Molecular Analysis of Recurrent Translocations in Mucoepidermoid Carcinoma

by

Elizabeth Gensterblum-Miller

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Cellular and Molecular Biology)
in the University of Michigan
2023

Doctoral Committee:

Associate Professor J. Chad Brenner, Chair
Professor Celina Kleer
Associate Professor Yu Leo Lei
Associate Professor Ryan Mills
Professor Marina Pasca di Magliano

Elizabeth Gensterblum-Miller

gensterb@umich.edu

ORCID iD: 0000-0002-8714-1257

Dedication

This dissertation is dedicated to the cancer patients who volunteered to be included in these experiments. I hope that these results will eventually benefit people just like you, who have yet to be diagnosed.

Acknowledgements

Throughout my studies, I've been trained and helped by so many people to get where I am today. I'd like to thank all my mentors, lab mates, friends, and family, for all your support.

First, I would like to thank Chad. I'm glad that we got to work together for these past five years. Thank you for your insight and support through all our successes and setbacks, especially when we all became work-from-home bioinformaticians for part of 2020. Thank you for your mentorship throughout this whole process.

Next, I would like to thank my committee members, Drs. Celina Kleer, Yu Leo Lei, Ryan Mills, and Marina Pasca di Magliano. Dr. Lei, thank you for your help in really using my model systems to their full potential. Dr. Kleer and Dr. Pasca di Magliano, thank you for offering your expertise, especially by supporting me learning new techniques that really make this research better. Dr. Mills, thank you for critiques and suggestions in tackling data analysis strategies. I want to thank each of you for your support, questions, and critiques, which have really made this research better.

To the CMB graduate program, including CMB directors Drs. Manoj Puthenveedu, Ben Allen, Marina Pasca di Magliano, and Ariella Shikanov, as well as administrators Lauren Perl and Erin King, thank you for helping me as I navigated this degree.

Next, I would like to thank the current and former members of the Brenner lab. To former members Nicole, Jackie, Brittany, and Sue, and especially Megan and Aditi, thank you for showing me the ropes, and helping me navigate both grad school and getting started with this research. To current lab members Apurva, Chrissy, Collin, Jiayu, John Henry, Michella, and

Behirda, I think we've been a great team, especially through the ups and downs of the past few years. To the undergraduate students that I have mentored Jacob, Becca, Elissa, and Anisha, thank you for the great work you've done, and I'm looking forward to seeing what you do next. I would also like to thank Drs. Thomas Carey, Heather Walline, and Steven Chinn for your insights throughout this project.

I would also like to thank my former scientific mentors. Dr. Amr Sawalha, thank you for your mentorship when I was fresh out of my Bachelor's degree, and for introducing me to the power of high-throughput research. Drs. Patrick Coit and Eliza Tsou, I'm happy that I got to learn from and work with you guys during that time. Dr. Linda Mansfield, at Michigan State University, thank you for the opportunity to learn in your lab. I learned so much about research from you and Dr. Bell, but most importantly, how cool research can be.

A big thank you to my friends, throughout this whole thing. I'm glad that you've been there for me this entire time, even when it was only online for a year there.

Next, thank you to my family. Thank you for encouraging me to pursue my interests and passions whatever those were.

Finally, a big thank you to my partner, Max, for supporting me through this whole process, and Rosie, our cat, for making sure I take regularly scheduled breaks.

Table of Contents

List of Tables

List of Figures

Abstract

Head and neck cancers include a diverse group of malignancies, and pathogenesis is driven by different recurring somatic mutations. In head and neck squamous cell carcinoma (HNSCC), these mutations include single nucleotide variants of several different genes, as well as HPV viral integration. By contrast, many salivary gland tumors are characterized by genomic translocations, resulting in frequent gene fusions. For example, mucoepidermoid carcinomas (MEC) have prevalent *CRTC1-MAML2* fusions, while hyalinizing clear cell carcinomas (HCCC) have prevalent *EWSR1-ATF1* fusions. Exploring the molecular phenotypes caused by these driver mutations and others will better explain the mechanisms of tumorigenesis and growth in salivary gland tumors, and is therefore necessary to identify potential targets for future patient treatments. In my thesis, I investigate the hypothesis that driver mutations, such as *CRTC1-MAML2* and *EWSR1-ATF1*, alter transcription regulation in salivary gland tumors, varying based on tumor type, fusion status, and grade.

