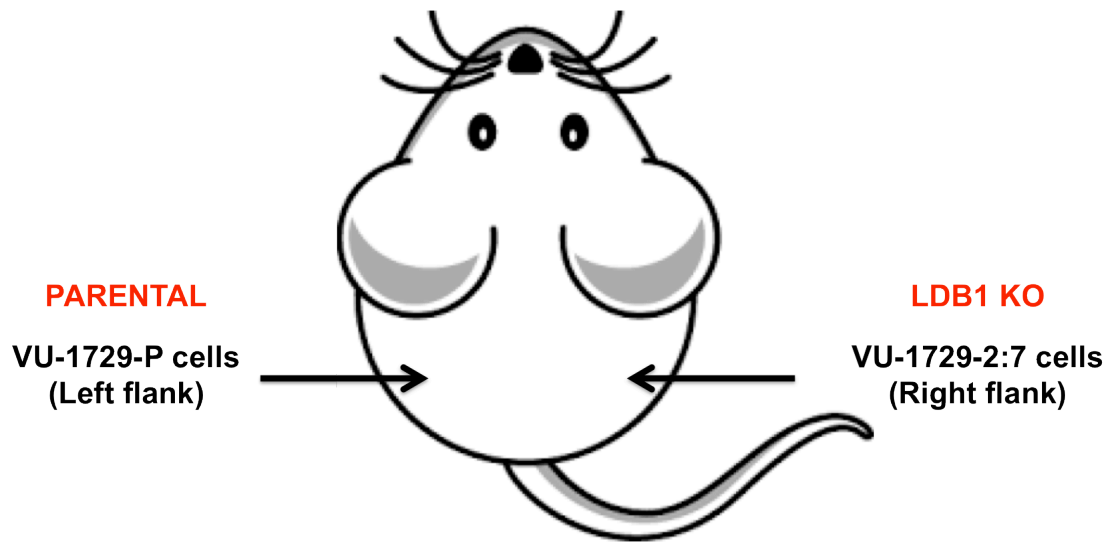
Figure 2.4 Schematic of xenograft injections. Dorsal flanks of athymic nude mice were injected subcutaneously with VU-SCC-1729-P cells on the left flank and VU-SCC-1729-2:7 cells on the right flank. Tumors were allowed to grow for 23 days.

Resulting tumors were measured over time using caliper measurements of tumor length and width. Measurements were taken for the following time points: Day 0, Day 5, Day 10, Day 13, Day 20, and Day 23. Tumor volume was quantified using the following formula (Bergers, 2000):

(Eq. 2.1) $$\textbf{Tumor Volume = (Length x Width)}^2 \textbf{ x (0.52)}$$

28

Tumors were harvested on Day 23 post-injection and processed for histology. Ten independent mice were used (n = 10). Xenograft tumors were dissected from mice and fixed in 10% formalin for 48 hours.

PCNA Immunohistochemistry

Tissue processing, *ET AL*&E staining, and immunohistochemistry for Cleaved Caspase 3 was performed by the Vanderbilt Translational Pathology Shared Resource. PCNA immunohistochemistry analysis was performed as described previously (see Brantley-Sieders *et al*. 2011 Cancer Res 71: 976-87; Youngblood *et al*. 2015 Mol Cancer Res 13: 524-37). Photomicrographs acquired on an Olympus CK40 inverted microscope through an Optronics DEI-750C charge-coupled-device video camera using CellSens capture software. CellSens software was also used to quantify the average percentage of PCNA or Cleaved Caspase 3 positive nuclei relative to total in 3 to 4 fields/sample (10-20X magnification) as described previously. Data are a representation of 5-7 independent tumors/condition. Standard statistical analysis is reported as mean +/- SEM and unpaired student t-tests were used to determine statistical significance of differences in cell proliferation. $P < 0.05$ was considered statistically significant (GraphPad Software, La Jolla, CA)

vWF Immunofluorescence

Tissue processing, H&E staining was performed by the Vanderbilt Translational Pathology Shared Resource. vWF immunofluorescence analysis was performed as described previously (see Brantley-Sieders *et al*. 2011 Cancer Res 71: 976-87;

Youngblood *et al*. 2015 Mol Cancer Res 13: 524-37). Photomicrographs acquired on an Olympus CK40 inverted microscope through an Optronics DEI-750C charge-coupled-device video camera using CellSens capture software. CellSens software was also used to quantify the microvascular density (vWF positive blood vessels) in 3 to 4 fields/sample (10-20X magnification) as described previously. Data are a representation of 5-7 independent tumors/condition. Standard statistical analysis is reported as mean +/- SEM and unpaired student t-tests were used to determine statistical significance of differences in mean microvessel density. P < 0.05 was considered statistically significant (GraphPad Software, La Jolla, CA)

RNA-seq Analysis

Total RNA was isolated from parental control (VU-SCC-1729-P) and CRISPR-cas9 LDB1 knockout (VU-1729-2:7) cells in biological triplicates (Qiagen RNeasy kit). RNA-seq libraries were generated and sequenced by the Vanderbilt Technologies for Advanced Genomics (VANTAGE) Core at Vanderbilt University, Nashville, TN. Bioinformatics was performed by Yan Guo (Vanderbilt University, Nashville, TN) and the significantly impacted pathways were analyzed using Advaita Bio's iPathwayGuide (Advaita Bio, Plymouth, MI; http://*et al*.advaitabio.com/ipathwayguide). RNA-seq data were submitted to Gene Expression Omnibus and assigned accession number GSE79183 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE79183).

Chromatin Immunoprecipitation

Standard chromatin immunoprecipitation (ChIP) followed formaldehyde fixation and sonication conditions described below in ChIP-exo. Protocol and reagents from EZ ChIP kit (Upstate Biotechnologies, 17-371) were used to lyse, wash, and elute immunoprecipitates, and DNA was purified using standard phenol/chloroform isolation. Real time-quantitative PCR (RT-qPCR) was performed using SYBR® Green Quantitative RT-qPCR Kit (BIO-RAD, 172-5140) was used to evaluate occupancy at target genes. Forward (F) and reverse (R) primers are listed for each target gene below.

CDC27

(F): CTGCAGCACCGTCATCCT

(R): GTAACGGTCGCTGGTGAGTT

CDC27-exon

(F): TCCCAACAATGAAAGAGTAAAGC

(R): TTGCAGAAGGGGAACAAATC

MIR3687

(F): GGATGCGTGCATTTATCAGA

(R): GTTTCTCAGGCTCCCTCTCC

CDH11-promoter

(F): TTAGGATTCTACTCAAGAGAACAGGA

(R): TTTATGTTGTTTCTTTTGTTTATCTGC

CDH11-exon

(F): TCCTGGACCTTGACAATGAA

(R): TGACTCTGGTGATGGGAACA

CKAP2L-promoter

(F): TGGGTATTCTATTGTGCGGATA

(R): CAGCATTATTCATAATAGCCAAAAA

HMGA2-intron

(F): AGCCAAGCCAGGCATCTAAG

(R): TCTTGTGTTGAGTTTAATTTCAGAGG

ChIP-exo

Chromatin immunoprecipitation followed by exonuclease digestion (ChIP-exo) was used to acquire genome-wide occupancy information for LDB1 and SSBP2 in VU-SCC-1729 cells as was previously developed (Rhee *et al.*, 2012) and described below.

Formaldehyde fixation and sonication

Adherent VU-SCC-1729 cells were grown in 1.75 L flasks under normal conditions to 85% confluency. Growth medium was removed and washed away with phosphate buffered saline (PBS). Cells were fixed in a 1% formaldehyde PBS solution for 10 minutes after which the reaction was quenched with excess glycine. Cross-linked cells were washed again with cold PBS (4°C), scraped off flask and pelleted into flasks containing ~150 million cells each. Cell pellets were stored at -80°C until ready for cell lysis and sonication.

Frozen pellets of cross-linked cells were thawed on ice, re-suspended in cell lysis buffer (final concentration 10mM Tris pH 8.0, 10mM NaCl, 0.5% NP40, plus complete protease inhibitors) and lysed on ice for 10 minutes. Cells were centrifuged (4°C) at 2500 rpm for 5 minutes and re-suspended in nuclear lysis buffer (final concentration 50mM Tris pH 8.0, 10mM EDTA, 1.0% SDS, plus complete protease inhibitors) and lysed on ice for 10 minutes. Cells were diluted with cold immunoprecipitation dilution buffer (final concentration 20 mM Tris ph 8.0, 2mM EDTA, 150mM NaCl, 1% Triton x-100) and transferred to 1.5 mL tubes containing 300 µL cell lysate each (~6.6 million cells/100 µL). Tubes were transferred to a chilled (4°C) biorupter (Diagenode Biorupter, B01020001) water bath and were sonicated on high for a total time of 40 minutes (4 x 10 minute sessions of 30 seconds ON/30 seconds OFF). Samples were put on ice between each session until water bath returned to 4°C. After sonication, samples were centrifuged at 14,000 rpm for 10 minutes at 4°C. Supernatants were transferred to new tubes and used for immunoprecipitation.

Chromatin immunoprecipitation

Antibodies (SSBP2 and LDB1) and sonicated extract were incubated overnight at 4°C. Magnetic Protein G beads (New England BioLabs, S1430S) were added to antibody/extract mixture and incubated for 1 hour at 4°C. Tubes were placed against a magnetic rack, supernatant was removed, and beads were washed for three minute-rotations in the following buffers: RIPA Salt buffer (final concentration 1X PBS, 1% NP40, 0.5% NaDeoxycholate, 0.1% SDS); LiCl Buffer (final concentration 100 mM Tris pH 8.0, 500 mM LiCl, 1.0% NP40, 1.0% NaDeoxycholate); FAT buffer (final

concentration 40 mM Tris pH 8.0, 7 mM EDTA, 56 mM NaCl, 0.2% SDS, 0.375%

TritonX-100); Tris-HCl butter (final concentration, 10 mM Tris pH 7.5). After last wash

(Tris-HCl buffer) is removed, on-bead library preparation began (see below).

Library preparation for Illumina

Polishing master mix (ddH$_2$0, 10x NEB2 buffer, 1 mg/mL BSA, 3 mM dNTPs, 3

U/µl T4 DNA polymerase) was added to the beads, pipeted to mix, and incubated in a

thermomixer at 12°C for 20 minutes at 1400 rpm. Tubes were placed against a magnet,

polishing mix removed, and washed with 10mM Tris buffer.

Kinase master mix (ddH$_2$0, 10x T4 DNA ligase buffer, 10 U/µl T4 polynucleotide

kinase) was added to the beads, pipeted to mix, and incubated in a thermomixer at 37°C

for 30 minutes at 1400 rpm. Tubes were placed against a magnet and kinase mix was

removed.

A-tailing master mix (ddH$_2$0, 10x NEB2 buffer, 3 mM dATP, 5 U/µl Klenow 3'-

5' exo minus) was added to the beads, pipeted to mix, incubated in a thermomixer at

37°C for 30 minutes at 1400 rpm. Tubes were placed against a magnet and A-tailing mix

was removed.

Master ligation mix (ddH$_2$0, 1mg/mL BSA, NEB 10x T4 DNA ligase buffer, NEB

400 U/µl T4 DNA ligase) was aliquoted to fresh tubes (one tube per sample) and

immediately mixed with an Index Adapter (different for each sample). Ligation mixes

(containing unique Index Adapters for each sample) were added to the beads, pipeted to

mix, incubated in a thermomixer at 25°C for 2 hours at 1400 rpm. Tubes were placed

against a magnet and ligation mix was removed. Samples were then washed three

different times for 3 minutes each with aspiration of supernatant in between using FAT buffer, LiCL Buffer, and 10 mM Tris-HCL buffer.

Following the last wash, Phi-29 master mix (ddH$_2$0, 1mg/mL BSA, 10x Phi29 reaction buffer, 3mM dNTPs, 10 U/µl Phi29 DNA polymerase) was added to the beads, pipeted to mix, and incubated in a thermomixer at 30°C for 20 minutes at 1400 rpm. Tubes were placed against a magnet and Phi-29 mix was removed.

Kinase master mix (ddH$_2$0, 1mg/mL BSA, 10x T4 DNA ligase buffer, 10 U/µl T4 PNK) was added to the beads, pipeted to mix, and incubated in a thermomixer at 37°C for 20 minutes at 1400 rpm. Tubes were placed against a magnet and kinase mix was removed.

Lambda exonuclease master mix (ddH$_2$0, 1mg/mL BSA, 10x lambda exonuclease reaction buffer, 5 U/µl lambda exonuclease) was added to the beads, pipeted to mix, and incubated in a thermomixer at 37°C for 30 minutes at 1400 rpm. Tubes were placed against a magnet and lambda exonuclease mix was removed.

RecJ$_f$ exonuclease master mix (ddH$_2$0, 1mg/mL BSA, 10x NEB2 buffer, 30 U/µl RecJ$_f$ exonuclease) was added to the beads, pipeted to mix, and incubated in a thermomixer at 37°C for 30 minutes at 1400 rpm. Tubes were placed against a magnet and RecJ$_f$ exonuclease mix was removed. Samples were washed overnight in FAT buffer at 4°C. The following morning, samples were washed 5 times with each of the following buffers: RIPA buffer, RIPA salt buffer, LiCl Buffer. After the last wash was removed, samples were eluted off the beads with ChIP elution buffer and then were incubated at 65°C for 15 minutes on a heat block. Tubes were placed against a magnet and the supernatant was transferred a new tube with the addition of 20 mg/mL proteinase K, then

incubated at 65°C for 30 minutes on a head block. To reverse crosslinks, samples were

placed on a heat block at 95°C for 10 minutes. DNA was then extracted by standard

phenol:chloroform method. Following ethanol precipitation dried DNA pellets were

resuspended in 10 µl ddH$_2$0 and transferred to 0.5 mL PCR tubes.

Primer master mix (ddH$_2$0, 1mg/mL BSA, 10xPhi29 reaction buffer, 3 mM

dNTPs, FX-15 primer) was added to resuspended DNA pellets and pipeted to mix.

Samples were run in the thermocycler using the following program: 95°C for 5 minutes,

40°C for 10 minutes, 30°C for five minutes, 30°C hold (add 1 µl of 10U/µl Phi29 DNA

polymerase at this step), 30°C for 20 minutes, 65°C for 10 minutes, and 4°C hold.

At the end of the thermocycler program, A-tailing master mix (ddH$_2$0, 10x NEB2

buffer, 3mM dATP, 5 U/µl Klenow 3' to 5' exo minus) was added to the samples,

pipeted to mix, and incubated in a thermocycler at 37°C for 30 minutes and then at 75°C

for 20 minutes.

At the end of the thermocycler program, ligation master mix (ddH$_2$0, 10x T4

DNA ligase buffer, 15 uM ExA1-58/13 adapter, 600 U/µl T4 DNA ligase) was added to

the samples, pipeted to mix, and incubated in the thermocycler at 25°C for 2 hours.

Samples were either frozen at this step (-80°C) or directly moved to the following

Ampure purification.

Ligation samples were combined with 1.8 volumes of AMPure beads and pipeted

for 20 seconds to mix. Samples were then placed on a magnetic rack and supernatant was

discarded. Beads were washed with 70% ethanol 3 times and dried at room temperature

for 10 minutes. DNA samples were then eluted from dry beads using 10 mM Tris buffer

and transferred to PCR tubes.

Q5 LM-PCR master mix (ddH$_2$0, 5x Q5 reaction buffer, 10 mM dNTPs, 20 μM P1.3 primer, 20 uM P2.1 primer, 2 U/μl Q5 hot start DNA polymerase) was added to the DNA samples, pipeted to mix, and run in the thermocycler with the following program: 98°C for 30 seconds for 1 cycle; 98°C for 10 seconds, 52°C for 30 seconds, 72°C for 20 seconds for 20 cycles, 72°C for 2 minutes, 4°C hold.

Resulting PCR products between 200-500 base pairs in length were purified by agarose gel and gel extraction kit (Qiagen, 28704). Purified DNA concentration was measured (Qubit fluorometer) and sent to the Vanderbilt VANTAGE sequencing core for high-throughput next generation sequencing. Resulting sequence files were returned to our lab in .bam format and analyzed in the sequencing analysis pipeline below.

Sequencing analysis pipeline

Sequencing files obtained from the Vanderbilt VANTAGE core were analyzed using Terminal command line. The specific tools used in the following analysis pipeline were either publically available scripts or custom scripts courtesy of Byran Venters (Vanderbilt University, Nashville, TN).

Sequencing files (.bam) were sorted using a SAM Tools script called sort and the command:

(Eq. 2.2)   Time /pathtoSamtools/samtools sort
/pathtosavedoutputPrefix/ input_sorted

The script output was a sorted .bam file, which was then indexed with a SAM Tools

script called index using the following command:

(Eq. 2.3)   **Time /pathtoSamtools/samtools index /pathtosavedoutputPrefix/ input_sorted.bam**

The script output was an indexed .bai file. The following step isolated only uniquely

mapped reads from the sorted.bam file by using a SAM Tools script called with the

following command:

(Eq. 2.4)   **time /pathtosamtools/samtools view –F 4 –q 1 –hb /pathto/input_sorted.bam > ./output.sorted_uniq.bam**

The output containing uniquely mapped reads was then converted to a .bed file using a

BED Tools script called bamtobed and the command:

(Eq. 2.5)   **time /pathtobedtools /bedtools bamtobed –i /pathto/input_sorted.bam > .pathto/output_sorted.bed**

The output .bed file was then sorted using BED Tools and the command:

(Eq. 2.6)   **time /pathtobedtools /bedtools sortedbed –i ./output_uniq.bed > output_uniq_sort.bed**

The .bed file containing uniquely mapped reads was indexed using a python script called

tabs2genetrack and the command:

(Eq. 2.7)   **time python /pathytotabs2genetrack /tabs2genetrack.py –i ./output_uniq.bed –f bed –o output_uniq.idx**

The output index file was used to call peaks using a custom python script called

multiprocess_gff3 and the command:

(Eq. 2.8)   **time python /pathto/genetrack/multiprocess_gff3.py –s 20 –e 40 ./output_uniq.idx**

The output was a .gff file containing information regarding the standard deviation

around the called peaks. A non-zero number for standard deviation indicated that there

were multiple reads around a specific unique region. Standard deviations of zero were

considered 'orphan' peaks, and were not able to be paired with another peak. The .gff

file containing both orphans and non-orphans was peak paired using a custom python

script called cwpair_gff3_rzr142 and the command:

(Eq. 2.9)   **time python /pathto/cwpair_gff3_rzr142.py –u 0 –d 80 –b 3 /pathto/input_peakscalled.gff**

The output was a simplified cwpair.gff file containing the paired peaks and the

coordinates of the midpoint peak value. The locations of these paired peaks were then

mapped to a meaningful reference segment. For this analysis, they were mapped to the

distance of the nearest human genome (HG19) transcriptional start site (TSS) using a

python script called map_segment_to_segment and the command:

(Eq. 2.10)   **time python /pathto/map_segment_to_segment.py –f**
                      /pathto/referenceTSS_hg19.csv –t
                      /pathto/input_S_cwpair.gff –S 3 –E 4 –u 2000 –d 2000

The output file is .gff file that contained the information of the paired peaks that were

within the defined upstream and downstream limits of a transcriptional start site within

the human genome. The sequence information at those regions was obtained using a

python script called fastaextract and the command:

(Eq. 2.11)   **time python /pathTo/fastaextract.py –g**
                      /pathTo/compiled_mm10orHg19_build.fa –u 40 –d 40
                      /pathTo/input_peakPairMidpoint.gff

The output was a fasta.fa file that contained the actual sequences for all of the paired

peaks. This file was evaluated further in the motif qualification pipeline below.

Motif qualification pipeline

The fasta.fa files contained DNA sequence information for the regions of

occupancy of LDB1 and SSBP2. Due to the nature of the algorithms used to "peak call"

peaks in the ChIP-exo sequencing analysis pipeline, ChIP-exo positive results were

returning two types of peaks: clustered peaks and distinct peaks. The clustered peaks

were unable to be validated on target genes for either LDB1 or SSBP2. The distinct

peaks were validated using RT-qPCR on target genes. Therefore, the motif qualification was used to further analyze only distinct peaks.

The "called peaks" were sorted first by rank order. This organized the "called peaks" with the most sequence tags (strongest/highest peaks compared to background) first and the "called peaks" with the fewest sequence tags (lowest peaks compared to background) last. The top peaks were checked graphically based on the following characteristics:

    i. Must have multiple sequence tags on both the forward and reverse strand (excluded orphans and likely PCR artifacts all mapping to the same nucleotide);

    ii. The distribution of the sequence tags must span several different nucleotides;

    iii. The peak must appear to be distinct, not a series of clustered peaks

    iv. The peak must appear in a promoter or intronic region compared to a proximal gene.

This checklist was followed until ~35 of the top "distinct peaks" that were in promoter regions and the ~35 top "distinct peaks" that were in intronic regions. Three files were generated containing the DNA sequences associated with the distinct peaks (1: Promoter peak sequences; 2: Intronic Peak sequences; 3: Promoter + Intronic peak sequences). Biostatistician Yan Guo analyzed each of the three files described above and found three statistically significant and repeatedly represented motifs within each group of DNA sequences.

The 3 motifs discovered from each of the 3 DNA sequence files were compared to a variety of known motif databases using the Benso Lab STAMP analysis (alignment, similarity, and database matching for DNA motifs; University of Pittsburgh, Pittsburgh, PA; http://*et al*.benoslab.pitt.edu/stamp). The outputs from these analyses were lists of likely binding motifs that may be present within particular peak regions and the corresponding transcription factors that are predicted to bind there.

CHAPTER III

EXPRESSION ANALYSIS OF LMO4, LDB1, SSBPS

Introduction

Investigation of the relevance of SSBPs, LDB1, and LMO4 in head and neck

carcinoma began with replication of the important findings of Mizunuma *et al*., 2003.

They reported that LMO4 and LDB1 were frequently and, importantly, concordantly

upregulated in oral cavity squamous cell carcinomas and both proteins were shown to

localize to the nuclei of cells at the invasive edge of these tumors (Mizunuma *et al*,

2003). The very significant association between LMO4 and LDB1 abundance, which

would not be obligate despite their interaction in cells, prompted the hypothesis that

LMO4 and LDB1 were protected by SSBPs from ubiquitylation and subsequent

proteasomal destruction in head and neck carcinoma analogous to that mechanism

reported in erythroid cells (Xu *et al*., 2007). The expression levels of LMO4 and LDB1

in oral cavity tumors were evaluated first to confirm the published expression results, and

then extended to include SSBP2 and SSBP3. The expression levels of all four proteins

were then evaluated in primary oropharyngeal tumors and lymph node metastases.

SSBP2 expression was further analyzed using head and neck carcinoma tissue

microarrays (TMAs). Finally, LMO4, LDB1, and SSBP protein levels were measured in

human oral cavity carcinoma cell lines.
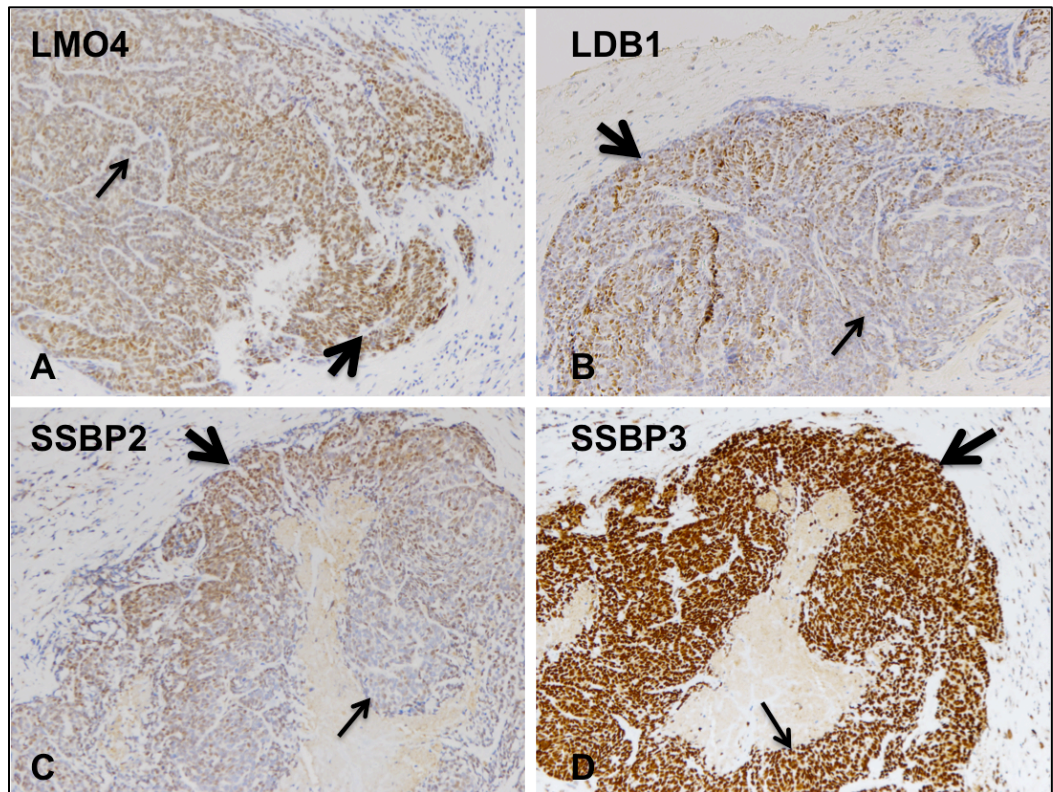
Figure 3.1 Immunolocalization of LMO4, LDB1, SSBP2, and SSBP3 in human oral cavity carcinoma. (A) LMO4, (B) LDB1, (C) SSBP2, and (D) SSBP3 protein expression was enriched at the invasive edge (thick arrows) compared to more central regions of tumor (thin arrows).

Immunolocalization of LMO4, LDB1, SSBP2, and SSBP3 in oral cavity carcinoma

Immunohistochemical analysis was carried out on sections of paraffin-embedded blocks from three males with squamous cell carcinoma of the oral cavity. Patients were between 55-58 years of age, were long-time smokers, and reported alcohol consumption that ranged from occasional to frequent. Clinical stage ranged from I-IV and tumors were moderately to well differentiated. All tumors were p16^{INK4A} negative and two had metastasized to cervical lymph nodes. Confirming published data (Mizunuma *et al*., 2003), LMO4 and LDB1 were expressed in the nuclei of carcinoma cells, with expression enriched in cells at the invasive edge (thick black arrows, Figure 3.1A and B). SSBP2 and SSBP3 also localized to nuclei of oral cavity carcinoma cells, and their expression was likewise increased in cells at tumor edge (thick black arrows, Figure 3.1C and D).

Immunolocalization of LMO4, LDB1, and SSBPs in oropharynx carcinoma

Primary oropharynx tumors

The analysis was expanded to HPV-positive squamous cell carcinomas of the oropharynx (n = 7). Samples included six males and one female aged 44-75. In contrast to the first group, five of the seven had never smoked, with alcohol consumption ranging from occasional to frequent. All tumors were p16^{INK4A}-positive, a surrogate for HPV positivity (Zhang *et al*., 1999), and three of the seven tumors had metastasized to cervical lymph nodes, which were available for study (See*: Metastatic lymph nodes*). Importantly, immunohistochemical staining for LMO4, LDB1, SSBP2 and SSBP3 (Figure 3.2A, C, E, and G) in these oropharyngeal carcinomas was identical to what was observed for the oral cavity carcinomas (Figure 3.1A, B, C, D). All four proteins localized to nuclei and were
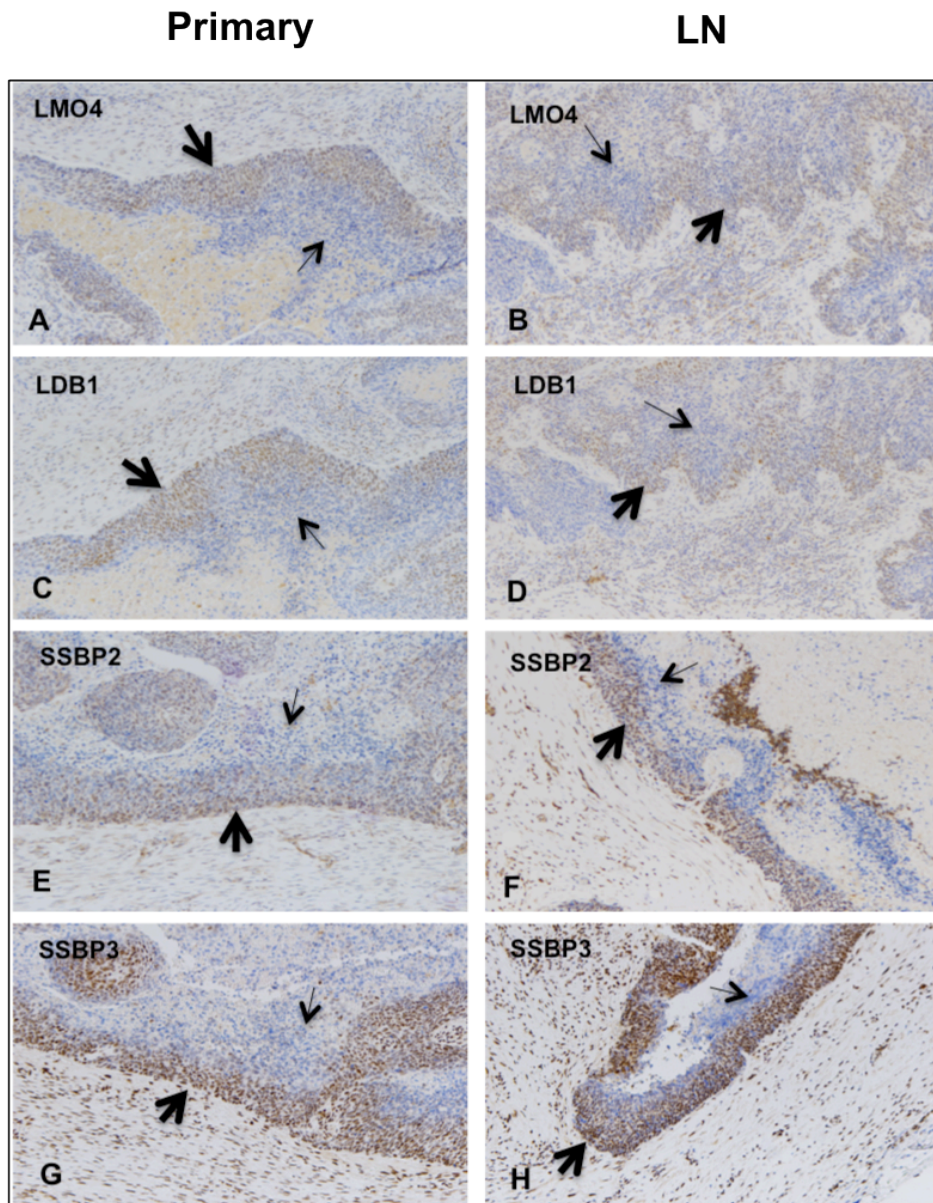
Figure 3.2 Immunolocalization of LMO4, LDB1, SSBP2, and SSBP3 in human oropharyngeal carcinoma and regional lymph nodes (LN). (A) LMO4, (C) LDB1, (E) SSBP2, and (G) SSBP3 were each increased at the invasive edge (thick arrows) compared to more central regions of tumor (thin arrows). (B) LMO4, (D) LDB1, (F) SSBP2, and (H) SSBP3 were detectable in regional lymph nodes for all four carcinomas in which paired samples were available and exhibited a similar distribution as in the primary tumor.

increased in abundance in cells at the invasive edge, indicating a shared function of these proteins in invasiveness despite the very different pathogenetic mechanisms underlying these carcinomas of the head and neck.

Metastatic lymph nodes

To assess their involvement in metastasis, paired lymph nodes and tumors (n = 3) were evaluated for expression of LMO4, LDB1, SSBP2, and SSBP3. Mirroring the expression pattern in primary tumors, each of these proteins were expressed in lymph nodes involved by tumor, with their intracellular location restricted to the cell nucleus (Figure 3.2B, D, F, and *ET AL*). For all three pairs of patient samples, the abundance of LMO4, LDB1, SSBP2 and SSBP3 was the same or higher in lymph nodes than in primary tumors, indicating that tumor cells attaining regional lymph nodes maintained if not increased their expression of LMO4, LDB1, SSBP2, and SSBP3. This lends support to the notion that these proteins regulate a genetic program involved in invasiveness in this tumor type.

Immunolocalization of SSBP2 in human head and neck carcinoma arrays

LMO4 overexpression in oral cavity carcinomas has also been reported to be associated with reduced tumor cell differentiation (Mizunuma *et al.*, 2003; Kwong, *et al.*, 2011). In order to explore whether other members of a predicted LMO4, LDB1, and SSBP-containing complex had a similar relationship with tumor cell differentiation, SSBP2 expression was analyzed in two human HNSCC tissue microarrays. The combined TMAs provided 96 tumor samples from mixed types of HNSCC, with the majority from oral cavity or oropharyngeal subtypes. Approximately 50% of poorly
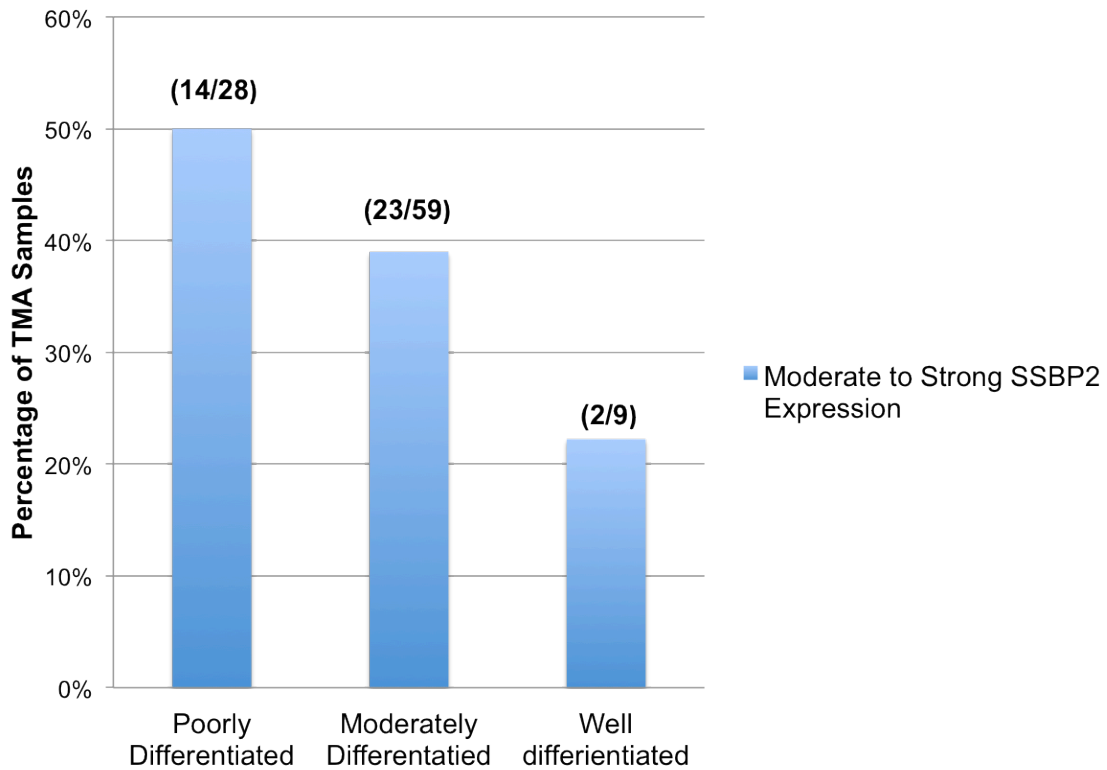
Figure 3.3 Percentage moderate/strong SSBP2 staining compared to differentiation. The percentages of moderate to strong SSBP2 staining present in various stages of head and neck tumor differentiation. Head and neck tumor samples were evaluated for SSBP2 expression with immunohistochemistry analysis. 50% of poorly differentiated tumors had moderate/strong SSBP2 staining. Moderately-differentiated and well-differentiated tumors showed a decrease (38% and 22%, respectively) in moderate to strong SSBP2 staining ($R^2 = 0.98$).

differentiated tumors had moderate-strong SSBP2 staining (Figure 3.3). Percentages of moderate-strong SSBP2 staining decreased in moderately differentiated tumors (37%) and poorly differentiated tumors (22%). Importantly, SSBP2 expression was inversely correlated with tumor differentiation in the TMA samples, which corresponds with the published LMO4-tumor differentiation relationship, and provides support that both proteins may be influencing the same tumor cell properties in HNSCC.

Expression of LMO4, LDB1, and SSBPs in oral cavity carcinoma cell lines

After analyzing LMO4, LDB1, and SSBP expression in oral cavity and oropharyngeal tumors and lymph nodes, their abundance was measured in a panel of human oral head and neck carcinoma cell lines. Applying highly quantitative immunoblot analysis to ten established oral cavity carcinoma lines, and the relative abundance of LMO4, LDB1, and SSBPs was determined. As predicted by semi-quantitative immunohistochemical analysis (Figure 3.1, 3.2), LMO4 and LDB1 showed a highly significant correlation in their protein levels ($R^2 = 0.90$, $P < 0.0001$) (Figure 3.4A). Highly significant correlations were discovered between SSBP and LDB1 abundance (SSBP2 *vs.* LDB1, $R^2 = 0.95$, $P < 0.0001$; SSBP3 *vs.* LDB1, $R^2 = 0.94$, $P < 0.0001$) (Figure 3.4B) and between SSBP and LMO4 abundance (SSBP2 *vs.* LMO4, $R^2 = 0.80$, $P < 0.0005$; SSBP3 *vs.* LMO4, $R^2 = 0.84$, $P < 0.0002$) (Figure 3.4C). Taken together with published data from transfected cells showing SSBPs protect LDB1 and LMO2 and LMO4 from proteasomal destruction (Xu *et al*., 2007), these results are compatible with SSBPs coordinately regulating turnover of LMO4 and LDB1 in this tumor type.
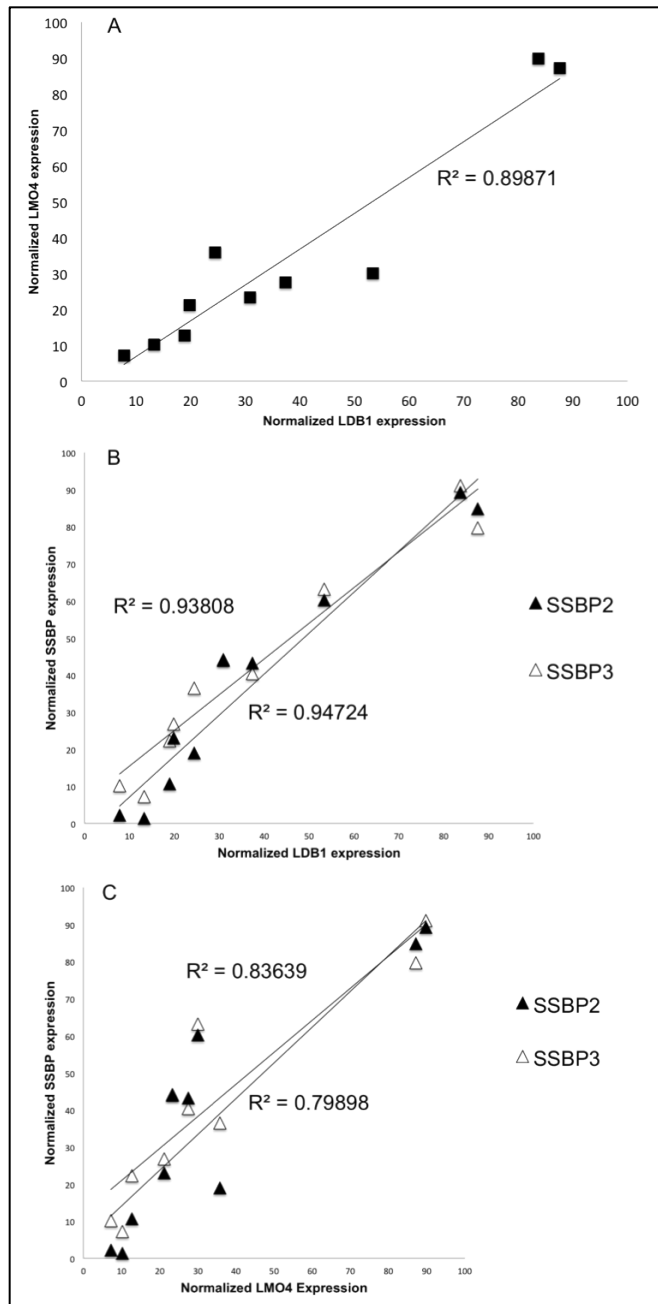
Figure 3.4 Relative expressions of LMO4, LDB1, SSBP2, and SSBP3 in panel of 10 human oral cavity squamous cell carcinoma lines. The relative abundance of LMO4, LDB1, SSBP2, and SSBP3 were found to be significantly interrelated using immunoblot analysis. (A) LMO4 and LDB1 abundance were significantly correlated ($R^2 = 0.90$, P < 0.0001, n = 10). (B) SSBP and LDB1 abundance were significantly correlated (black triangles SSBP2: $R^2 = 0.95$, P < 0.0001, n = 10; white triangles SSBP3: $R^2 = 0.94$, P < 0.0001, n = 10). (C) SSBP and LMO4 abundance were significantly correlated (black triangles SSBP2: $R^2 = 0.80$, P < 0.0005, n = 10; white triangles SSBP3: $R^2 = 0.84$, P <0.0002, n = 10).