In this thesis, I begin by using molecular techniques to differentiate two salivary gland tumors, MEC and HCCC, which are difficult to differentiate by standard histopathology approaches. Using RNA sequencing (RNAseq), I identify a 354 gene signature that differentiates both malignancies. These genes are significantly enriched for an ATF1 binding motif, consistent with the *EWSR1-ATF1* fusion found in HCCC. These differentially expressed genes include *IGF1R*, *SGK1*, and *SGK3*, which are elevated in HCCC tumors. This, and other differentially expressed genes in this signature, describe examples of differing molecular pathology between MEC and HCCC.

I then seek to further understand the genetic underpinning of MEC. Within MEC tumors, the most common somatic translocation forms the CRTC1-MAML2 fusion. I map the CRTC1-MAML2 breakpoint in four MEC-derived cell lines, via long-read sequencing. I also identify a series of genomic translocations leading to this fusion and uncover a *TERT* promoter rearrangement in NCI-H292. Subsequent *TERT* break apart FISH reveals *TERT* copy number increase and translocation events in all four cell lines tested. These experiments reveal complex genomic rearrangement leading to *CRTC1-MAML2* formation and a novel *TERT* driver mutation. Thus, I discover and validate *TERT* as a novel MEC driver.

While the *CRTC1-MAML2* fusion is the most common MEC driver mutation, patients with *CRTC1-MAML2*, or less commonly *CRTC3-MAML2*, positive tumors have a better prognosis. Therefore, using RNAseq on 48 MEC tumors, I identify gene expression patterns associated with tumor *CRTC1/3-MAML2* fusion status and grade. Gene expression signatures associated with fusion status are enriched for gene sets involving cellular respiration, including oxidative phosphorylation and the electron transport chain. Moreover, changes to T and B cell infiltration are associated with *MAML2* fusion status and grade, respectively. Therefore, I perform spatial RNA sequencing to measure the effect of CRTC1-MAML2 activity throughout the MEC tumor microenvironment. I identify spatial overlap between CRTC1-MAML2 associated gene expression and many other transcripts, including *VEGFA* and *CTNNB1*. These data suggest that *CRTC1-MAML2*-associated gene expression affect a variety of biological processes throughout the tumor microenvironment.

Overall, these data describe a pattern of gene regulation dependent on tumor type, fusion status, and grade. These gene expression changes, coupled with novel driver mutations, such as *TERT* translocation, affect multiple cancer phenotypes throughout the tumor microenvironment.

These biological processes play a role in the molecular etiology of HCCC and MEC tumors, uncovering several pathways which are opportunities to advance targeted therapies, which may improve the survival of MEC and HCCC patients.

Chapter 1 Roles of Bioinformatics Advances in Head and Neck Cancer Genetics Research

1.1 Protecting Tumors by Preventing Human Papillomavirus Antigen presentation: Insights from Emerging Bioinformatics Algorithms[1]

1.1.1 Abstract

Recent developments in bioinformatics technologies have led to advances in our understanding of how oncogenic viruses such as the human papilloma virus drive cancer progression and evade the host immune system. Here, we focus our review on understanding how these emerging bioinformatics technologies influence our understanding of how human papilloma virus (HPV) drives immune escape in cancers of the head and neck, and how these new informatics approaches may be generally applicable to other virally driven cancers. Indeed, these tools enable researchers to put existing data from genome wide association studies, in which high risk alleles have been identified, in the context of our current understanding of cellular processes regulating neoantigen presentation. In the future, these new bioinformatics approaches are highly likely to influence precision medicine-based decision making for the use of immunotherapies in virally driven cancers.

1.1.2 Introduction

Human papilloma virus (HPV) has garnered public health attention, as HPV infection has been associated with 99.7% of all cases of cervical cancer (1). However, HPV is also highly

[1] This section was published in Cancers (Basel) in collaboration with J. Chad Brenner.

associated with head and neck squamous cell carcinoma (HNSCC). HNSCC affects epithelial cells in the region of the mouth and throat spanning the nasal cavity and oral cavity, through the larynx (2). The region of the head and neck that is most commonly infected by HPV is the oropharynx. In meta-analysis, approximately 20% of all HNSCC cases are HPV-positive, and the overall incidence of HNSCC in young patients continues to rise (3, 4). In the oropharynx, however, 40–60% of all cancers are associated with HPV infection (5) and the proportion of HPV-related HNSCC cases in the United States is increasing, which is thought to be due to changes in both sexual behaviors and smoking habits (6). Importantly, however, the molecular mechanisms by which HPV infected cells evade the host immune system have only recently begun to come into the limelight, which is largely due to the recent emergence of novel bioinformatics techniques.