Conclusions

The expression analyses in this chapter resulted in many important findings. First, the published results of LMO4 and LDB1 having concordant upregulation in oral cavity squamous cell carcinomas and localization to the nuclei of cells at the invasive edge of these tumors were confirmed. Second, SSBP2 and SSBP3 were also found to have concordant expression and co-localization with LMO4 and LDB1 in both oral cavity tumors and oropharyngeal tumors, and all four proteins of interest had maintained or increased expression in lymph node metastases. Third, SSBP2 was found inversely correlated with tumor differentiation, similar to reported correlation between LMO4 and reduced tumor cell differentiation. Finally, the concordant expression of LMO4, LDB1, and SSBPs was also present across a panel of human oral cavity carcinoma cell lines. Together, these findings support the hypothesis that LMO4, LDB1, and SSBPs are important for the progression of head and neck carcinoma.

CHAPTER IV

EFFECTS OF *LDB1* GENE-TARGETED HUMAN OCC CELLS

Introduction

After expression analyses were completed for LMO4, LDB1, and SSBPs (Chapter III) and correlations between the protein levels and tumor phenotypes were observed, loss of function studies were executed to measure direct biological and mechanistic effects of the proposed complex. A brief invasion analysis was conducted to determine whether the abundance of these proteins correlated with cellular invasiveness. In general, the invasive potential of the cell lines measured correlated with levels of LMO4, LDB1, and SSBPs (data not shown). Indeed, the most invasive cell line, VU-SCC-1729, exhibited the highest levels of LMO4, LDB1, SSBP2, and SSBP3 and was selected for use in loss-of-function studies. LDB1, which serves as an adapter protein that simultaneously contacts LMO proteins and SSBPs through their LIM interaction domain and LDB1/Chip conserved domain, respectively, was selected for targeting based on the hypothesis that a loss of this protein would best disrupt the putative DNA-binding complex(es) containing LMO4, LDB1, and SSBPs (Chapter 1, Figure 1.2).

CRISPR/cas9 mediated deletion of *LDB1*

CRIPSR/cas9-mediated gene targeting was used to disrupt both exons 1 and 2 of *LDB1* to maximize the chance of successful reduction in the encoded protein. Whereas targeting of exon 1 resulted in only a partial reduction in LDB1 expression, targeting of

52
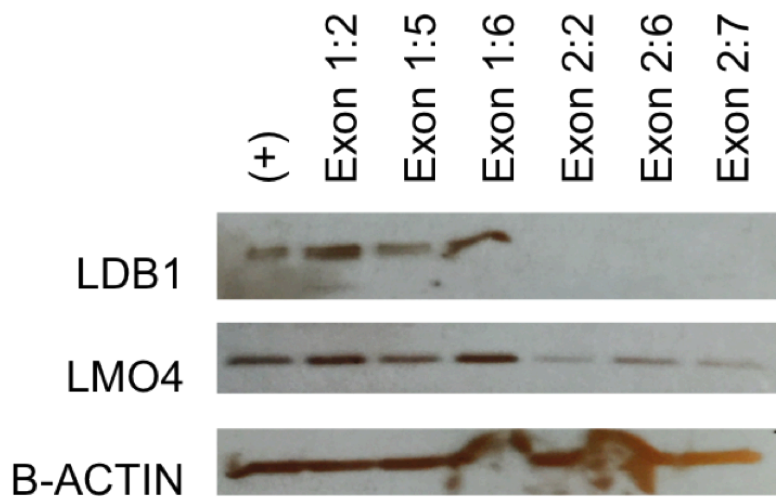
Figure 4.1 Reduced LDB1 expression in VU-SCC-1729 cells decreases LMO4 abundance. LDB1 and, to a lesser extent, LMO4 protein abundance were reduced in VU-SCC-1729 cells in which the LDB1 levels were reduced by CRISPR-cas9-mediated deletion (VU-1729 Exon 2: clone 2; clone 6; clone 7) but not in cells in which gene targeting was not successful (VU-1729 Exon 1: clone 2; clone 5; clone 6).

exon 2 resulted in complete loss of expression (Figure 4.1). The reduction in LDB1

protein levels in exon 2-targeted cells was also associated with significant reduction in

LMO4 expression (Figure 4.1) as predicted by previous studies (Xu *et al.*, 2007).

Loss of LDB1 reduces proliferation of OCC cells *in vitro*

The developed cell lines with successful loss of LDB1 (and as a result, reduced

LMO4) were used to test whether reduction in LDB1 and LMO4 had an effect on cell

proliferation or viability. A significant and progressive decrease in the number of LDB1

knockout cells (VU-1729-2:7) adherent to tissue culture plastic compared to the parental

line was detected beginning on day 2 without a change in cellular viability (Figure 4.2) (P

= 0.053), implicating LDB1 and LMO4 in tumor cell proliferation.

Loss of LDB1 reduces invasion of OCC cells through matrigel

To investigate whether these proteins had some role in invasion, VU-SCC-1729

LDB1 knockout and parental cell lines were evaluated for their ability to traverse a layer

of basement membrane in Boyden chambers *in vitro*. Compared to both the parental line

and the vector control line (VU-SCC-1729+), the LDB1 knockout line VU-1729-2:7 was

much less able to penetrate Matrigel (P = 0.0011 and P = 0.0057, respectively). In accord

with expression data, the invasiveness of the exon 1-targeted line VU-1729-1:2 was

significantly reduced compared to the parental line (P = 0.0114) but less so compared to

the exon 2-targeted line (Figure 4.3). Thus, reduction of LDB1 and LMO4 protein

abundance and, by inference, an LDB1/LMO4/SSBP transcriptional complex, decreased

cellular invasiveness in reconstituted basement membrane.
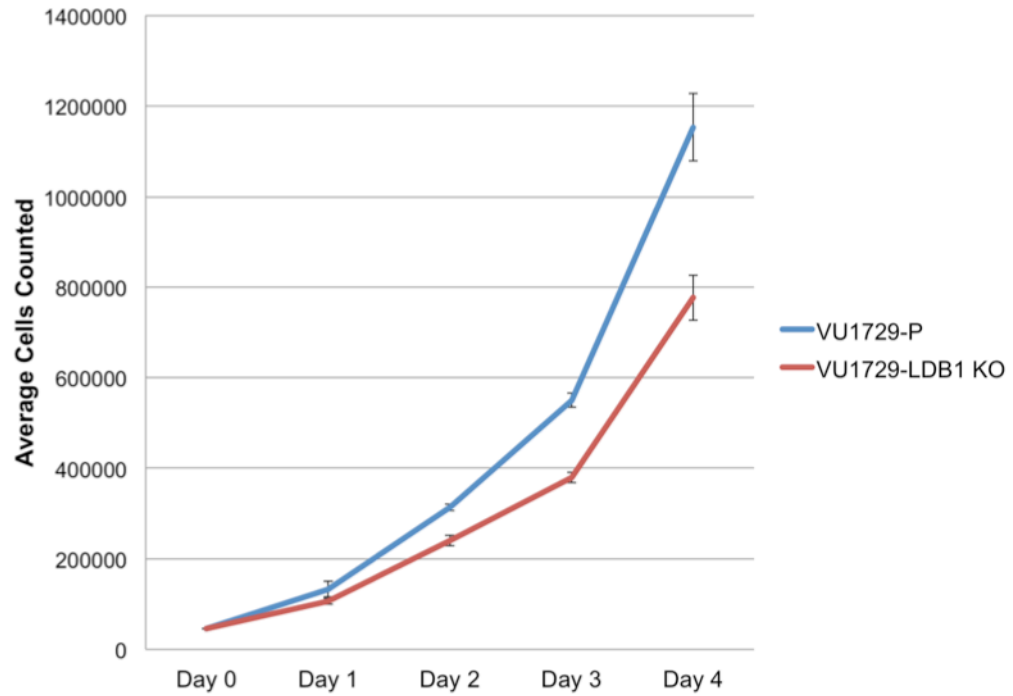
Figure 4.2 Loss of LDB1 expression in VU-SCC-1729 cells reduced growth in vitro. Cells with a reduction in LDB1 protein abundance (VU-1729-2:7, red line) also had decreased growth rate in vitro compared to VU-SCC-1729 cells (blue line) beginning on Day 2 ($P = 0.053$).
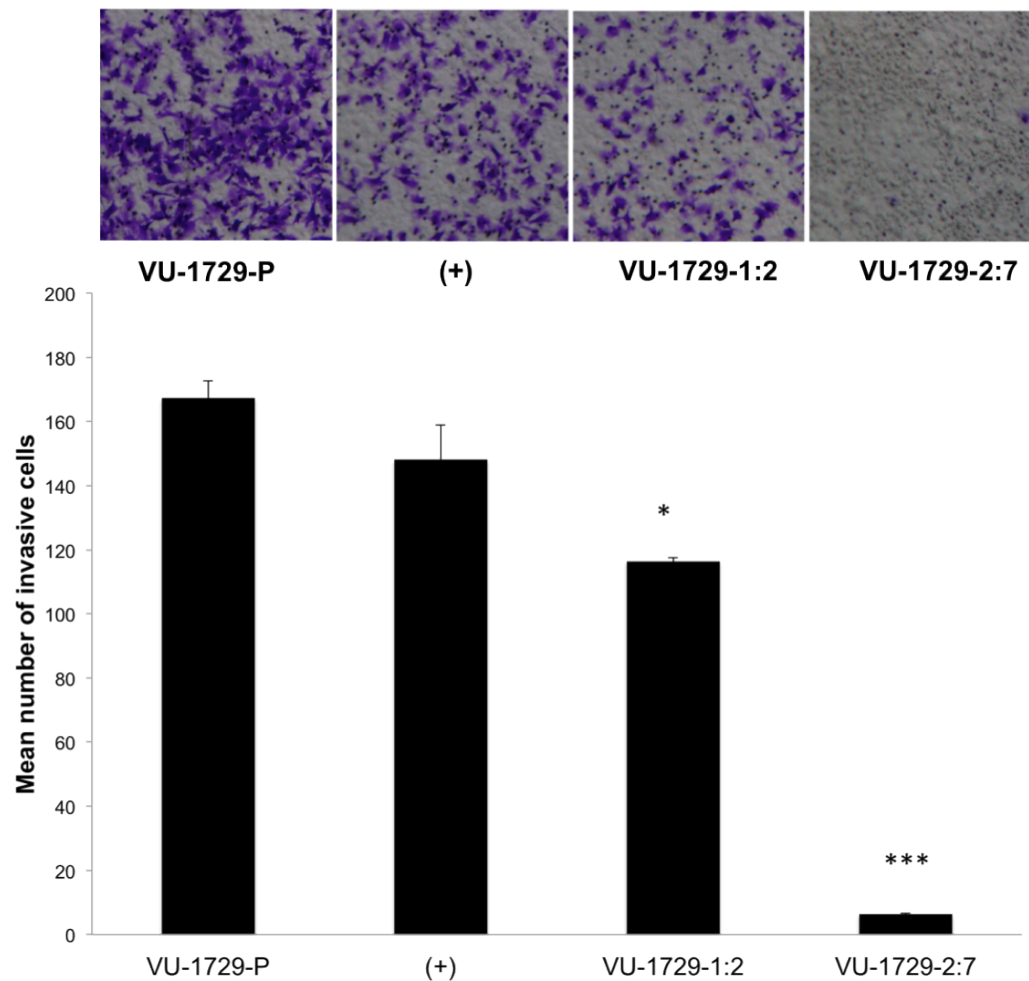
Figure 4.3 Loss of LDB1 expression in VU-SCC-1729 cells significantly reduces cellular invasiveness in 2-D Matrigel invasion assay. Invasion through a layer of reconstituted basement membrane (Matrigel) was compared for VU-SCC-1729 cells (VU-1729-P), a knockout vector control (+), a cell clone with only modest LDB1 knockdown (VU-1729-1:2), and cells in which LDB1 protein expression was effectively eliminated (VU-1729-2:7).

Loss of LDB1 reduces OCCs invasion through organotypic co-cultures

To evaluate invasiveness in a more physiologic context, parental (VU-SCC-1729) and LDB1 knockout (VU-1729-2:7) cell lines were tested in an organotypic reconstruction permitting two-dimensional visualization of the oral cavity carcinoma cells invasive into a surrounding cells in addition to basement membrane. Tumor cells were co-cultured on a a mixture of embryonic fibroblasts, collagen, and Matrigel as previously described (Andl *et al*., 2010). For both lines cells emanating from the epithelial layer and invading the fibroblastic layer as individual cells were noted (Figure 4.4, thick arrows) characteristic of epithelial-mesenchymal transition (EMT) in which cells lose their affinity for and attachment to each other and gain invasive functions. In the parental line, in addition, clusters of cells still adherent to each other (Figure 4.4, thin arrows) were seen invading the fibroblastic layer, a distinct form of invasiveness termed collective cell migration (or invasion). Importantly, the number of these clusters was very significantly reduced in the LDB1 knockdown cell line VU-1729-2:7 compared to parental cells ($P < 0.0001$) (Figure 4.4). Together, data from two different assays of tumor cell invasiveness showed that depletion of LDB1 and LMO4 significantly reduced that function and suggest that invasion of oral cavity carcinoma cells through epithelium and extracellular matrix is critically dependent on LDB1 protein abundance.
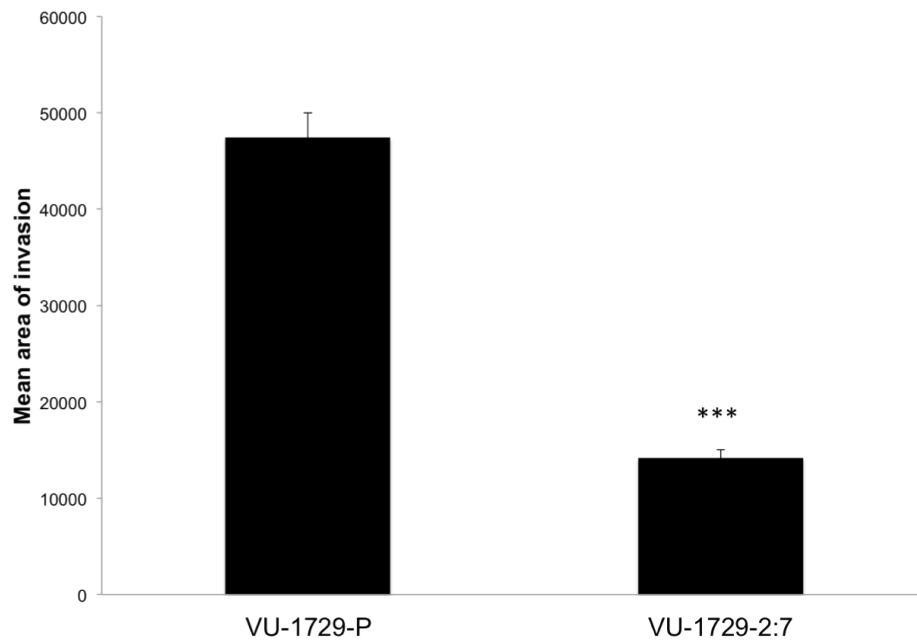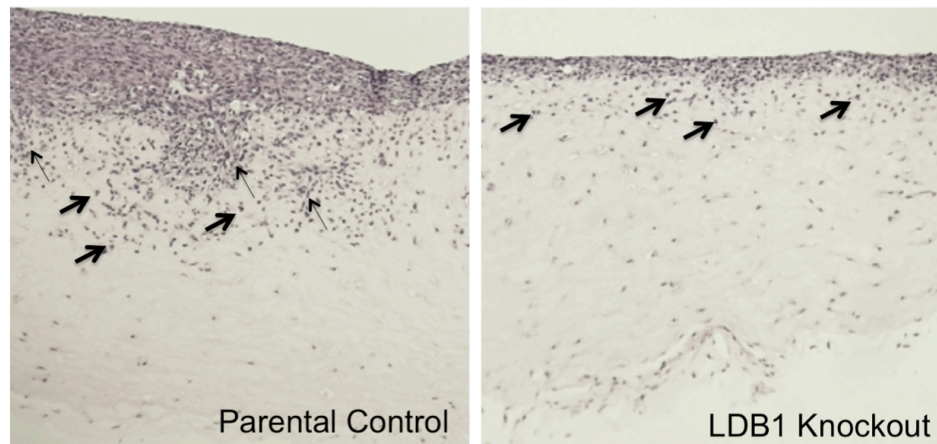
Figure 4.4 Loss of LDB1 expression in VU-SCC-1729 cells significantly reduces cellular invasiveness in 3-D co-culture assay. Invasion through a fibroblast monolayer and into extracellular matrix was compared for parental and VU-1729-2:7 cells. The highly invasive cell line VU-1729-P showed significant penetration of the collagen/fibroblast matrix by clusters of cells (thin arrows) and individual cells (thick arrows). LDB1 knockout cells showed significantly reduced invasion, with only a few individual cells invading noted (thick arrows) and no larger collections of cells. *, P < 0.01; ***, P < 0.0001.

Tumor xenografts of *LDB1* gene-targeted human OCC cells

To better evaluate the function of the LIM-only and LIM domain interacting protein in head and neck carcinoma cells *in vivo*, limiting numbers of parental and LDB1 knockout (VU-1729-2:7) cells were injected into nude mice and tumor volumes measured over a period of 23 days. Although tumor volumes did not differ significantly over the first five days ($P = 0.766$), tumors derived from LDB1 knockout cells were significantly different on Day 10 ($P = 0.002$) and were progressively smaller than those derived from parental cells on Day 13 ($P = 0.00096$), Day 20 ($P = 0.0000194$), and Day 23 ($P = 0.0000345$) (Figure 4.5A). Thus, LDB1 and LMO4 protein abundance are also critical for tumor growth *in vivo*.

To determine whether this difference in tumor growth affected apoptosis, proliferation, or both, tumors were harvested on Day 23, sectioned, and stained with antibodies to caspase-3 and PCNA, respectively. Tumors derived from LDB1 KO cells did not stain differently from parental tumors in expression of the apoptotic marker caspase-3 ($P = 0.2406$) (Figure 4.5B) but did exhibit significantly lower labeling with antibody to the proliferation marker PCNA (Figure 4.5C). This was entirely consistent with the reduced proliferation of LDB1-deficient VU-SCC-1729 cells *in vitro* (Figures 4.3 and 4.4).

Effect of LDB1 deletion on tumor vascularization

In addition to being smaller in size, tumors derived from the LDB1 KO cells appeared less vascularized compared to those derived parental VU-SCC-1729 cells (see Figure 4.5A insert).

A

B

C
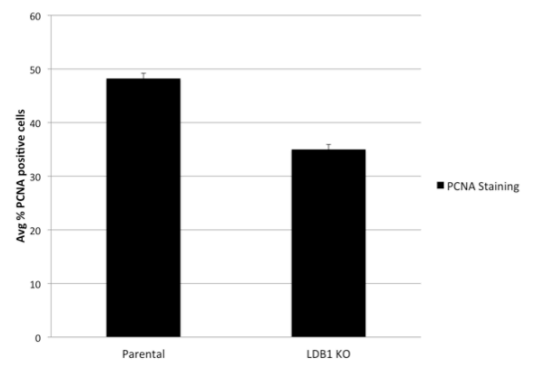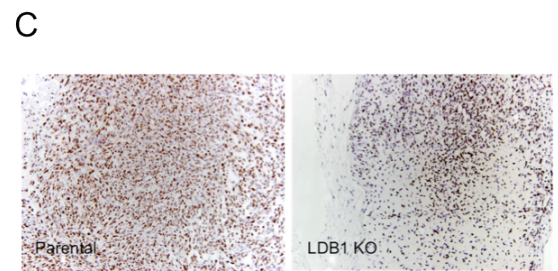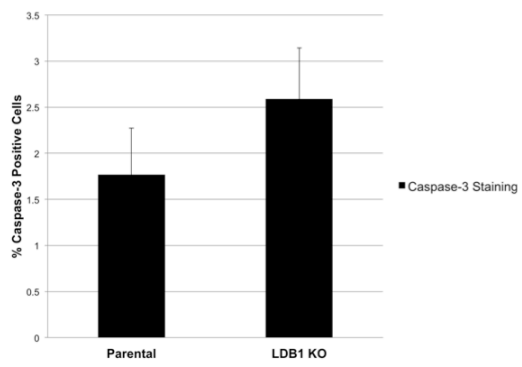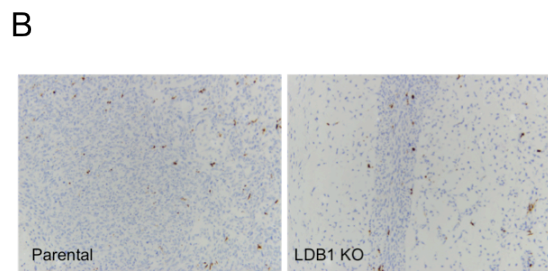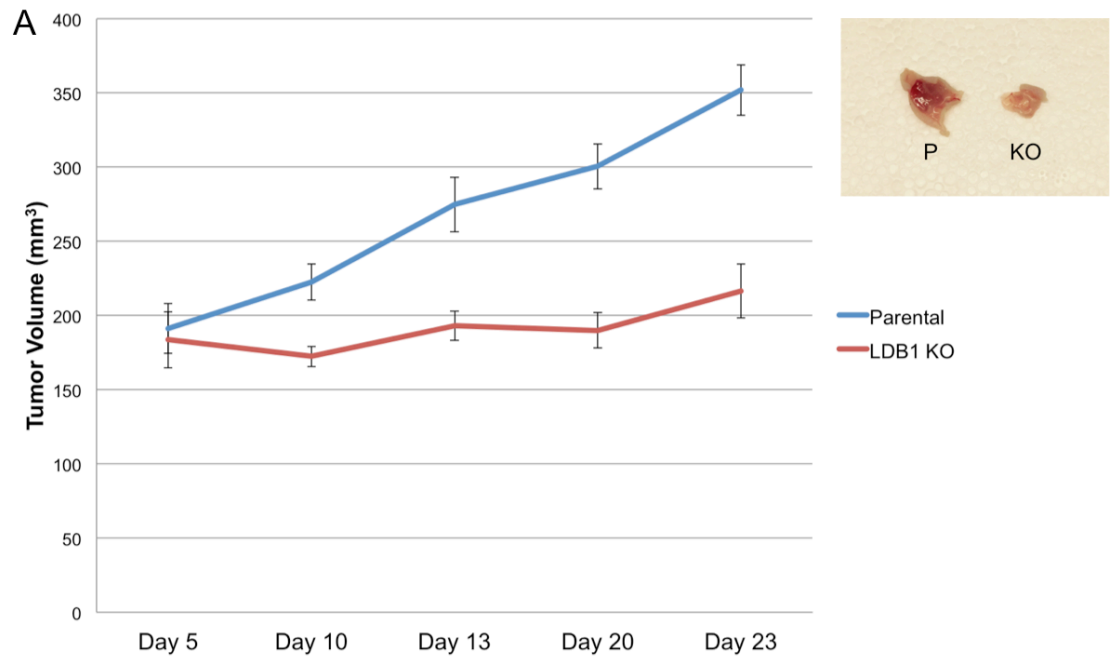
Figure 4.5 (opposite page) Loss of LDB1 expression in VU-SCC-1729 cells significantly reduces tumor growth in nude mice. (A) Tumor growth in nude mouse xenografts was compared in VU-SCC-1729 cells and cells in which LDB1 protein expression was effectively eliminated (LDB1 KO). Tumor volumes were compared beginning on Day 5 (P = 0.766). A difference in tumor size was detected on Day 10 (P = 0.002), with this difference becoming progressively greater from Day 13 (P = 0.00096), to Day 20 (P = 0.0000194), to Day 23 (P = 0.0000345). (B) Caspase-3 staining for apoptosis was not significantly different in LDB1 KO *vs.* parental tumors (P = 0.2406). (C) PCNA staining for proliferation was significantly reduced in LDB1 KO tumors compared to tumors derived from parental VU-SCC-1729 cells (P = 0.00028).

To evaluate this observation, the two sets of tumors were stained with an antibody for an endothelial cell marker, Von Willebrand Factor (vWF). Indeed, LDB1 KO tumors exhibited significantly decreased vWF staining compared to parental tumors ($P = 0.00196$) (Figure 4.6). Collectively, the results of these xenotransplantation assays demonstrated that loss of LDB1 protein, and presumably the transcriptional complex(es) to which it contributed in association with LMO4 and SSBPs, reduced the growth and vascularization of tumors derived from human head and neck carcinoma cells.

RNA-seq analysis of *LDB1* gene-targeted human OCC cells

In order to begin an investigation on what genetic targets were altered by the loss of LDB1, and therefore potentially regulated by a proposed complex containing LMO4, LDB1, and SSBPs, RNA-seq analysis was carried out on LDB1 knockout (VU-1729-2:7) and control (VU-SCC-1729-P) cell lines. These RNAs were ranked according to the level of significance of the difference in their abundance between knockout and parental lines and then grouped according to the cellular processes to which they potentially contributed. The iPathway analysis found 135 RNAs that were significantly changed with a threshold set of $P < 0.05$ and log fold change of 2. The differentially expressed genes discovered in the screen were found to contribute to a number of processes important in the promotion or spread of cancer including: transcriptional misregulation in cancer, ECM-receptor interaction, TGF-ß signaling pathway, PI3K-Akt signaling pathway, pathways in cancer (Figure 4.7).
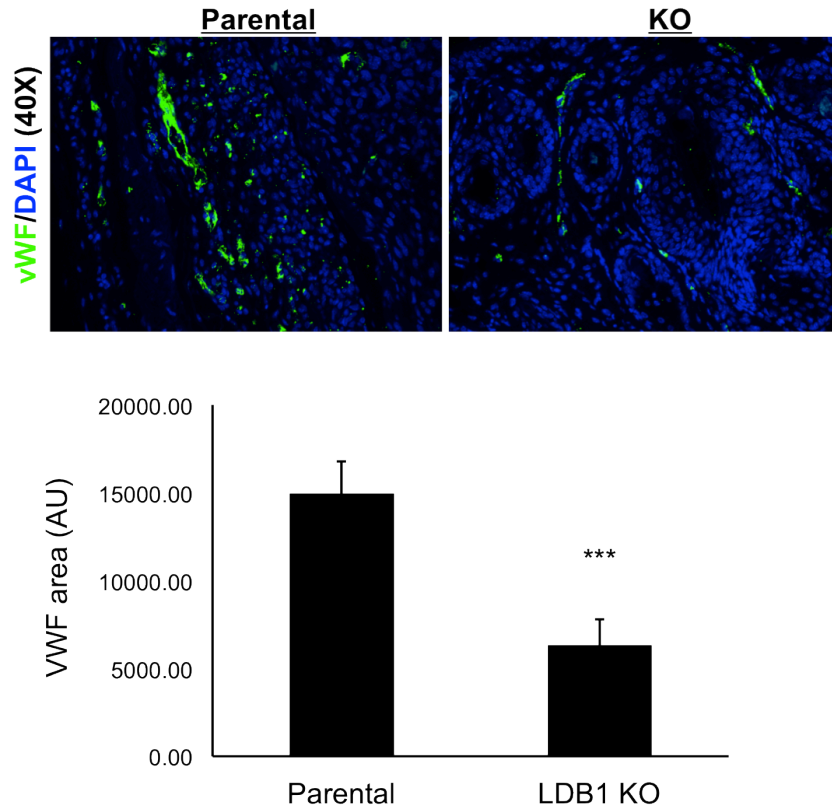
Figure 4.6 Loss of LDB1 expression in VU-SCC-1729 cells significantly reduces angiogenesis in nude mice. Endothelial cells in nude mice xenografts were compared for VU-SC-1729 (parental) and VU-1729-2:7 cells in which LDB1 protein expression was effectively eliminated (LDB1 KO). Endothelial cells are stained with Von Willebrand Factor (VWF) (green immunofluorescence) and nuclei are stained with DAPI (blue immunofluorescence). Angiogenesis was significantly decreased in LDB1 KO tumors (p = 0.00196).

Conclusions

Loss of function studies targeting LDB1 in the invasive oral carcinoma cell line resulted in several important findings regarding the biological impact of a proposed complex containing LMO4, LDB1, and SSBPs. First, as predicted, a loss of LDB1 also reduced LMO4 protein levels. These oral cavity cells with a loss of LDB1, and subsequent LMO4 reduction, retained cell viability but experienced a decreased proliferative advantage. The same cells were less invasive in two different *in vitro* invasion assays. Xenografts comparing control cells with LDB1 reduced oral cavity cells showed loss of LDB1 resulted in a growth and vascularization disadvantage. These biological results, in combination with changes in RNA-expression in response to a loss of LDB1, highlight the direct role of LDB1, LMO4, and likely SSBPs in a variety of pathways important for the progression of HNSCC.
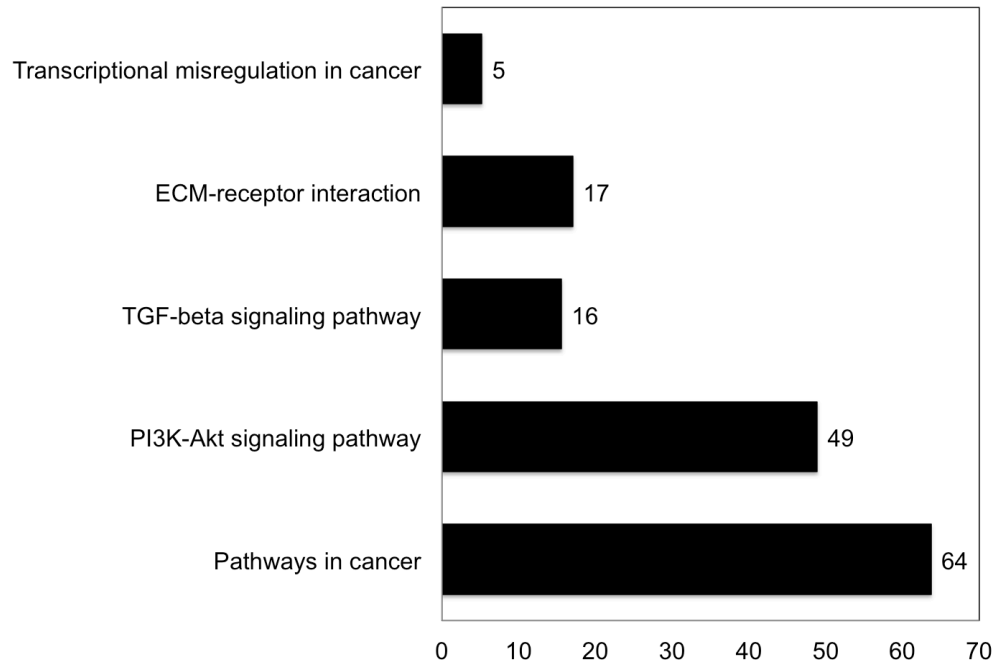
Figure 4.7 Significantly impacted pathways identified with iPathway Guide for differentially expressed genes in VU-SCC-1729 vs. VU-1729-2:7 cells. Numbers represent percentage of genes in each category that were differentially expressed using a threshold of 0.05 for statistical significance and 2 for absolute log expression change. P-values corresponding to the pathways are as follows: ECM-receptor interaction (P = 0.0000527), pathways in cancer (P = 0.001), TGF-beta signaling pathway (P = 0.002), transcriptional misregulation in cancer (P = 0.002), PI3K-Akt signaling pathway (P = 0.003).

CHAPTER V

SSBP2 AND LDB1 DNA-OCCUPANCY IN HUMAN OCC CELLS

Introduction

In order to determine whether LDB1 and SSBP2 may contribute to a common
DNA-binding complex that regulated transcriptional targets involved in proliferation,
invasiveness, and angiogenesis, genome-wide ChIP-exo was performed on VU-SCC-
1729-P cells. This experiment, which by design allows for the generation of a next
generation DNA sequence library containing all locations that were bound by the
antibody used, allowed for genome-wide analysis of DNA occupancy of both LDB1 and
SSBP2 with base-pair resolution. There were many regions that contained signals from
both LDB1 and SSBP2, and those peaks were investigated further. There were also two
distinct types of peaks that were isolated by this technique, the first was classified as
'clustered peaks' and the second was classified as 'distinct peaks'. Both peak classes
contained overlapping signals from LDB1 and SSBP2 antibodies, and were further
validated and investigated through conventional ChIP analysis.

Genome-wide ChIP-exo of SSBP2 and LDB1 in OCC cells

Clustered peaks

The first of the two peak classes that was isolated through the sequencing analysis
pipeline was termed 'clustered'. Clustered regions contained both large numbers of
LDB1 and SSBP2 reads and had multiple single peaks directly proximal to one another
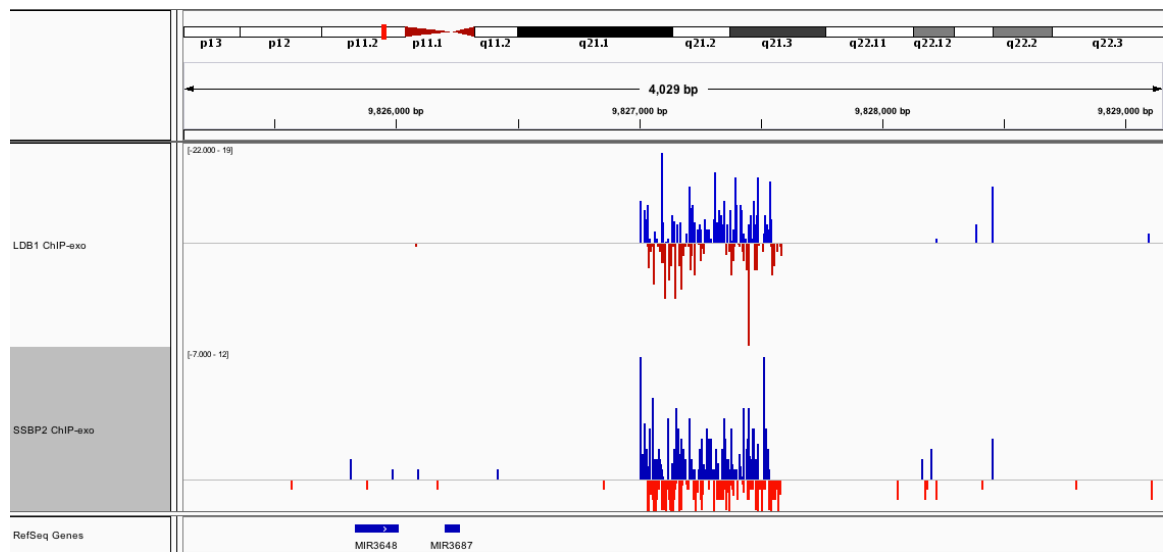
Figure 5.1 Clustered peak in the *MIR3687* promoter. Shown is ~4 Kb window comparing LDB1 (top track) and SSBP2 (middle track) ChIP-exo results in relation to the Hg19 reference genome (bottom track). Signals derived from the forward strand (blue) and from the reverse strand (red) are clustered one after another in approximately 8 clustered peaks upstream of *MIR3687*.

(Figure 5.1). Graphically, these clusters appeared to be stretches of forward and reverse reads that spanned several hundred base pairs. In many cases, the clustered peaks were within the boundaries of potential regulatory regions of genes. A large clustered peak is shown upstream of two miRNAs, *MIR3687* and *MIR3648*, surrounded by smaller orphan reads or non-called peaks (figure 5.1). A second large clustered peak was identified both upstream of *CDC27*, and within an exon of *CDC27* (data not shown). These three clustered peaks (*MIR3687, CDC27, CDC27* exon) were tested in conventional ChIP assays to validate LDB1 and SSBP2 occupancy. Importantly, none of the clustered peaks showed increased occupancy relative to normal IgG (figure 5.3) and were eliminated from further analysis.