Over 100 HPV strains have been found in humans; however, only a few strains have been associated with an increased cancer risk, which are termed high-risk strains (7). The most frequent of these types is HPV-16, which is associated with 80–90% of all HPV-positive HNSCC cases (5, 8). In the oral mucosa of cancer-free women, infections of low-risk HPV strains are quickly cleared, while infections of high-risk strains, such as HPV-16, continue to linger (9). Therefore, the factors utilized by high-risk HPV to circumvent efficient immune clearance are thought to contribute to the oncogenic activity of high-risk HPV types.

Several potential mechanisms may contribute to HPV-driven immune evasion. For example, while HPV-positive HNSCC is characterized by decreased presentation of the major histocompatibility complex, class I (MHC class I) on the cell surface, the molecular mechanisms leading to MHC class 1 suppression have only recently begun to emerge (10). The dysregulation of MHC class I presentation has a profound impact on the T cell-mediated anti-tumor immune response, as MHC class I expression is directly correlated with cytotoxic T cell and natural killer

cell infiltration in solid tumors (11, 12). HPV-positive tumors are associated with increased cytotoxic T lymphocyte infiltration (13). HPV-positive HNSCC tumors also have an increased cytotoxic T lymphocyte-mediated anti-tumor immune response compared to HPV-negative tumors; however, this immune response is negatively correlated with expression of the HPV E7 protein (14, 15). HPV-positive tumors are also associated with other immune changes that are not seen in HPV-negative tumors (16, 17). For example, B cells are less active in HPV-positive HNSCC tumors, and, interestingly, while B cell infiltration is increased in HPV-positive tumors, B cell-mediated antibody production is decreased (18). HPV positivity is associated with decreased activity in multiple immune cell types. Therefore, recovery of the anti-HPV immune response could improve anti-tumor MHC class I mediated immune response.

Recovery of MHC class I presentation may be poised as an ideal target for immunotherapy for HPV driven cancers, as this increased immune activity could improve tumor clearance. However, previous studies have been limited by the genomic complexity of the MHC class I region. As bioinformatics and sequencing techniques tailored to the MHC genetic locus have been recently developed and improved, and as publicly available MHC class I variation data has expanded over time, these limitations have decreased. Therefore, this review focuses on understanding how recently developed tools and resources impact our understanding of the mechanisms by which high risk HPV drives cancer development and impacts MHC class I presentation.

1.1.3 Normal MHC Class I Activity

The MHC class I, which in humans is occasionally also referred to as the human leukocyte antigen (HLA), presents antigens on the surface of normal nucleated cells (19). The closely related complex MHC class II is involved in antigen-presenting immune cells, and is beyond the scope of

this review. The MHC class I presents antigens generated from either endogenous or exogenous peptides on the surface of the cell. When exogenous antigens are presented, such as antigens generated from viral proteins, or neoantigens generated from mutated proteins, the immune system mounts a cytotoxic T lymphocyte and natural killer cell mediated immune response (19). This was first described in the context of viral infection, where it was demonstrated that antigens generated from viral proteins would trigger an immune response, but only when they are presented by the MHC class I (20). The MHC class I contains a heavy chain and a light chain; the light chain is transcribed by the *B2M* gene, while the heavy chain is transcribed from either the *HLAA*, *HLAB*, or *HLAC* genes (19).

Once expressed, MHC class I resides in the endoplasmic reticulum, where small antigens associate with MHC class I. The MHC-antigen complex is subsequently shuttled into the Golgi apparatus, and then continues to the cell surface (21). At the cell surface, the MHC class I interacts with either the T cell receptor presented by cytotoxic T lymphocytes, or the killer immunoglobulin-like receptor on natural killer cells. If the lymphocyte or natural killer cell recognizes the antigen as exogenous, and therefore potentially pathogenic, the cell will then mount an immune response targeting the exogenous antigen (14). In a normal cell, T lymphocytes will interact with viral antigens, killing the host cell. Therefore, downregulation of the MHC class I is one strategy by which virus-infected cells can avoid immune attack. In the case of oncogenic viruses, this could cause decreased efficacy of the anti-tumor immune response.

1.1.4 Neoantigen Identification Using Bioinformatics Methods

Recent advances in bioinformatics techniques have shed light on interactions between various HPV types and neoantigen presentation. The considerations for neoantigen prediction software have been previously reviewed (22-27). Many neoantigen prediction pipelines are available, using

a variety of prediction methods (**Table 1-1**). The most common method of neoantigen prediction, used in programs such as NetMHC, focuses on modeling the interaction between a predicted neoantigen and the MHC class I molecule (28). This method uses a neural network to predict the binding affinity of putative neoantigens, based on the HLA allelotype of the sample. Alternative methods include predicting location of peptide cleavage, and interaction with TAP proteins (29-31). Recently developed programs, including NeoPredPipe and NetTepi, use both MHC class I binding affinity and T cell receptor affinity, which leads to improved prediction accuracy (32, 33). In addition, multiple pipelines have been developed that integrate each step of neoantigen prediction into a single workflow, including TIminer, CloudNEO, and pVAC-seq (34-36). Some of these pipelines are specialized to identify neoantigens from specific types of mutations. For example, ScanNEO identifies neoantigens caused by insertion and deletion mutations, while INTEGRATE-Neo was developed to identify neoantigens caused by gene fusions (37, 38).