Distinct peaks

The second of the two peak classes that was isolated through the sequencing analysis pipeline was termed 'distinct'. Like clustered peaks, distinct peaks contained large numbers of both LDB1 and SSBP2 reads, however they did not have multiple runs of forward and reverse reads directly proximal to one another (Figure 5.2). Graphically, these distinct peaks were defining a much smaller region (~150 base pairs) and were within the boundaries of potential regulatory regions of genes. A large distinct peak is shown upstream of *CDH11* (figure 5.2). The distinct peak in the promoter of *CDH11*, and a control region within an exon of *CDH11*, were tested in conventional ChIP assays to validate the distinct peak. The distinct peak upstream of the *CDH11* TSS had increased occupancy compared to both normal IgG and to the control exon region (figure 5.3). The analysis continued with only data derived from distinct peaks.
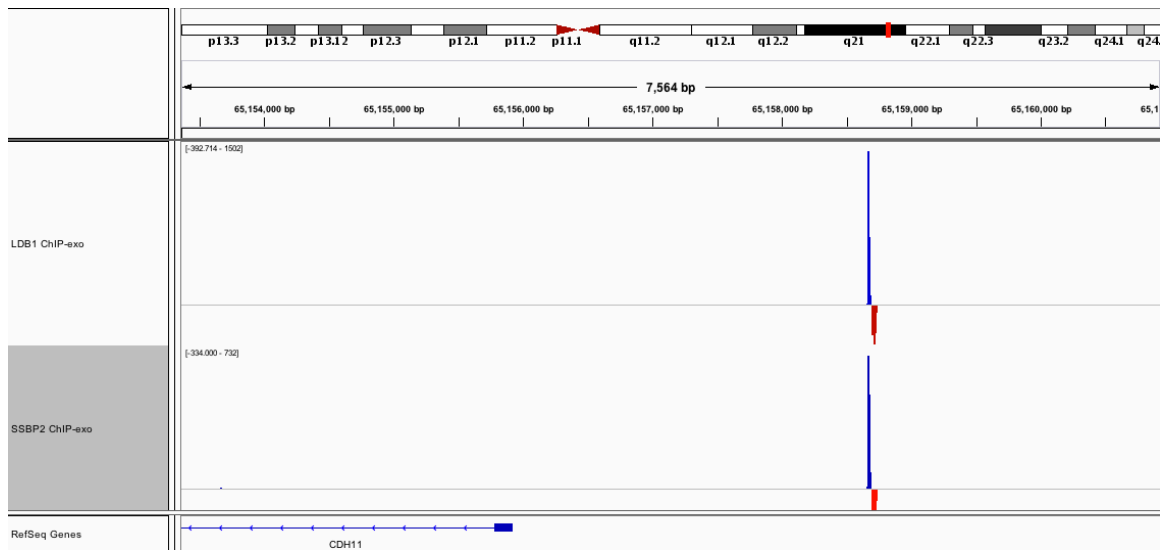
Figure 5.2 Distinct peak in the *CDH11* promoter. Shown is ~7.5 Kb window comparing LDB1 (top track) and SSBP2 (middle track) ChIP-exo results in relation to the Hg19 reference genome (bottom track). Signals derived from the forward strand (blue) and from the reverse strand (red) are in one distinct location upstream from the *CDH11* transcriptional start site.
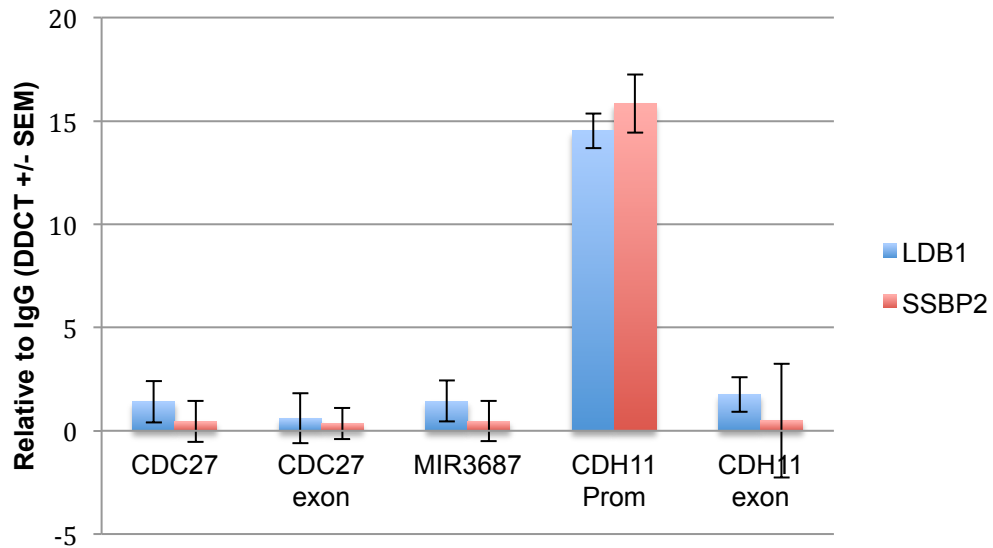
Figure 5.3 Validation of clustered and distinct peaks. LDB1 and SSBP2 ChIP-exo signals were validated using conventional ChIP analysis followed by qPCR. Probes were designed to validate clustered peak in *MIR3687* promoter and distinct peak in *CDH11* promoter regions. *CDC27, CDC27* exon, and *CDH11* exon probes were designed as negative controls.

Validated target genes

CDH11 promoter

CADHERIN-11 (CDH11), a member of the cadherin superfamily whose normal role is maintaining cell-cell adhesions in osteoblastic differentiation, has also been found to be overexpressed in 15% of breast cancers and essential for tumor progression in others (Assefnia *et al.*, 2014). Although the importance of CDH11 in head and neck carcinoma is unknown, the link to general tumor progression and importance in cell-cell adhesion make it an interesting potential target for a LDB1, SSBP2-containing complex in this disease.

A distinct ChIP-exo peak was found in the promoter region of *CDH11*, located approximately 2.5 kilobases upstream of the TSS. Forward (blue) and reverse (red) reads from LDB1 and SSBP2 ChIP-exo sequencing results were graphed in the integrated genomics viewer (IGV, Broad Institute) (Figure 5.4A). The same viewing window was opened using UCSC ENCODE genome browser (Figure 5.4B). Viewable within the ENCODE window are two tracks, layered activating marks and DNA hypersensitivity regions, which can be seen all along the promoter region of *CDH11*. Although the signals for activating marks are decreasing as they approach the region occupied by LDB1 and SSBP2, the DNA hypersensitivity marks remain present there, indicating that the peak is located in an open and active transcriptional region. A third track open within the ENCODE genome browser is the transcription factor ChIP-seq results, which graphically shows where different transcription factors have been found to bind (Figure 5.4B). No known transcription factors within the ENCODE ChIP-seq database appear

Figure 5.4 Distinct peak upstream of *CDH11* promoter aligned with UCSC Encode genome browser. Window displayed is chr16:65,155,397-65,159,185. (A) LDB1 (top track) and SSBP2 (middle track) ChIP-exo results in relation to the Hg19 reference genome (bottom track). Signals derived from the forward strand (blue) and from the reverse strand (red) are in one distinct location upstream from the *CDH11* transcriptional start site. (B) Encode browser tracks from Top to Bottom: UCSC genes, layered H3K27ac activating marks, DNase hypersensitivity regions, and transcription factor ChIP-seq data.

directly in the peak region, however, there are many transcription factor binding locations proximal to the peak.

CKAP2L promoter

Cytoskeleton Associated Protein 2-Like (CKAP2L) is a mitotic spindle protein critical for neural stem and progenitor cells. CKAP2L has been linked to cell cycle progression in neural progenitor cells and deletions affecting CKAP2L have been associated with various cancers including head and neck carcinoma (Klivmov, 2013). These links made it an interesting potential target for a LDB1 and SSBP2-containing transcriptional complex in this disease.

The distinct peak found in the promoter region of *CKAP2L* was located approximately 4 kilobases upstream of the TSS. Forward (blue) and reverse (red) reads from LDB1 and SSBP2 ChIP-exo sequencing results were graphed in the IGV (Figure 5.5A). The same viewing window was opened using UCSC ENCODE genome browser (Figure 23B). For this promoter, the ENCODE track marking layered activating marks (H3K27ac) has a small signal peak proximal to the LDB1 and SSBP2 peak. The track showing the DNA hypersensitivity regions is offset slightly, although still very proximal to the corresponding ChIP-exo peak (Figure 5.5B). The third track open within the ENCODE genome browser indicating known transcription factors within the ENCODE ChIP-seq database appear directly in the peak region (Figure 5.5B). Together, these three tracks indicate likelihood that the LDB1/SSBP2 ChIP-exo peak is located in an open and active transcriptional region.
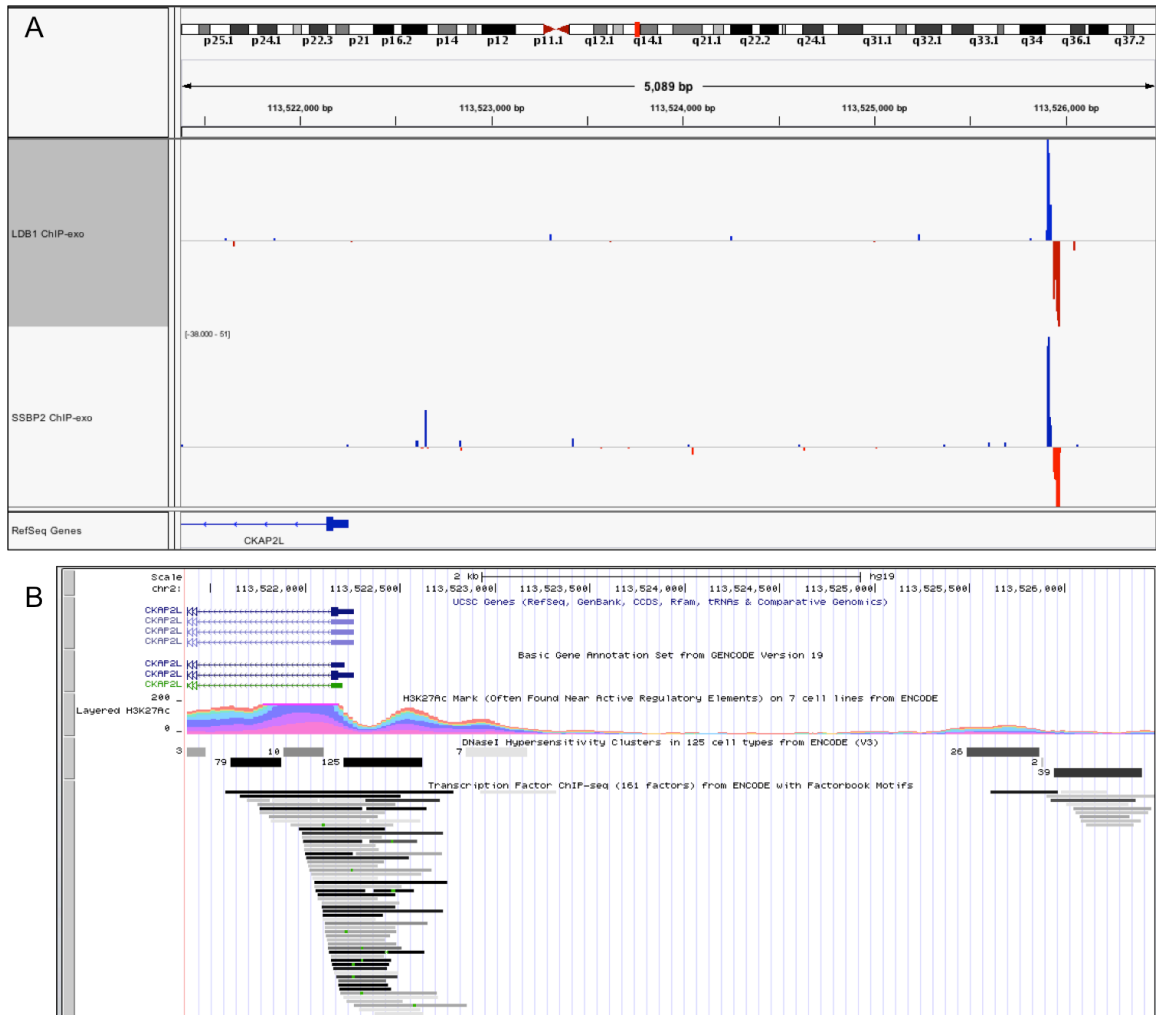
Figure 5.5 Distinct peak upstream of *CKAP2L* promoter aligned with UCSC Encode genome browser. Window displayed is chr2:113,521,377-113,526,475. (A) LDB1 (top track) and SSBP2 (middle track) ChIP-exo results in relation to the Hg19 reference genome (bottom track). Signals derived from the forward strand (blue) and from the reverse strand (red) are in one distinct location upstream from the *CKAP2L* transcriptional start site. (B) Encode browser tracks from Top to Bottom: UCSC genes, layered H3K27ac activating marks, DNase hypersensitivity regions, and transcription factor ChIP-seq data.

74

HMGA2 intron

High Mobility Group AT-Hook 2 (HMGA2) is an architectural factor and essential component of the enhancesome, which is normally expressed only in early development. The expression of HMGA2 has been associated with a variety of poor-prognosis tumors, and has been implicated to promote breast cancer metastasis and recurrence of squamous cell carcinoma, although the mechanism is unknown (Boo *et al*., 2005). These associations made it an interesting potential target for a LDB1 and SSBP2-containing transcriptional complex in this disease.

The distinct peak for *HMGA2* differed from the first two promoter peaks in that it was found in the second intron. Forward (blue) and reverse (red) reads from LDB1 and SSBP2 ChIP-exo sequencing results were graphed in the IGV (Figure 5.6A). The same viewing window was opened using UCSC ENCODE genome browser (Figure 5.6B). For this intronic peak, the first ENCODE track marking layered activating marks (H3K27ac) has a small signal peak directly in line with the LDB1 and SSBP2 peak. The second track showing the DNA hypersensitivity regions is also directly in line with the ChIP-exo peak (Figure 5.6B). The third track depicting the transcription factor ChIP-seq database results shows several binding regions directly in line with peak region (Figure 5.6B). Together, these three tracks indicate a strong likelihood that the LDB1/SSBP2 ChIP-exo peak is located in an open and active transcriptional region.
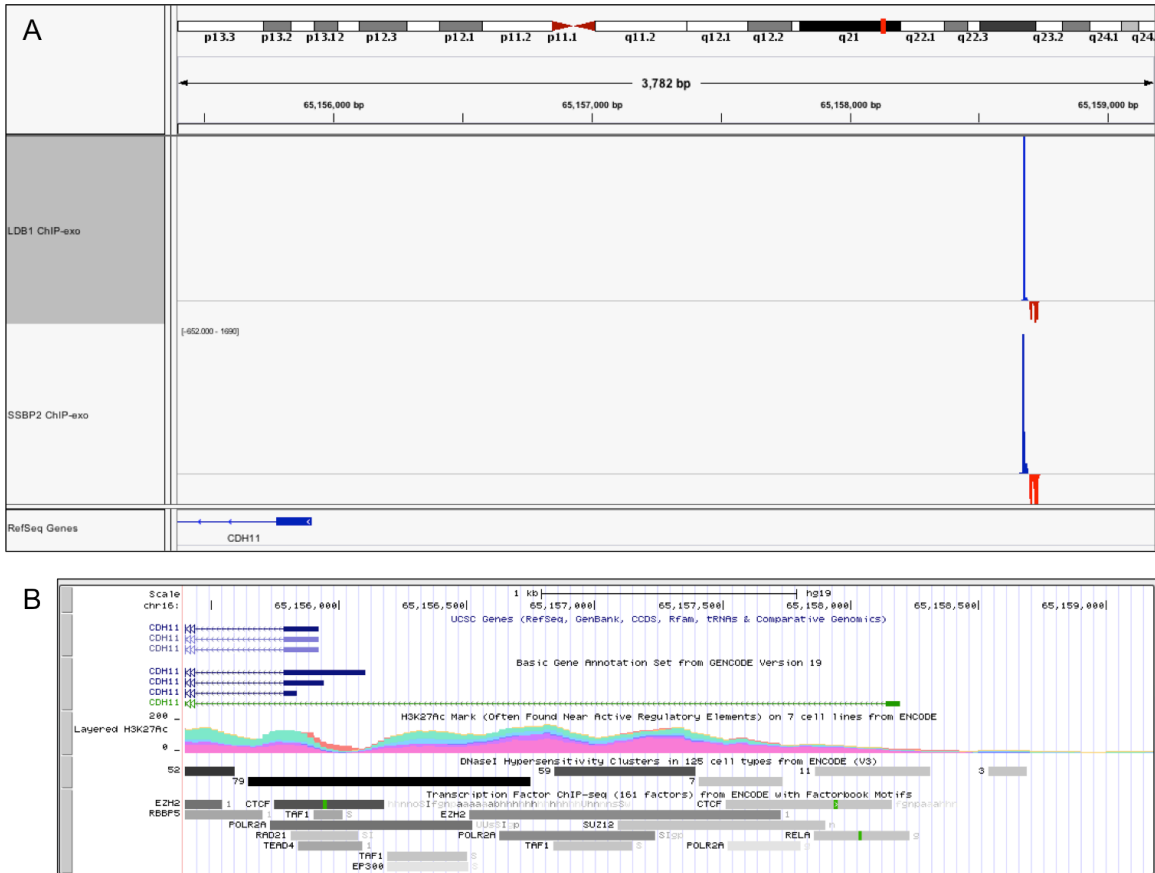
Figure 5.6 Distinct peak upstream of *HMGA2* intron aligned with UCSC Encode genome browser. Window displayed is chr12:66,271,330-66,274,135. (A) LDB1 (top track) and SSBP2 (middle track) ChIP-exo results in relation to the Hg19 reference genome (bottom track). Signals derived from the forward strand (blue) and from the reverse strand (red) are in one distinct location inside the *HMGA2* second intron. (B) Encode browser tracks from Top to Bottom: UCSC genes, layered H3K27ac activating marks, DNase hypersensitivity regions, and transcription factor ChIP-seq data.

Target gene occupancy reduced in LDB1 KO cells

Several factors made the *CDH11*-promoter, *CKAP2L*-promoter and *HMGA2*-intron promising target regions for validation of the ChIP-exo experiments. First, all three distinct peaks were co-occupied by LDB1 and SSBP2 (Figures 5.4A, 5.5A, 5.6A). Second, all three distinct peaks were in or proximal to regions of activating acetylation marks and DNA hypersensitivity when compared with the UCSC ENCODE genome browser (Figures 5.4B, 5.5B, 5.6B). Third, all three distinct peaks were in line or proximal with at least one reported transcription factor binding site based ChIP-seq data in a variety of cell lines as reported in the UCSC ENCODE genome browser (Figures 5.4B, 5.5B, 5.6B). Together, these factors contributed support for the existence of a multiprotein complex containing both LDB1 and SSBP2 that can bind to these various regulatory regions. The conventional ChIP assay was carried out in the invasive head and neck carcinoma cell line VU-SCC-1729 cells, where endogenous levels for LDB1 and SSBP2 levels were previously found to be high (Chapter III, figure 3.4). As predicted, *CDH11*-promoter, *CKAP2L*-promoter, and *HMGA2*-intron were all found to have increased LDB1 and SSBP2 occupancy compared to normal IgG in VU-SCC-1729-P cells (Figure 5.7), ultimately validating the ChIP-exo and sequencing analysis pipeline results.

Next, the occupancy of LDB1 and SSBP2 at the *CDH11* promoter, *CKAP2L* promoter, and *HMGA2*-intron was evaluated in VU-1729-LDB1 KO cells (figure 5.7). The loss of LDB1 in the invasive head and neck cell line did, in fact, reduce the occupancy of both LDB1 and SSBP2 in all three of the target regulatory regions.
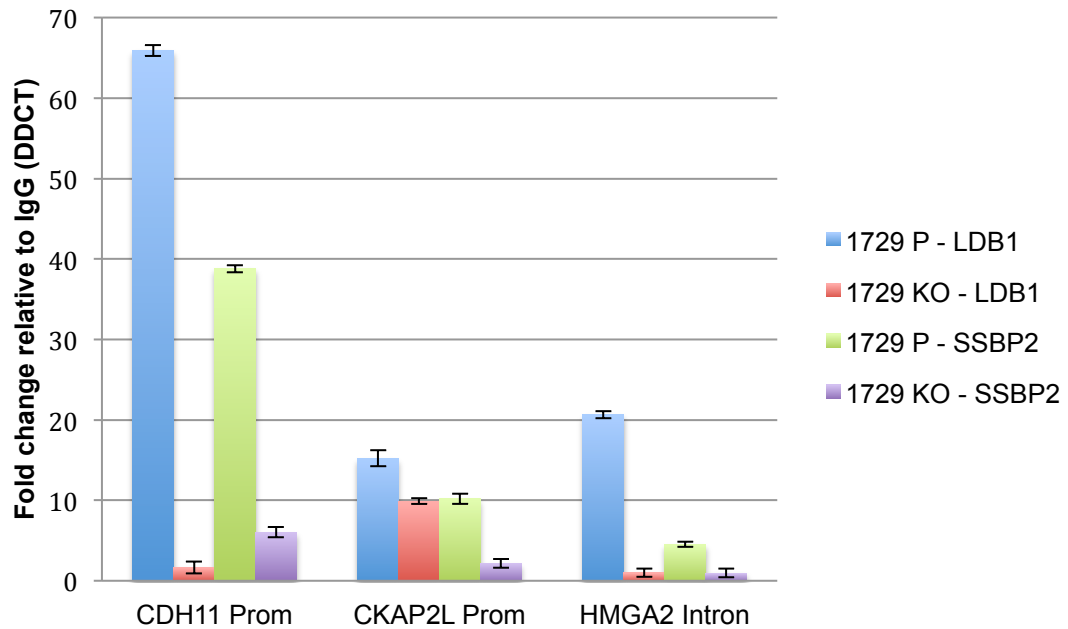
Figure 5.7 LDB1 and SSBP2 occupancy is reduced in VU-1729-LDB1 KO cells. Three distinct peaks discovered in the ChIP-exo analysis (*CDH11* promoter, *CKAP2L* promoter, *HMGA2* intron) were validated in conventional ChIP analysis using VU-1729-P cells for LDB1 (blue bars) and SSBP2 (green bars). VU-1729-LDB1 KO cells also evaluated for occupancy at the same regions. Both LDB1 (red bars) and SSBP2 (purple bars) had reduced occupancy in the LDB1 KO cells compared to the parental cells.

These results reinforce the likelihood of LDB1 and SSBP2 participating in a common DNA-binding complex in head and neck carcinoma cells.

Conclusions

Genome-wide ChIP-exo analysis of LDB1 and SSBP2 occupancy in the invasive head and neck carcinoma line VU-SCC-1729 had several important findings. LDB1 and SSBP2 were found to co-occupy two subtypes of peaks. Although clustered peaks looked promising due to the large number of total reads in the region and for the pattern of forward and reverse reads in both LDB1 and SSBP2 assays, they were unable to be validated in conventional ChIP analysis. This lead to the conclusion that clustered peaks are artificially enriched background signals with no probable biological significance. The co-occupancy of LDB1 and SSBP2 for three distinct target peaks was, however, validated with conventional ChIP analysis. Furthermore, the occupancies of LDB1 and SSBP2 were significantly reduced in VU-1729-LDB1 KO cells compared to VU-SCC-1729-P cells. Overall, these validation results support the genome-wide ChIP-exo and subsequent sequencing analysis pipelines, and supported the use of those data in further studies for motif discovery (Chapter VI).

CHAPTER VI

IDENTIFICATION OF SSBP2 AND LDB1 BINDING LOCI

Introduction

One of the key methods for investigating gene regulatory regions and the multiprotein complexes that bind them is motif analysis. DNA motifs are amino-acid sequence patterns that are widespread, non-random, and have biological significance. Many transcription factors regulate multiple genes by binding to a common sequence motif within a given gene's regulatory regions. The proximal promoter, a standard regulatory region, is located directly upstream of the transcriptional start site (TSS) of a gene and contain many different motifs to which factors bind to signal transcription initiation. The distal promoter and intronic regions of a gene can also contain motifs that recruit regulatory factors. Unlike the proximal promoter elements, these regions can be several kilobases upstream or downstream of the TSS.

There are many combinatorial ways that these regulatory regions and the factors that bind to them can alter gene expression. Although LMO4, LDB1, and SSBPs are adapter that stabilize multiprotein transcriptional complexes, they themselves do not directly contact DNA. Additionally, the DNA-binding components of complexes containing LMO4, LDB1, and SSBPs in head and neck carcinoma cells have yet to be determined. In order to investigate how increased LDB1 and SSBPs (and presumably LMO4) in the cell are influencing a pro-oncogenic phenotype in head and neck

carcinoma, the DNA sequence motifs to which those proteins may be recruited to target genes were evaluated.

Here within, motif analysis was used to discover patterns within the sequence in locations that were occupied by LDB1 and SSBP2 in the invasive head and neck carcinoma cell line VU-SCC-1729 as indicated by the genome-wide ChIP-exo studies described in Chapter V. To discover motifs, biostatistician Yan Guo (Vanderbilt University, Nashville, TN) used statistical motif analyses to determine base-pair patterns within groups of sequences. The newly discovered motifs were then compared to a database containing known transcription factor binding motifs (Jaspar families database). The motifs and their locations were then considered for a potential regulatory role in promoting a pro-oncogenic or invasive phenotype. The motifs were also directly compared with changes in RNA expression when LDB1 expression is lost, and with validated ChIP gene targets.

Motif discovery

Promoter peak motifs

The DNA sequences from the top 37 co-occupied, distinct promoter peaks were selected from the LDB1 and SSBP2 ChIP-exo results (Table 6.1). These sequences were compared to one another using bio-statistical analyses to determine the top three significantly repeated motifs (biostatistician Yan Guo, Figure 6.1). These three motifs were inputted into STAMP analysis (algorithm written by Shaun Mahony, University of Pittsburgh, Pittsburgh, PA), which compared each of the discovered motifs to existing transcription factor databases. The results of the STAMP analysis were familial motif

trees that highlighted the most similar known motif with the discovered motif counterpart (Figure 6.1). The discovered promoter peak motifs included well-known E-box motifs and a homeobox motif. Specific transcription factors that were linked to these binding motifs included C2H2, MEF2A, and RUNX1 (Figure 6.1).

Intronic peak motifs

The DNA sequences from the top 33 co-occupied, distinct intronic peaks were selected from the LDB1 and SSBP2 ChIP-exo results (Table 6.2). These sequences were compared to one another using biostatistical analyses to determine the top three significantly repeated motifs (biostatistician Yan Guo, Figure 6.2). These three motifs were inputted into STAMP analysis, which compared each of the discovered motifs to existing transcription factor databases. The results of the STAMP analysis were familial motif trees that highlighted the most similar known motif with the discovered motif counterpart (Figure 6.2). Like the promoter peaks, the discovered intronic peak motifs also included well-known E-box motifs. Homeobox motifs, however, were not statistically represented in the intronic peak regions. Specific transcription factors that were linked to these binding motifs included FOXD1, ARNT, and RUNX1 (Figure 6.2).

Gene	Promoter Peak DNA Sequence
CDH11	AATGTTCATAGCAGCCTTATTTATAATAGCCAAAATCTGGAAACAACACAAATACCCAT
RANGAP1	AGCATGGTGGCTCACAACCAACTGTAATGCCAG
PRPF38B	AATGACAATAAGCACCTGAAAAAATGTTCAACATCATTAG
LMXA1	GTGGAAAAAGGCCTGAGTTCAAATCCCAGCACTGACATTA
ANKRD6	CACATGAAACAAGGCAAAAGACACCATCAATTAACAAAAGAGGAAACAGGACTAGC
SMIM10	AGCAATCCTCCTGTCTCTGCCTCCTGAGTGCTGGGATTACAGGTGTGAGC
TSIX	ACTCTTTATCCAATTTGTTAGTCTATGTCTTTTAATTGGGGCACTTAGCC
MAPK10	CTGGAGCTAGACAAGGATGCCCACTCTCACCACTCCTTTTCAACATAGTACTGGAAGTCCTA
PAQR5	GATTGGGCATGGTGTCTCTTCACAGCCATGGGC
BBOX1	GAAAACAGTATGGTGGTTCCTCAGAAAATTGAAAATAGAGCTCT
CKAP2L	TTTATTTGCCATTCATCAGTTGATGGACATCTAGGTTCTTTCCACTTTTTGGCTATTATGAATAATGCTGTTATGAACT
SLC5A9	TCAGCATCATTTGTTGAAAATGCTGTCATTTTTCCACTGA
IFNA2	TGGTCATGATGAATGATTGTTTTAATGTGTTCTTGAATTCAGTTTG
UBR7	GATATGTTGCCCTGGCTGTTCTGGAACTCCTGACC
SNORD11B	GAAAGGAGGCCAGGTGTGGTGGCTCAAGCCTTTAATCCCAGCACTTGGGAGGCTGAGGCAGGCGGATCACCTGAGGTCA
PIK3IP1	CCTGCCTCAGCCTCCCAAGTAGCTGGGATTAAAGGTGTGCGCCACCACGCCTGGCTAAATTTTTTTT
CALB2	CTACTCTTAACCACTGAGTCATCCTGCCT
PIGW	GTCAGGAGTTTGAGACCAGCCTTGTCTACAGAGTGAGACTCCATCTCTAAAAT
ZFP2	CATCTCGCCATGTTGCCCAGGCTGTCCTGGAACTTGCGGTCTTGAGTGA
CDKN2AIP	AGACAGGGTCTTGCCATGTAGACTAGGCTGGCCTTGAACCCCTGGGCTCAA
ERICH1	CACACAATCACATCATTAGATGCAGACAAAGCATTTGACAAAATCCAACACCCATTCATGATTTAAA
CRISP1	CCTCCAGTTTCATTCTTTTTACTCAGAATAGTTTTTGCTATTCTAGGTATTTTGTGGTTTCATATAAATTTTAGGA
ADAM7	CATTCTTTTTACTCAGAATAGTTTTTGCTATTCTAGGTATTTTGTGGTTTCATATA
ELTD1	GAACAGTAGCCCCAGTGTAGAGGAATGCCCGTGAATGGAG
CLCN1	GAACAGTAGCCCCAGTGTAGAGGAATGCCCGTGAATGGAG
ITGB3	GACTAGGGTGGGAGATGGGCTGGAGAGATGGGTCAGGGCCAGTTGTCTGGCAT
SNAR-H	AATGGAAACACAGCATACCAAAACCTTTGGGACACAGGGAAAGCAATACTAAGAAGGAAGCTTATAGGTATAAATGCCT
ARHGEF12	AATGGAAACACAGCATACCAAAACCTTTGGGACACAGGGAAAGCAATACTAAGAAGGAAGCTTATAGGTATAAATGCCT
TM9SF4	GATTTACATTTCCCTGATGATTAAGGATGTTGAGCACCTTT
IL36A	GCATCAGGAACTAGAAAATATCCTGAGTGTTGGGCAGTG
MYOM3	CAGGAGTTTGAGGCTAGCCTGGTCAACAGGGTGAAACCTG
GNLY	TTTTTTTTTTTTTTTTAAAGATTTATTCATTTTATTGC
KY	CTGTCTCTCTGCTATGTCTGAAGACAGCTTCG
PRSS38	CTTTTCCTCTAAAATCAGGGACAAGACAAGGATGCCTACT
MTF2	ACACAAATGTTCATAGCAGTCTTATTTATAATAGCCCCAG
CCDC125	ATGGATGGAACTAGAAAACATTATGCTAA

Table 6.1 Promoter peak DNA sequences. Gene names and associated DNA sequences for distinct ChIP-exo peaks found in the promoter regions of listed genes. E-box motifs (red), homeobox motifs (blue).
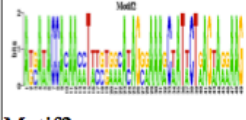
Figure 6.1 Promoter discovered motifs. (Top) Top three discovered motifs in the list of promoter peaks. (Bottom) Motif Tree comparing known transcription factor motifs to discovered motifs. The known transcription factors associated with the discovered motifs were C2H2, MEF2A, and RUNX1.

Combined promoter and intronic motifs

After analyzing the promoter and intronic peak regions separately, the two lists of sequences were combined and re-analyzed. The re-evaluation was used to determine if there were other motifs that were originally below the significant threshold due to low numbers of starting sequences. The combined DNA sequences from tables 1 and 2 (69 total sequences) were analyzed together to determine the top three significantly repeated motifs among them (biostatistician Yan Guo, Figure 6.3). The three discovered motifs were, once again, inputted into STAMP analysis and compared to a known transcription factor database (Jaspar Families database). The STAMP analysis familial motif tree highlighted the closest similar known motif with the discovered motif counterpart (Figure 6.3). As was reported for the both the individual promoter and intronic motif familial trees, E-box motifs were found (Figure 6.3). Importantly, by combining the promoter and intronic peak sequence lists, the resolution to see the homeobox motifs, which were present in the promoter peak sequence list, was lost. The specific transcription factors motifs included FOXD1, RUNX1 and C2H2 (Figure 6.3).

Gene	Intronic Peak DNA Sequence
CCNYL1	AGTCTCACTGTGTTGACCAGGCTGGCCTCGAACTCCTGG
PLEK2	ACAGGTGGTTTTCAGTTTTTCGCCACATAAATGA
TMEM62	TGCGGTGGGACCCTTGCGGGTGTTGGGCCCCAC
IGSF11	AAAAAAAATCACAGAAAATGAGAAAATATCTT
LOC253044	CCCTCCAGCAAACACAGAAGTGGCTGCTGTGCC
FAM89A	GGCATCCGAATGGCCTCTCCTCCCATTGAGATGGTATTAC
LINC00630	TCACATGGGTAGCTCTATTTTCAGTTTTCTGAGGAACCTCCAAACTGTTC
ELMSAN1	AGAGGCAAAGCTGAGCAAGCCAGGGAAAGGGCCTG
FLJ27354	TTTAGGGTCTCACTCTGTAGATCAGGCTGGAGTGCAATGT
PLEKHO2	TACAGGCGTGCGCCACCACCGCCCGGCTATTT
C14orf142	ATATGTTGCCCTGGCTGTTCTGGAACTCCTGACCT
PINK1	TAAAATGCAGTGCCCTTAACCACTGAGCCAGGCTGCCTCC
TMOD3	TGATTGAAATGTGGCTGCCCTTGCATTCACCGG
RCC1	TGCTTTCTCGTATTTTTGTGATTTGGGTCATTTTTATTTT
C1orf177	CAAGTGTGGTACTGAACCTCTGAACCTGAAGGTTTTAAGG
C4orf19	GCCTGCTCTTAACCATTGAGCCATCCTGTGT
ITGB7	AAGGTAGAAAGAACCCAGATGTCCATCAACTGATGAATGGATA
MYO19	AGGTCAGGAGTTTGAGACCAGCCTTGTCTACAGAGTGAGACTCCATCTCT
HIP1	TGAGGTCAGGAGTTCGAGGCCAGCCTGGTCAACAAGTGAGACCCTGTCTC
PLEKHG6	CTCTGTCCCCTTACAGGTTCAGAGGTTGAGCTATTCCTGCCCC
ACOT4	TAACCAGTTTCCAAGTTTTTCAGTGATTTTAAGAT
MMP20	CCTGAATTAACAGTGTGTATTTTCATTTTCCTTAGTTATGAGTT
F2RL3	CGGAGGTGGTGATGGTGAGTGGTCCTGGCTTTGGGGTG
TNFSF4	GAGGATAAGCTGTCTTCAGACACAAGAGAATCGTTACTTT
TATDN1	CTCCCCTTCCCCAGAAGTGGGAGTGGGTGGGAGGGACTGAAACTTCC
CPXCR1	TCCACCATGGAATTCAGTAAGAATGTTCCCTGCCCATAAT
BMP10	GGCATGGCTCTCAAGACCAGCTATATCTGGGTCTTAACC
IGSF21	GTGTGAGCTTGTGTCACTGTGGGTGTGAGGTGTTAGGGGG
ADCYAP1R1	TGACAGCCACCCCCACAGTGACACACAGAGAGGGAATGA
FANCA	TGGGATTAAAGGCGTGCGCCACCACACCC
RDX	AACTTCATCTTCAATCATTAATCAAGAAAATCTGTACACAAACA
NRD1	GCACTTTGAGAGGCAGAGGCAGGCAGATCTCTTGAGGTCA
CAP1	TCCATAGAATTGGAGGCTGATTATGGGAATGAAGAGGTAA
HMGA2	AAGGAATGCAGGTGAGATGGCTCAGCTGATAAACTCCCGGCCT

Table 6.2 Intronic peak DNA sequences. Gene names and associated DNA sequences for distinct ChIP-exo peaks found in the intronic regions of listed genes. E-box motifs (red), homeobox motifs (blue).
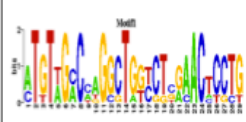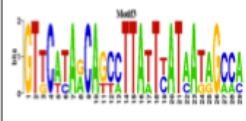
86

Motif 1
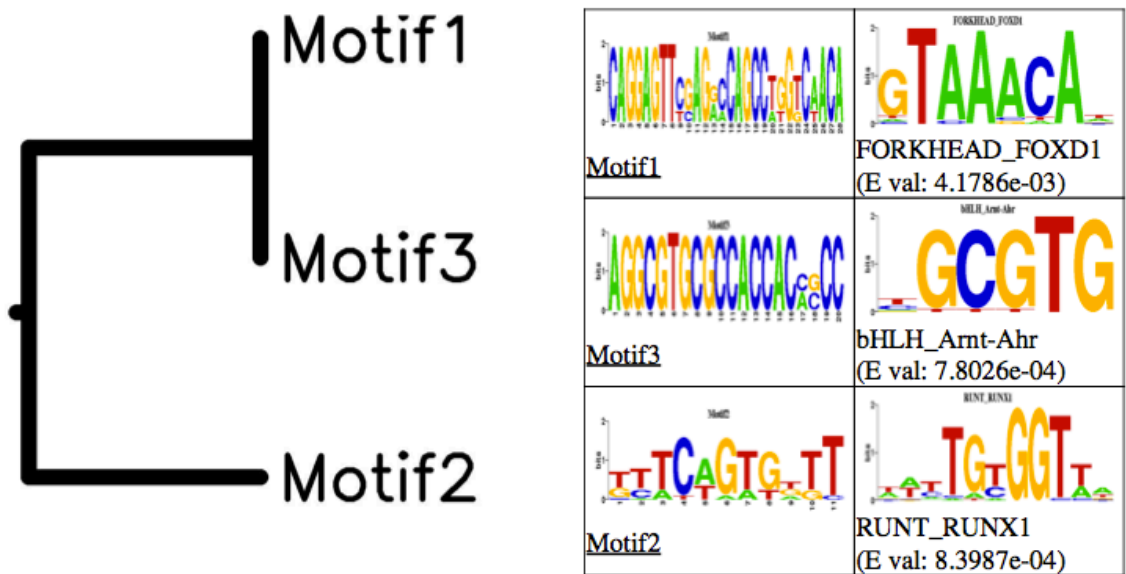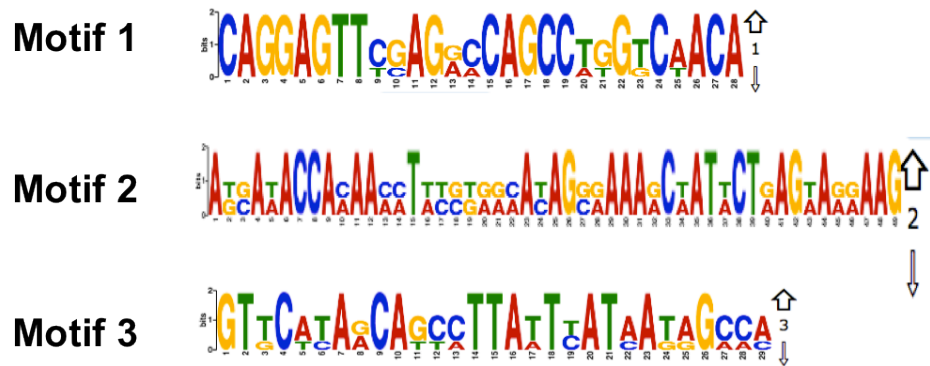
Motif 2

Motif 3

Figure 6.2 Intronic motifs. (Top) Top three discovered motifs in the list of intronic peaks. (Bottom) Motif Tree comparing known transcription factor motifs to discovered motifs. The known transcription factors associated with the discovered motifs were FOXD1, Arnt-Ahr, and RUNX1.