However, each prediction technique has limited efficacy, and predicted neoantigens are not reliably detected in vivo (39). This is largely because most programs perform neoantigen predictions based on limited aspects of the neoantigen presentation pathway. For example, the widely used prediction program NetMHC predicts neoantigens based on binding affinity to the MHC class I, but does not include neoantigen interaction with T cell receptors, RNA expression of putative neoantigens, or interactions with the neoantigen loading machinery. Programs that integrate multiple neoantigen prediction techniques, such as NeoPredPipe and NetTepi, may increase stringency of the prediction. However, additional pipelines are needed, which can improve prediction accuracy by incorporating additional neoantigen loading pathways into the neoantigen prediction algorithm. Therefore, in vivo validation of predicted neoantigens is necessary. In addition, it has been previously noted that because less binding data is available for

rarer HLA haplotypes, often neoantigen prediction for these haplotypes is less accurate (23). This problem is the motivation for programs such as NetMHCcons, which uses binding information from highly similar peptides to inform the binding patterns of less common haplotypes, which results in a modest increase in prediction accuracy (40). Recent advances have been made to increase the accuracy of neoantigen prediction tools, though additional methods are still needed.

1.1.5 Neoantigens in HNSCC

In addition to presenting viral antigens at the cell surface, the MHC class I proteins also present tumor neoantigens. These neoantigens are typically an eight to eleven amino acid peptide sequence that are generated from proteins containing either missense mutations or alternative splicing events (23). These then undergo antigen loading onto the MHC class I, via similar pathways used in viral antigen loading. In many cancers, identification of neoantigens has attractive therapeutic implications because the presentation of these neoantigens can elicit an anti-tumor immune response. In addition, vaccines that target these neoantigens are an attractive immunotherapy, and have been shown improve T cell mediated immune response in melanoma patients (41, 42). Clinical trials have been started for neoantigen vaccines in many additional cancer types, including glioma, bladder carcinoma, non-small cell lung cancer, pancreatic cancer, and triple negative breast cancer, (summarized in Chu *et al*. (43)).

Neoantigen identification in HNSCC is an attractive target because HNSCC tumors have a moderately high mutational burden in comparison with other cancers (44, 45), so a high number of putative neoantigens are possible (46). Indeed, HNSCC is found to have increased tumor mutational burden, as well as increased predicted neoantigen burden, when compared to other solid tumor types (47). This study predicts putative neoantigens across a large patient cohort using the prediction program NetMHCpan, identifying neoantigens that are predicted to bind to the MHC

class I with high affinity. However, studies such as this one are limited by the prediction methods currently available. Programs such as NetMHCpan use a neural network to predict antigen binding efficiency to the MHC class I, but do not predict other aspects of neoantigen processing, such as peptide cleavage, and neoantigen-T cell interaction. The specific putative neoantigens identified have not been shown to be biologically relevant, as it is not known if they will activate T cells, triggering a T cell-mediated anti-tumor immune response. Only a small fraction of these predicted neoantigens will actually play a role in immune regulation, so the actual biological role of specific neoantigens must be validated. One recent study identified neoantigens in ten HNSCC patients. Neoantigens were predicted using a similar pipeline as the previous study, but then each neoantigen was validated by testing each neoantigen's ability to activate patient CD8+ T cells (48). While 15 predicted neoantigens were identified, this study found that only two are able to trigger an immune response. Another study identified predicted neoantigens in HNSCC tumor samples, and treated patient T cells with these neoantigen peptides (49). They then measured T cell activation to determine the immunogenicity of each neoantigen. The authors described one specific neoantigen, the result of a gene fusion, that elicits an immune response, of all predicted neoantigens tested. These results suggest that gene fusion-derived neoantigens cause an increased immune response, compared to missense-derived neoantigens. Each of these studies successfully identify patient-specific neoantigens in HNSCC, which could be useful in future studies developing personalized immunotherapies, including neoantigen vaccines. However, they also illustrate current limitations in predicting the biological relevance of specific neoantigens, which is largely due to the complexity of both neoantigen presentation and neoantigen-mediated T cell activation.