E-box and homeobox motifs

The two well known motifs, E-box and homeobox, were found within the discovered motifs by comparing them to a known transcription factor database (Jaspar Families Database). Each sequence was then analyzed individually to determine which actual target gene regions contained an E-box or homeobox motif. The peak sequences were divided into three motif groups: a group that contained E-box motifs, a group that contained homeobox motifs, and a third group that contained sequences that did not have either an E-box or a homeobox motif called "other" (Table 6.3). The sequences within those groups were directly compared to the genes that had also changed in the VU-1729 LDB1 knockout RNA-seq results (Chapter IV, Figure 4.7). Importantly, 31% of genes with promoter peak regions and 32% of intronic peak regions corresponded with actual changes in the gene expression based on the RNA-seq analysis. Within those groups of genes, there were even smaller percentages of peaks that contained either an E-box or a homeobox motif (summarized in Table 6.3). Although the majority of the gene peak regions that corresponding with gene expression changes were classified as 'other', the DNA sequences for all three validated ChIP-exo target genes contained either a homeobox or an E-box motif. Specifically the CDH11 promoter, CKAP2L promoter, and HMGA2 intronic, were found to contain a homeobox motif, a double homeobox motif, and an E-box motif, respectively. (Table 6.4).
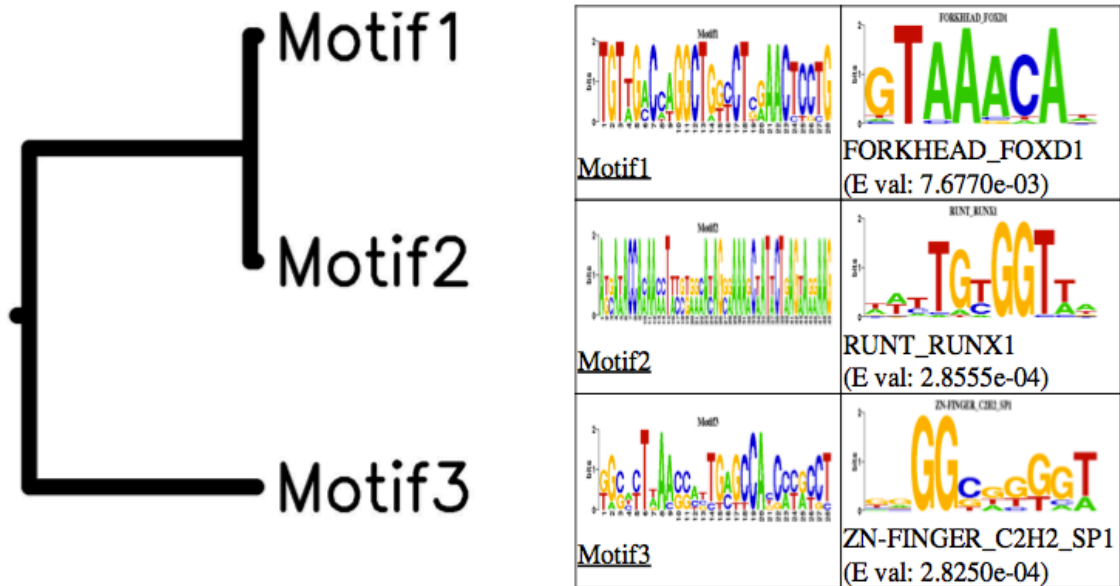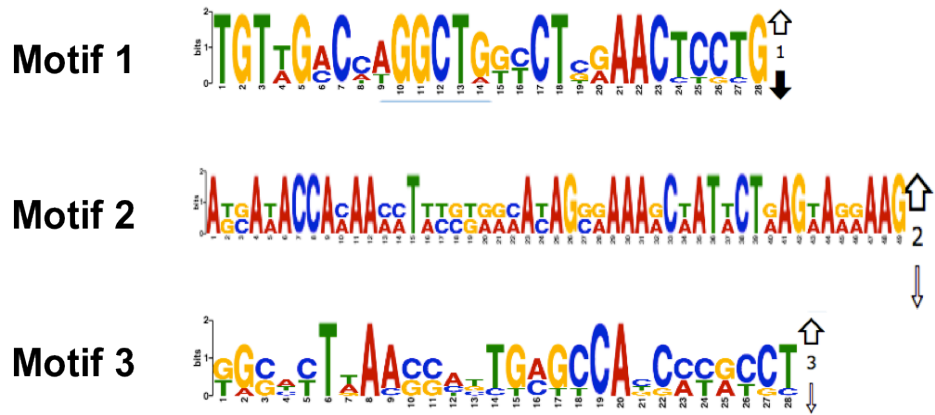
Figure 6.3 Promoter and intronic combined motifs. (Top) Top three discovered motifs in the combined list of promoter and intronic peaks. (Bottom) Motif Tree comparing known transcription factor motifs to discovered motifs. The known transcription factors associated with the discovered motifs were FOXD1, RUNX1, and C2H2.

Motif Type	ChIP-exo Peaks Overlapping with RNA expression	Promoter Peaks	%	Intronic Peaks	%
E-box	Total peaks with changed RNA expression in LDB1 KO cells	3 out of 11	27.27	3 out of 11	27.27
	Total peaks with *decreased* RNA expression in LDB1 KO cells	1 out of 3	33.33	2 out of 3	66.67
Homeobox	Total peaks with changed RNA expression in LDB1 KO cells	5 out of 11	45.45	2 out of 11	18.18
	Total peaks with *decreased* RNA expression in LDB1 KO cells	2 out of 5	40.00	1 out of 2	50.00
Other	Total peaks with changed RNA expression in LDB1 KO cells	7 out of 11	63.64	6 out of 11	54.55
	Total peaks with *decreased* RNA expression in LDB1 KO cells	3 out of 7	42.86	4 out of 6	66.67
All	Total peaks with changed RNA expression in LDB1 KO cells	11 out of 36	30.56	11 out of 34	32.35
	Total peaks with *decreased* RNA expression in LDB1 KO cells	5 out of 11	45.45	8 out of 11	72.73

Table 6.3 Top ChIP-exo peaks in promoter and intronic regions found by Rank Order. The top 36 promoter and 34 intronic peaks as found by Rank Order in ChIP-exo of VU-SCC-1729 cells were compared with changes in RNA expression as a result of LDB1 knockout in the same cells. Percentages of peak sequences containing an E-box or a homeobox that were discovered within promoter peaks and intronic peaks are shown comparatively for VU-SCC-1729 cells and VU-1729-2:7 LDB1 knockout cells. Percentages of peak sequences that did not contain an E-box or a homeobox motif that were discovered within promoter peaks and intronic peaks are also shown (other).

Genes found in ChIP-exo; validated in ChIP	Decrease LDB1 Occupancy in LDB1 KO (p-value)	Decrease SSBP2 Occupancy in LDB1 KO (p-value)	Possible Binding Motif	Sequence	log2FoldChange (DESeq2) in LDB1 KO	pValue(DESeq2)
CDH11 Promoter	< 0.0001	< 0.0001	Homeobox motif	ATTA or TAAT	-2.319702027	8.69E-68
CKAP2L Promoter	0.0071	0.0007	Double Homeobox motif	ATTA spacer(4) TAAT	N/A	N/A
HMGA2 2nd intron	< 0.0001	0.0047	Double e-box motif	CAGGTG spacer(9) CAGCTG	-2.491362405	2.03499E-75

Table 6.4 Known motifs in validated target genes. Three regions co-occupied by LDB1 and SSBP2 that were discovered in ChIP-exo analyses and validated using conventional ChIP analysis were evaluated for specific transcription factor binding motifs. *CDH11* promoter and *CKAP2L* promoter regions both contained homeobox motifs. *HMGA2* intronic region contained a double E-box motif.

Conclusions

The motif discovery and subsequent analyses were used to determine what DNA sequences an LDB1, SSBP2, and likely LMO4-containing complex might be binding to as transcriptional regulators. These studies on promoter and intronic peaks returned two different well-known motifs of interest, the E-box motif and the homeobox motif. While the combined analysis reduced the ability to find a statistically significant motif specific to promoter regulatory regions, the finding informs future analyses of this type that the regulatory regions should be separated during analysis in order to be able to maintain resolution of the potentially important targets.

The identities of the actual DNA-binding proteins that participate in the multiprotein complex containing LDB1, SSBP2, and likely LMO4, requires further investigation, however, the presence of the known motifs provide some clues. The homeobox motif and the E-box motif in the validated ChIP target genes points to the identity of the DNA-binding protein being either an LIM-Homeobox (LHX) or a basic helix-loop-helix (bHLH) transcription factor. Importantly, both LHX and bHLH proteins are known to bind to LIM-domain containing proteins and are reported with abilities to induce EMT (Natarjan *et al*., 2014, Yang *et al*., 2008).

Finally, the overlap found between ChIP-exo peak regions and changes in RNA expression after a loss of LDB1 may indicate direct targets occupied by an LDB1, SSBP2 containing complex. Further occupancy studies will be necessary to confirm direct targets containing these proteins of interest.

CHAPTER VII

DISCUSSION

LMO4, LDB1, and SSBPs promote invasion of HNSCC cells by EMT or CCI

There are many different regulatory mechanisms through which epithelial cellular identity and morphology can be reprogrammed, causing an otherwise healthy or normal cell to transform into a malignant one. Inappropriate alterations in gene expression networks allow for otherwise *in situ* epithelial cells to acquire motile and invasive abilities. Genes involved in maintaining cell polarization and cell-cell adhesion (E-CADHERIN) are downregulated or lost completely and genes that characterize an un-polarized mesenchymal phenotype (N-CADHERIN, VIMENTIN, MMP3, MMP9) are upregulated (Natarajan *et al*., 2014; Graves *et al*., 2014). Regulatory gene programs have been implicated in two processes thought to mediate tumor cell motility, invasiveness, and finally metastasis: 1) epithelial-mesenchymal transition (EMT) (Natarajan *et al*., 2014) and 2) collective cell migration (CCM) (Friedl *et al*., 2009) or collective cell invasion (CCI) (Etemad-Moghadam *et al*., 2015).

This body of work has shown that the abundance of LIM domain protein 4 (LMO4) and LIM domain-binding protein (LDB1) is increased in a majority of squamous cell carcinomas of the head and neck from multiple sites of origin and that their aberrant expression is caused, at least in part, by concomitant upregulation of two putative single-stranded DNA-binding proteins, SSBP2 and SSBP3. These proteins block LDB1 and LMO4 ubiquitylation and proteosomal destruction and thereby reduce their turnover.

Increased expression of LMO4 and LDB1 in the cell is correlated with changes in oral squamous cell morphology and function. This work reports the direct involvement of LMO4, LDB1, and SSBPs in the promotion of tumor cell invasiveness. In a recent study (Etemad-Moghadam *et al.*, 2015), an intermediate phenotype between EMT and CCI were reported for 92 oral squamous cell carcinomas (OSCCs). Our LDB1 knockout studies revealed a similar combinatory mode of invasion where both single invading cells and collective invading cell groups were found in the same oral squamous cell epithelial line. A significant reduction in invasive abilities as a result of Ldb1 loss in oral cavity carcinoma cells possibly implicates LMO4, LDB1, and SSBPs involvement in regulating a phenotypic EMT boundary in this tumor type.

A pro-oncogenic role of LMO4, LDB1, and SSBPs in HNSCC

The findings within this work add to a considerable literature showing that LMO proteins are pro-oncogenic, likely in association with LDB1, and establish in addition that these proteins control multiple aspects of tumor biology. The role of SSBPs in cancer, in contrast, has been less clear. While one study found a highly significant association between increased SSBP2 expression and poorer overall survival in glioblastoma (Xiao *et al.*, 2012), another showed inhibition of prostate cancer cell growth by SSBP2 with increased epigenetic silencing of its promoter with tumor progression (Liu *et al.*, 2008), more compatible with its function as a tumor suppressor. Similarly, there is a suggestion that SSBP2 acts as tumor suppressor in acute myeloid leukemia (Liang *et al.*, 2005), and it is possible that SSBPs can act as either oncoprotein or tumor suppressor, depending on the context.

Influence of LMO4, LDB1, and SSBPs on HNSCC is independent of HPV status

There is considerable heterogeneity in carcinomas of the head and neck, despite the fact that greater than 90% of them are squamous cell carcinomas. Importantly, carcinomas originating in the oropharynx are often positive for a high-risk type of human papillomavirus (HPV), whereas oral cavity carcinomas (and other less common subtypes like nasopharyngeal carcinomas) are generally HPV-negative (Leemans *et al*., 2011; Marur and Forastiere, 2016). Even though HPV-positive head and neck carcinomas show different genetic alterations, have different epidemiologic patterns, and demonstrate significantly different responses to chemotherapy and irradiation (Hobbs *et al*., 2006), the fact that no difference was detected in either the frequency of overexpression or intratumoral localization of LMO4, LDB1, and SSBP2 and SSBP3 between oral cavity and oropharyngeal carcinomas suggest the effects of these proteins on tumor cell biology is independent of virus. More specifically, LMO4, LDB1, and SSBPs may regulate some fundamental features of HNSCC tumor progression or invasion regardless of how they arise.

Potential role of LMO4, LDB1, SSBPs in head and neck cancer stem cells

Importantly, an LMO protein, specifically LMO2, LDB1, and SSBP2 have each been found to be required for the emergence or persistence of hematopoietic stem cells during development (Yamada *et al*., 1998) and Ldb1 has also been shown to be required for survival of stem cells in intestinal epithelium (Dey-Guha, 2009). This work presents a regulatory role of LMO4-, LDB1-, and SSBP-containing transcriptional complex(es) in metastatic potential and angiogenesis, both of which have been ascribed to cancer stem

95

cells (CSCs). A unifying hypothesis of these data, in consideration with the critical roles in developmental stem cells, is that LMO4-, LDB1-, and SSBP-containing transcriptional complex(es) affect stem cells, both malignant and normal. A preliminary observation supporting this notion was that approximately 80% of 20 head and neck CSC markers were co-occupied by LDB1 and SSBP2 in the genome-wide ChIP-exo analysis described in Chapter VI. Additionally, about 40% of those same CSC markers were changed in the RNA-seq analysis as a result LDB1 knockout *in vitro.* CSCs have been detected in a range of epithelial malignancies, including squamous cell carcinomas of the head and neck, and may be, in fact, regulated by the proposed complex.

LMO4, LDB1, SSBPs: promoters of angiogenesis in HNSCC

A role for LMO4 and/or LDB1 in invasiveness, but not angiogenesis, was predicted from immunohistochemical localization of these proteins in biopsies of human oral cavity carcinomas (Mizunuma *et al.,* 2003 and Kwong *et al.,* 2011). This work not only confirmed the previous findings and extended the analysis to confirm that SSBP2 and SSBP3 also had a role in HNSCC cell invasiveness (Chapter III), but also reported the novel finding that the role of LMO4, LDB1, and SSBPs also include the promotion of angiogenesis in this tumor type (Chapter IV). While LMO2 has been implicated in angiogenesis during normal development, these studies provide the first evidence of LMO4's involvement in tumor angiogenesis, which has implications for the other tumor types in which it is also overexpressed.

LMO4-, LDB1-, SSBP-containing complexes require further investigation

Although SSBP2, SSBP3, and, likely, SSBP4 can bind stretches of single-stranded DNA and were first discovered as a result of this property (Bayarsaihan *et al*., 1996), this single-stranded DNA-binding activity is entirely dispensable for its LDB1 and LMO protein stabilizing function. SSBP2 and SSBP3 have been shown to form multiprotein complexes with LDB1, LMO2, and heterodimers of basic helix-loop-helix transcription factors in erythroid cells and with LDB1 and LIM-homeodomain transcription factors in pituitary cells, respectively (Cai *et al*., 2008). Here within this body of work is evidence that in carcinomas of the head and neck, LMO4, is the critical member of an SSBP-containing multiprotein regulatory complex. An important question, however, that remains unanswered by this work is what are the identities of the sequence-specific transcription factors that recruit the complex to DNA and of the genes occupied and regulated by the complexes.

The genome-wide ChIP-exo studies (Chapter VI) began to answer this question by elucidating actual co-occupied gene targets of LDB1 and SSBP2, but further analysis including antibodies to LMO4 and to proposed DNA-binding transcription factors (likely bHLH or LIM-HD proteins) is required. Future studies should include ChIP-exo studies with antibodies of predicted and likely direct DNA binding proteins (E-box- and homeobox-binding transcription factors) to determine if such transcription factors co-occupy the same regions as LDB1 and SSBP2 across the genome.

Although bHLH and LIM-HD transcription factors have been well implicated in promoting EMT in epithelial cancers, there may be other types of transcription factors present in LMO4-, LDB1-, SSBP-containing multiprotein regulatory complexes that are

important for head and neck cancer. Interestingly, the majority of the analyzed ChIP-exo peaks were not E-box or homeobox motifs. Perhaps these 'other' motifs are, in fact, true targets of an LDB1, SSBP2-containing complex, however, they may not be directly involved in promoting the pro-oncogenic phenotype. Further analyses of the 'E-box, homeobox and other' types of motifs, including both occupancy studies and additional transcription factor ChIP-exo analysis will be crucial for further classifying these separate types of binding sites. Many of the unclassified 'other' motifs may, in fact, be novel findings, and upon further investigation could reveal important updates to the known transcription factor databases.

Conclusion

This body of work has addressed the hypothesis that LMO4, LDB1, and SSBPs participate in multiprotein transcriptional complexes that regulate head and neck cancer tumor cell biology. Expression data and loss of function studies have helped to highlight several biological effects of these proteins that are critical for tumor cell progression including increased invasion, proliferation, lymph node metastasis, and angiogenesis. Importantly, this work has also started the identification of genome-wide binding patterns of complexes containing LDB1 and SSBP2, although further investigations on the direct DNA-binding transcription factors are necessary. The relevance of LMO4, LDB1, and SSBPs in the maintenance of cancer stem cells, which may have wider implications outside of this tumor model alone, should also be further investigated. In sum, this body of work has contributed a better understanding of the mechanisms by which LMO4, LDB1, and SSBPs regulate tumor cell biology, but the important future studies suggested

here will be necessary to fully elucidate the functions of LMO-, LDB1-, SSBP-

containing complex(es). A more complete understanding of the genome-wide targets and

the regulatory pathways they affect may aid the development of novel therapies to treat

patients with invasive carcinomas of the head and neck.

REFERENCES

Agrawal N, F. M., Pickering CR, Bettegowda C, Chang K, Li RJ, *et al*. (2011). "Exome sequencing of head and neck squamous cell carcinoma reveals inactivating mutations in NOTCH1." Science **333**(6046): 1154-1157.

Aifantis, I., E. Raetz, *et al*. (2008). "Molecular pathogenesis of T-cell leukaemia and lymphoma." Nature Reviews Immunology **8**(5): 380-390.

Albini, A., Y. Iwamoto, *et al*. (1987). "A rapid in vitro assay for quantitating the invasive potential of tumor cells." Cancer research **47**(12): 3239-3245.

Andl, C. D., K. M. McCowan, *et al*. (2010). "Cathepsin B is the driving force of esophageal cell invasion in a fibroblast-dependent manner." Neoplasia **12**(6): 485-498.

Andl, C. D., T. Mizushima, *et al*. (2003). "Epidermal growth factor receptor mediates increased cell proliferation, migration, and aggregation in esophageal keratinocytes in vitro and in vivo." Journal of Biological Chemistry **278**(3): 1824-1830.

Armeanu-Ebinger (2011). " Differential expression of invasion promoting genes in childhood rhabdomyosarcoma." Int J Oncol. **38**(4): 1-8.

Baatenburg de Jong, R. J., J. Hermans, *et al*. (2001). "Prediction of survival in patients with head and neck cancer." Head & neck **23**(9): 718-724.

Bayarsaihan D, S. R., Lukens LN. (1998). "Cloning and characterization of a novel sequence-specific single-stranded-DNA-binding protein." Biochem J. **331**(Pt 2): 447-452.

Benson E, L. R., Eisele D, Fakhry C. (2014). "The clinical impact of HPV tumor status upon head and neck squamous cell carcinomas." Oral Oncol. **50**(6): 565-574.

Bergers G, B. R., McMahon G, Vu TH, Itoh T, Tamaki K, *et al*. (2000). "Matrix metalloproteinase-9 triggers the angiogenic switch during carcinogenesis." Nature Cell Biol. **2**: 737-744.

Bjerkvig R, J. M., Miletic Niclou SP. (2009). "Cancer stem cells and angiogenesis." Sem Cancer Biol. **19**(5): 279-284.

Boussadia, O. d., S. Kutsch, *et al*. (2002). "E-cadherin is a survival factor for the lactating mouse mammary gland." Mechanisms of development **115**(1): 53-62.

Brantley-Sieders, D. M., C. M. Dunaway, *et al*. (2011). "Angiocrine factors modulate tumor proliferation and motility through EphA2 repression of Slit2 tumor suppressor function in endothelium." Cancer research **71**(3): 976-987.

Cai Y, X. Z., Nagarajan L, Brandt SJ. (2008). "Single-stranded DNA-binding proteins regulate the abundance and function of the LIM-homeodomain transcription factor LHX2 in pituitary cells." Biochem Biophys Res Commun. **373**(2): 303-308.

Chai RC, L. D., Verma M, Punyadeera C. (2015). "Current trends in the etiology and diagnosis of HPV-related head and neck cancers." Cancer Med. **4**(4): 596-607.

Chytil, A., M. A. Magnuson, et al. (2002). "Conditional inactivation of the TGF,ÄêŒ≤ type II receptor using Cre: Lox." Genesis **32**(2): 73-75.

Dey-Guha I, M. M., Phillips M, Westphal . (2009). "Role of ldb1 in adult intestinal homeostasis." Int J Biol Sci. **5**(7): 686-694.

Du L, C. X., Cao Lu L, Zhang F, Bornstein S, et al. (2016). "Overexpression of PIK3CA in murine head and neck epithelium drives tumor invasion and metastasis through PDK1 and enhanced TGFβ signaling." Oncogene: doi: 10.1038.

Du L, S. J., Weems A, Lu SL. (2010). "Role of phosphatidylinositol-3-kinase pathway in head and neck squamous cell carcinoma." J Oncol. **2012:450179**(doi: 10.1155).

Duncan LD, . M., Carlson ER, Heidel RE, Kang E, Webb D. (2013). "p16 immunohistochemistry can be used to detect human papillomavirus in oral cavity squamous cell carcinoma." J Oral Maxillofac Surg. **71**(8): 1367-1375.

Elloul, S., M. Bukholt Elstrand, et al. (2005). "Snail, Slug, and Smad,Äêinteracting protein 1 as novel parameters of disease aggressiveness in metastatic ovarian and breast carcinoma." Cancer **103**(8): 1631-1643.

Etemad-Moghadam, S. and M. Alaeddini (2015). "Invasion Phenotypes of Oral Squamous Cell Carcinoma." Applied Immunohistochemistry & Molecular Morphology **23**(8): e12-e16.

Ferronha, T., M. A. Rabadon, et al. (2013). "LMO4 is an essential cofactor in the Snail2-mediated epithelial-to-mesenchymal transition of neuroblastoma and neural crest cells." The Journal of Neuroscience **33**(7): 2773-2783.

Floor S, v. S. ., Larsimont D, Dumont JE, Maenhaut C. (2011). "Cancer cells in epithelial-to-mesenchymal transition and tumor-propagating-cancer stem cells: distinct, overlapping or same populations." Oncogene: doi: 10.1038.

Grtz, G. G., K. Bucher, et al. (1998). "The oncogenic T cell LIM protein Lmo2 forms part of a DNA binding complex specifically in immature T cells." The EMBO Journal **17**(16): 4594-4605.

Güngör C, T.-I. N., Ma Drung A, Tursun B, Ostendorff HP, et al. (2007). "Proteasomal selection of multiprotein complexes recruited by LIM homeodomain transcription factors." Proc Natl Acad Sci U S A **104**(38): 15000-15005.

Hanahan, D. and R. A. Weinberg (2000). "The hallmarks of cancer." cell **100**(1): 57-70.

Ho YT, Y. J., Li TC, Lin JJ, Lin JG, Lai KC, *et al*. (2009). "Berberine suppresses in vitro migration and invasion of human SCC-4 tongue squamous cancer cells through the inhibitions of FAK, IKK, NF-κB, u-PA and MMP-2 and -9." Cancer Lett. **279**(2): 155-162.

Hosono, S., . Kajiyama, *et al*. (2007). "Expression of Twist increases the risk for recurrence and for poor survival in epithelial ovarian carcinoma patients." British Journal of Cancer **96**(2): 314-320.

Jemal, A., F. Bray, *et al*. (2011). "Global cancer statistics." CA: A Cancer Journal for Clinicians **61**(2): 69-90.

Kim S- K. E.-J., Hitomi M, Oh S- Jin X, Jeon -M, *et al*. (2015). "The LIM-only transcription factor LMO2 determines tumorigenic and angiogenic traits in glioma stem cells." Cell Death Diff. **22**(9): 1517-1525.

Kuo CL, L. K., Ma YS, Weng SW, Lin JP, Chung JG. Gallic acid inhibits migration and invasion of SCC-4 human oral cancer cells through actions of NF-κB, Ras and matrix metalloproteinase-2 and -9. Oncol Rep. 2014;32(1):355-61 (2014). "Gallic acid inhibits migration and invasion of SCC-4 human oral cancer cells through actions of NF-κB, Ras and matrix metalloproteinase-2 and -9." Oncol Rep. **32**(1): 355-361.

Kwong, R. A., C. J. Scarlett, *et al*. (2011). "LMO4 expression in squamous cell carcinoma of the anterior tongue." Histopathology **58**(3): 477-480.

Lechner, M., G. M. Frampton, *et al*. (2013). "Targeted next-generation sequencing of head and neck squamous cell carcinoma identifies novel genetic alterations in HPV+ and HPV tumors." Genome medicine **5**(5): 1.

Lee, S.-K., L. Jurata, *et al*. (2005). "The LIM domain-only protein LMO4 is required for neural tube closure." Molecular and Cellular Neuroscience **28**(2): 205-214.

Leemans, C. R., B. J. Braakhuis, *et al*. (2011). "The molecular biology of head and neck cancer." Nature Reviews Cancer **11**(1): 9-22.

Li J, K. Y., Wang, Clise-Dwyer K, Klumpp SA, Liang   *et al*. (2014). "Requirement for Ssbp2 in hematopoietic stem cell maintenance and stress response." J Immunol: 1950: doi: 10.4049.

Li L, J. R., Cui K, Lee JY, Cohen T, Gorivodsky M, *et al*. (2011). "Nuclear adaptor Ldb1 regulates a transcriptional program essential for the maintenance of hematopoietic stem cells." Nature Immunol. **2**(2): 129-136.

Li S, L. Q. (2014). "Cancer stem cells and tumor metastasis (Review)." Int J Oncol. **44**(6): 1806-1812.

Li S, L. Q. (2015). "Cancer stem cells, lymphangiogenesis, and lymphatic metastasis." <u>Cancer Lett.</u> **357**(2): 438-447.

Li Z. J., Hong S. (2014). "ANO1 as a marker of oral squamous cell carcinoma and silencing ANO1 suppresses migration of human SCC-25 cells." <u>Med Oral Patol Oral Cir Bucal.</u> **19**(4): 313-319.

Liang S. S., Nagarajan L. (2005). "SSBP2, a candidate tumor suppressor gene, induces growth arrest and differentiation of myeloid leukemia cells." <u>Oncogene</u> **24**(16): 2625-2634.

Liu J- N. J., Sun   Lee J, Kim MS, Ostrow KL, *et al*. (2008). "ssDNA-binding protein 2 is frequently hypermethylated and suppresses cell growth in human prostate cancer." <u>Clin Cancer Res.</u> **14**(12): 3754-3760.

Ma S, G. X., Beh PSL, Wong KY, Chan YP, Yuen HF, *et al*. (2006). "The significance of LMO2 expression in the progression of prostate cancer." <u>J Pathol.</u> **211**(3): 278-285.

Marur, S. and A. A. Forastiere (2016). <u>Head and neck squamous cell carcinoma: Update on epidemiology, diagnosis, and treatment</u>. Mayo Clinic Proceedings, Elsevier.

Marur S, B. B. (2014). "Oropharyngeal squamous cell carcinoma treatment: current standards and future directions." <u>Curr Opin Oncol.</u> **26**(3): 252-258.

Matthews, J. M. and J. E. Visvader (2003). "LIM‚Äêdomain‚Äêbinding protein 1: a multifunctional cofactor that interacts with diverse proteins." <u>EMBO reports</u> **4**(12): 1132-1137.

Mehrotra, R. and S. Yadav (2006). "Oral squamous cell carcinoma: etiology, pathogenesis and prognostic value of genomic alterations." <u>Indian journal of cancer</u> **43**(2): 60.

Mizunuma, H., J. Miyazawa, *et al*. (2003). "The LIM-only protein, LMO4, and the LIM domain-binding protein, LDB1, expression in squamous cell carcinomas of the oral cavity." <u>British Journal of Cancer</u> **88**(10): 1543-1548.

Montañez-Wiscovich M, S. D., Landis M, Visvader J, Andersen B, Keri R. (2009). "LMO4 is an essential mediator of ErbB2/HER2/Neu-induced breast cancer cell cycle progression." <u>Oncogene</u> **doi: 10.1038**.

Morcillo, P., C. Rosen, *et al*. (1997). "Chip, a widely expressed chromosomal protein required for segmentation and activity of a remote wing margin enhancer in Drosophila." <u>Genes & development</u> **11**(20): 2729-2740.

Mukhopadhyay, M., A. Teufel, *et al*. (2003). "Functional ablation of the mouse Ldb1 gene results in severe patterning defects during gastrulation." <u>Development</u> **130**(3): 495-505.

Nair, U., H. Bartsch, *et al*. (2004). "Alert for an epidemic of oral cancer due to use of the betel quid substitutes gutkha and pan masala: a review of agents and causative mechanisms." <u>Mutagenesis</u> **19**(4): 251-262.

Nakagawa, H., T. C. Wang, *et al*. (1997). "The targeting of the cyclin D1 oncogene by an Epstein-Barr virus promoter in transgenic mice causes dysplasia in the tongue, esophagus and forestomach." <u>Oncogene</u> **14**(10): 1185-1190.

Natarajan, J., C. Chandrashekar, *et al*. (2014). "Critical biomarkers of epithelial-mesenchymal transition in the head and neck cancers." <u>Journal of cancer research and therapeutics</u> **10**(3): 512.

National Cancer Institute (NCI). (2016). "Head and Neck Cancer." Retrieved from http://*et al*.cancer.gov/types/head-and-neck/

Nieto MA, C. A. (2012). "The epithelial-mesenchymal transition under control: global programs to regulate epithelial plasticity." <u>Sem Cancer Biol.</u> **22**(5-6): 361-368.

Ochoa, S. D., S. Salvador, *et al*. (2012). "The LIM adaptor protein LMO4 is an essential regulator of neural crest development." <u>Developmental biology</u> **361**(2): 313-325.

Ohira, M., S. Oba, *et al*. (2005). "A review of DNA microarray analysis of human neuroblastomas." <u>Cancer letters</u> **228**(1): 5-11.

P., R. (2009). "Collective cell migration." <u>Ann Rev Cell Develop Biol.</u> **25**: 407-429.

Panwar A, B. R., Lydiatt WM, Ganti AK. "Human papilloma virus positive oropharyngeal squamous cell carcinoma: a growing epidemic." <u>Cancer Treat Rev.</u> **40**(2): 215-219.

Parfenov M, P. C., Gehlenborg N, Freeman SS, Danilova L, Bristow CA, *et al*. (2014). "Characterization of HPV and host genome interactions in primary head and neck cancers." <u>Proc Natl Acad Sci U S A</u> **111**(43): 15544-15549.

Park H-S, H. M., Umehara T, Guo L-T, Osborne EM, Benner J, *et al*. (2011). "Expanding the genetic code of Escherichia coli with phosphoserine." <u>Science</u> **333**(6046): 1151-1154.

Peinado, H. c., D. Olmeda, *et al*. (2007). "Snail, Zeb and bHLH factors in tumour progression: an alliance against the epithelial phenotype?" <u>Nature Reviews Cancer</u> **7**(6): 415-428.

Prince MEP, A. L. (2008). "Cancer stem cells in head and neck squamous cell cancer." <u>J Clin Oncol.</u> **26**(17): 2871-2875.

Rabbitts, T. H. (1998). "LMO T-cell translocation oncogenes typify genes activated by chromosomal translocations that alter transcription and developmental processes." <u>Genes & development</u> **12**(17): 2651-2657.

Rhee, H. S. and B. F. Pugh (2012). "ChIP exo method for identifying genomic location of DNA binding proteins with near single nucleotide accuracy." Current Protocols in Molecular Biology: 21.24. 21-21.24. 14.

Rothenberg, S. M. and L. W. Ellisen (2012). "The molecular pathogenesis of head and neck squamous cell carcinoma." The Journal of clinical investigation **122**(122 (6)): 1951-1957.

Scully, C. "Oral health in America: a report of the Surgeon General." (2000): 1-308.

Shalem, O., N. E. Sanjana, *et al*. (2014). "Genome-scale CRISPR-Cas9 knockout screening in human cells." Science **343**(6166): 84-87.

Slaughter, D. P., . . Southwick, *et al*. (1953). "Field cancerization, in oral stratified squamous epithelium. Clinical implications of multicentric origin." Cancer **6**(5): 963-968.

Sum, E. ., B. Peng, *et al*. (2002). "The LIM domain protein LMO4 interacts with the cofactor CtIP and the tumor suppressor BRCA1 and inhibits BRCA1 activity." Journal of Biological Chemistry **277**(10): 7849-7856.

Sum, E. ., D. Segara, *et al*. (2005). "Overexpression of LMO4 induces mammary hyperplasia, promotes cell invasion, and is a predictor of poor outcome in breast cancer." Proceedings of the National Academy of Sciences of the United States of America **102**(21): 7659-7664.

Sun Z-J, C. ., Chen G, Wang R, Jia J, Chen X-M, *et al*. (2010). "LMO2 promotes angiogenesis probably by up-regulation of bFGF in endothelial cells: an implication of its pathophysiological role in infantile haemangioma." Histopathol. **57**(4): 622-632.

Tabor, M. P., R. . Brakenhoff, *et al*. (2001). "Persistence of Genetically Altered Fields in Head and Neck Cancer Patients Biological and Clinical Implications." Clinical Cancer Research **7**(6): 1523-1532.

Tian Y, W. N., Lu Z. (2010). "Repression of Lim only protein 4-activated transcription inhibits proliferation and induces apoptosis of normal mammary epithelial cells and breast cancer cells. Clin Exp Metastasis." Clin Exp Metastasis **doi: 10.1007**.

Tripic T, D. ., Cheng Zhang Vakoc CR, Gregory GD, *et al*. (2009). "SCL and associated proteins distinguish active from repressive GATA transcription factor complexes." Blood **113**(10): 2191-2201.

Vidal, L. and M. L. Gillison (2008). "Human papillomavirus in HNSCC: recognition of a distinct disease type." Hematology/oncology clinics of North America **22**(6): 1125-1142.

Visvader, J. E., D. Venter, *et al*. (2001). "The LIM domain gene LMO4 inhibits differentiation of mammary epithelial cells in vitro and is overexpressed in breast cancer." Proceedings of the National Academy of Sciences **98**(25): 14452-14457.

Visvader JE, M. X., Fujiwara Hahm K, Orkin SH. (1997). "The LIM-domain binding protein Ldb1 and its partner LMO2 act as negative regulators of erythroid differentiation." <u>Proc Natl Acad Sci U S A.</u> **94**(25): 13707-13712.

Wang, T., J. J. Wei, *et al*. (2014). "Genetic screens in human cells using the CRISPR-Cas9 system." <u>Science</u> **343**(6166): 80-84.

White RA, M. S., Wang XJ. (2010). "TGFβ signaling in head and neck squamous cell carcinoma." <u>Oncogene</u> **29**(40): 5437-5446.

Wu M. C., Keller CA, Mishra T, Pimkin M, Blobel GA, *et al*. (2014). "Dynamic shifts in occupancy by TAL1 are guided by GATA factors and drive large-scale reprogramming of gene expression during hematopoiesis." <u>Genome Res.</u> **24**(12): 1945-1962.

Xiao D. P., Rice T, McCoy LS, Smirnov I, Patoka JS, *et al*. (2012). "SSBP2 variants are associated with survival in glioblastoma patients." <u>Clinical Cancer Research</u> **18**(11): 3154-3162.

Xu, Z., X. Meng, *et al*. (2007). "Single-stranded DNA-binding proteins regulate the abundance of LIM domain and LIM domain-binding proteins." <u>Genes & development</u> **21**(8): 942-955.

Xu Z, H. S., Chang L-S, Agulnick AD, Brandt SJ. (2003). "Identification of a TAL1 target gene reveals a positive role for the LIM domain-binding protein Ldb1 in erythroid gene expression and differentiation." <u>Mol Cell Biol.</u> **23**(21): 7585-7599.

Yamada, Y, A. J. Warren, *et al*. (1998). "The T cell leukemia LIM protein Lmo2 is necessary for adult mouse hematopoiesis." <u>Proceedings of the National Academy of Sciences</u> **95**(7): 3890-3895.

Yamada Y, P. R., Forster A, Rabbitts TH. (2000). "The oncogenic LIM-only transcription factor Lmo2 regulates angiogenesis but not vasculogenesis in mice." <u>Proc Natl Acad Sci U S A.</u> **97**(1): 320-324.

Yamada Y, P. R., Forster A, Rabbitts TH. (2002). "The LIM-domain protein Lmo2 is a key regulator of tumour angiogenesis: a new anti-angiogenesis drug target." <u>Oncogene</u> **21**(9): 1309-1315.

Yoshida, S., N. Furukawa, *et al*. (2009). "Expression profiles of genes involved in poor prognosis of epithelial ovarian carcinoma: a review." <u>International Journal of Gynecological Cancer</u> **19**(6): 992-997.

Youngblood V, W. S., Song, Walter D, Hwang, Chen J, *et al*. (2015). "Elevated Slit2 activity impairs VEGF-induced angiogenesis and tumor neovascularization in EphA2-deficient endothelium." <u>Mol Cancer Res.</u> **13**(3): 524-537.

Yu M, R. L., Xie H, Schindler Y, Moran TB, Cheng Y, et al. (2009). "Insights into GATA-1-mediated gene activation versus repression via genome-wide chromatin occupancy analysis." Mol Cell. **36**(4): 682-695.

Yuen, Ä., C. Ä. Chua, *et al*. (2007). "Significance of TWIST and E-cadherin expression in the metastatic progression of prostatic cancer." Histopathology **50**(5): 648-658.

Zeisberg, M. and E. G. Neilson (2009). "Biomarkers for epithelial-mesenchymal transitions." The Journal of clinical investigation **119**(6): 1429.

Zhang, S., A. A. Postigo, *et al*. (1999). "Active transcriptional repression by the Rb,E2F complex mediates G1 arrest triggered by p16 INK4a, TGF-B, and contact inhibition." cell **97**(1): 53-61.