1.1.6 MHC Class I Genotype May Affect HPV Antigen Presentation

As part of normal MHC function, HPV antigens are generated from translated peptides, then loaded onto the MHC class I antigen binding groove for surface presentation. Individual allelotypes have unique binding affinities to HPV antigens, and, therefore, HPV neoantigen load is also expected to vary between individual patients. In fact, recent studies have shown that the frequency and prognosis of HPV-positive cervical cancers are both associated with specific HLA haplotypes. For example, an association between HLA-B*07 and HPV-positive cervical cancer has been identified in multiple ethnic groups (50-52). This association has been identified in three independent populations, which each perform PCR sequence typing to identify HLA types present in each individual: 42 British cervical cancer patients and 946 healthy controls; 50 cervical cancer patients and 89 healthy controls from the Hubei province in China; and 141 high grade, 202 low grade, and 202 HPV-positive healthy patients in the United States. This led to the hypothesis that specific haplotypes may be more amenable to transition to HPV-driven cancer following HPV infection because certain haplotypes may not present as many HPV peptides on the cell surface.

Interestingly, to put the hypotheses generated by these papers in context using recently developed informatics tools, we used the NetMHC 4.0 to define the distribution of HPV peptide neoantigen presentation load across all known HLA haplotypes. Indeed, the predicted HPV16-based antigen load was highly dependent on the HLA allelotype of the infected individual (**Figure 1-1**). Surprisingly, however, the number of predicted neoantigens did not necessarily correlate the reported increased risk of HPV-positive cervical cancer (**Figure 1-1**) as different high-risk haplotypes had a widely varying predicted neoantigen load. This lack of a correlation could be due to many factors. For example, intergenic variants or pathway mutations may play a role in neoantigen presentation that is not described by the HLA haplotype alone. Also, recent research has also shown that HPV oncoproteins also directly regulate MHC expression as described below,

as such it is possible that certain HPV haplotypes are also more stringently regulated by HPV oncoproteins. Therefore, additional research is necessary to determine the role of individual mutations, and their effects on HLA regulation.

1.1.7 The Role of MHC Class I Changes in HNSCC

Early studies have documented that the MHC class I is lost in approximately 50% of HNSCC tumors (53). Though these studies do not classify HPV status, more recent studies have identified MHC class I loss in both HPV-positive and HPV-negative HNSCC, as HPV-positive tumors had a lower surface level of MHC class I than HPV-negative tumors (10). This study measured MHC class I surface expression by immunohistochemistry of paraffin-embedded biopsies, using 27 HPV-positive and 68 HPV-negative tumors. There is some debate, however, about the prognostic effects of MHC class I loss (10, 54). Despite these initial data, large-scale studies, as well as meta-analyses, are still necessary to determine the association of MHC class I expression and prognosis. Although the studies completed to date demonstrate that MHC class I loss is detected in HNSCC, regardless of HPV status, the specific mechanism of MHC class I loss in both tumor subtypes remains incompletely understood. Thus, further work is also necessary to determine the processes involved in MHC loss, and the differences between HPV-positive and negative MHC loss.

Recent sequencing analysis of HNSCC tumors has identified one mechanism of MHC inactivation. Somatic mutations in both HLA proteins and related pathway proteins have been documented in HNSCC. In a large HNSCC patient cohort published by the TCGA cancer atlas (*n* = 515), 9% (45/515 of all HNSCC patients have somatic mutations in the MHC class I heavy chain genes (*HLA-A, HLA-B,* and *HLA-C*), or the MHC class I light chain gene (β-2 microglobulin, *B2M*) (**Figure 1-2**). These mutations include nonsense mutations, which lead to loss of the MHC class I, as well as several missense mutations. Of particular note, 61% (11/18) from this cohort have

missense mutations found in the antigen binding groove of the MHC class I protein, which are predicted to lead to altered neoantigen binding specificity within the tumor (**Figure 1-2**).

In multiple types of cancer, including HNSCC, enrichment of loss-of-function mutations, including missense mutations in the antigen binding groove and nonsense mutations, has been documented (47, 55). Moreover, in multiple cancer types, MHC class I loss-of-function mutations, including mutations in the HLA genes and β-2 micro-globulin, are associated with decreases in markers of CD8+ T cell activity (47), suggesting that the mutations identified in the HNSCC TCGA project as well as in HNSCC cell line panels (56, 57) also lead to decreased CD8+ T-cell activity in this tumor type.

In addition to mutations in the MHC class I protein, mutations in the pathways that control transcriptional regulation, neoantigen loading, or protein trafficking of the MHC class I have been found in 21% of HNSCC patients in the TCGA cancer atlas (106/515) (**Figure 1-2**). These mutations could putatively affect MHC class I expression or antigen binding, but many have not yet been individually validated. As such, additional studies are needed to generate a comprehensive list of all proteins involved with MHC class I gene expression, especially in squamous tissues, as it has been shown that a wide variety of transcription factors are involved in HLA regulation and that the role of these transcription factors is highly dependent on cell type (58). Consequently, the role of these neoantigen pathway mutations in MHC class I-mediated processes, such as neoantigen presentation, has not been fully described. Moreover, bioinformatics tools predicting the effect of these pathway mutations on antigen presentation are not currently available. Therefore, additional studies are needed to determine the specific transcription factors that are necessary for HLA transcriptional regulation in HNSCC and to determine the functional role of antigen presentation pathway mutations.

1.1.8 Considerations for MHC Class I Mutations

Limitations in describing the genetic regulation of the MHC class I are often due to the high density of variants in the MHC genetic region. Additional whole-genome sequencing studies are needed to identify somatic mutations that affect the activity of MHC class I in HNSCC patients. While exonic mutations certainly play a role in the activity of the MHC class I peptide, it has been shown that intergenic mutations and variation play a significant and complex role in the transcriptional regulation of the MHC class I (58). The role of variation within the MHC genomic region is incompletely understood, which is partially due to unique technical considerations when studying this genomic region.

The MHC genomic region refers to a four megabase portion of chromosome 6 that contains all MHC class I genes, as well as the MHC class II, which is involved in a separate pathway of antigen presentation by immune cells. This region is one of the most genetically complicated regions of the human genome, with a high density of variation between individuals. MHC genetic variation is generalized into named haplotypes for each MHC gene; currently, 13,680 validated MHC class I gene haplotypes have been described worldwide (59). Moreover, the MHC genomic region is characterized by a complex network of gene regulation, which is highly dependent on MHC genetic variation (58, 60)

In HNSCC, the effect of MHC class I genetic variation on anti-cancer immune activity and HPV-MHC interactions has not been described. In addition, transcriptional regulation of MHC class I is highly cell type-specific, so many known mechanisms of MHC class I transcriptional regulation have not been validated in the context of HNSCC (58). Interesting future directions of study to characterize this regulation include: functionally relevant intergenic somatic mutations, the effect of HLA haplotypes in HNSCC neoantigen presentation, and the transcription factors

involved in MHC class I regulation in HNSCC. Before the role of HPV proteins can be fully described in individual patients, the role of intrinsic variation between individuals will have to be considered.

1.1.9 HPV E7 Inhibits MHC Class I Gene Expression

MHC class I loss in HPV-positive cancers could be directly caused by specific HPV proteins with the MHC genetic locus. Two HPV proteins, E6 and E7, are important oncogenes, and play a role in the development of HNSCC. HPV E6 and E7 are constitutively expressed throughout viral infection (61). E6 degrades p53, an important tumor suppressor gene critical for cell cycle arrest (62). Interestingly, in HPV-negative HNSCC tumors, which lack E6 activity, p53 loss of function mutations occur at a much higher frequency than in HPV-positive cases (63). In addition, E7 interacts with retinoblastoma protein (RB), another tumor suppressor (64). However, E7 protein also has a significant role in aberrant gene regulation of infected cells, and E7 is directly linked to dysregulation of MHC class I transcription.

In HEK-293 cell lines infected with high-risk HPV strains, MHC mRNA levels are much lower than in samples infected with low-risk HPV strains (65). Moreover, E7 proteins from high-risk HPV strains are directly associated with MHC class I expression (12). HPV E7 is reported to bind to, and repress, HLA promoters (66, 67). However, these reports were published before high-throughput techniques revealed a complex network of trans-regulation in the MHC class I genetic region (58). Therefore, the implications of these promoter binding patterns must be reconsidered in the greater context of complicated regulation that is the hallmark of the MHC region. For example, E7 may interact with distal enhancers, as well as the HLA promoters. Alternatively, proteins or mutations at these distal enhancers may play a role in regulating E7 inhibition of HLA gene expression.

E7 has been shown to interact with the *HLA-A* promoter, and could interact with other promoters and enhancers in the HLA genetic region. While directly interacting with the promoters of HLA genes, E7 could recruit specific epigenetic editor proteins, such as histone deacetylases or DNA methyltransferases, which directly alter the epigenetic patterns at each HLA promoter, which would result in repressed transcription of each HLA gene. E7 has been shown to recruit histone deacetylases, which remove open chromatin marks, reducing the region's openness to transcription (65). In addition, transcriptional repression of MHC class I genes can be reversed by treatment with a DNMT1 inhibitor (68). DNMT1, or DNA methyltransferase 1, maintains DNA methylation patterns, and also produces de novo methylation events. Because HLA transcription increases after DNMT1 inhibition, HLA transcriptional repression could be mediated, at least partially, by DNA methylation. Moreover, this same study describes that after recovery of MHC class I expression by DNMT inhibitor treatment, immunotherapy efficacy improves in a mouse HPV-16-positive tumor model (69). These promising results demonstrate that recovery of MHC class I presentation is an attractive target for improving immunotherapy treatments.

1.1.10 The Next Steps of HPV-Targeting Therapies

High-risk strains of HPV have been shown to inhibit MHC class I surface presentation in cervical cancer. Constitutively expressed viral proteins E6 and E7 inhibit expression of MHC class I proteins, while E5 prevents trafficking of the MHC class I to the cell membrane. Decreased MHC class I presentation on the cell surface then interferes with immune attack of the infected cell. However, it remains unclear how this interference with immune response can be rescued in the context of HNSCC. These studies have the potential to improve treatment of HPV, through the identification and inhibition of HPV-specific pathways. For example, increasing expression of the MHC class I by inhibition of HPV proteins E5 or E7 could improve patient outcomes.

Previous studies describing the role of HPV in MHC class I regulation have been successful in proposing specific mechanisms, but they have been limited by the variability of both the MHC protein and the MHC genomic region. These issues have been addressed by recent bioinformatics techniques, including sequence alignment workflows and antigen binding affinity predictions. By implementing these methods, future studies will be able to characterize the effect of HPV within the complex MHC class I regulatory network that has been described. However, new tools are also needed to improve antigen prediction accuracy and to further address limitations caused by MHC genetic variability. Therefore, the use of bioinformatics and sequencing methods developed for working with the MHC may better inform treatment for HPV-positive cancers.

1.2 Diverse Bioinformatics Strategies in the Study of HPV-Negative Tumors

In the preceding section, I summarize currently available bioinformatics methods useful to determine how HPV affects HNSCC tumorigenesis. These methods include several antigen and neoantigen prediction pipelines and sequencing challenges important when studying the MHC genomic locus and MHC class I antigen presentation activity. Utilizing a diversity of sequencing methods is therefore critical to detecting HPV-mediated development in an HPV-positive tumor.

As in HPV-positive HNSCC, many cancers include complex gene mutation patterns, which can affect gene regulation in dramatic ways. For example, mucoepidermoid carcinoma (MEC), a salivary gland malignancy, has been found to harbor genome rearrangement driver mutations, including a translocation leading to the formation of CRTC1-MAML2 fusion(70). Wild type CRTC1 and MAML2 are each transcription factors individually; the fusion typically contains the CRTC1 CREB binding domain and the MAML2 transcriptional activation domain (70, 71). The CRTC1-MAML2 fusion has been shown to affect gene regulation, upstream of the transcription factors PPARγ, CREB, and MYC (72-74). However, the extent of genome rearrangements in

MEC tumors is not known, in part due to limitations in commonly used whole exome sequencing methods. The mechanism by which the CRTC1-MAML2 fusion is reproducibly generated, the extent of additional translocation events, and the effects these mutations have on gene regulation throughout the tumor are open questions within the study of MEC tumorigenesis and growth. Each of these questions will require the use of emerging bioinformatics and sequencing strategies to determine the biological processes involved.

1.3 Figures

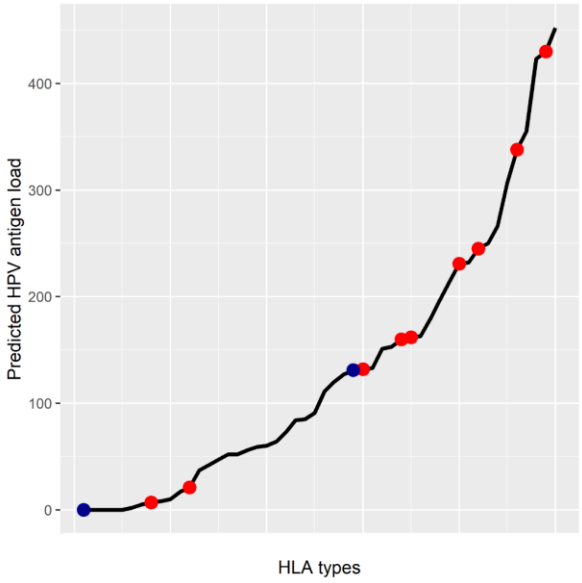


Figure 1-1. The predicted HPV antigen load of all common MHC class I haplotypes.

All possible antigens were generated from the canonical peptide sequences encoded from HPV16, and each antigen's binding affinity to each HLA haplotype was predicted using NetMHC 4.0 (28). Predicted HPV antigen load is the sum of all HPV antigens predicted to strongly bind to each MHC class I protein (binding affinity < 500 nM). All MHC haplotype available from the immune epitope database are included in this analysis (68). Specific HLA alleles are highlighted. Red points correspond to HLA alleles associated with increased risk of cervical cancer, and blue points are protective alleles for cervical cancer (75).

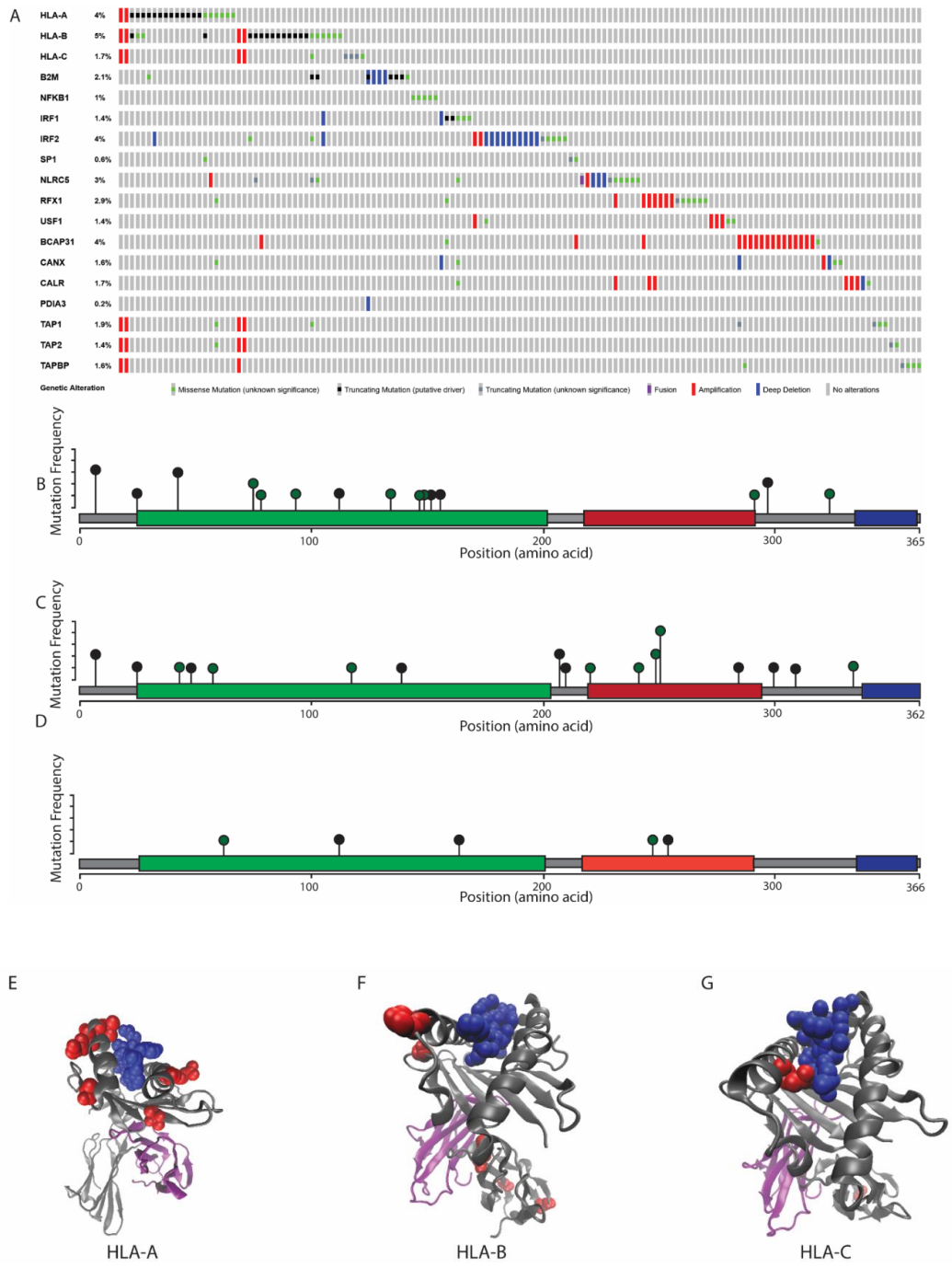


Figure 1-2. Mutation frequency and localization of MHC class I complex and related proteins.

(**A**) mutation frequency of members of the MHC. HLA-A, B, and C, are paralogs of the MHC heavy chain, and β-2 microglobulin (*B2M*) forms the MHC light chain. 8% of HNSCC tumors have a somatic mutation in one or more HLA gene, and 2% have a mutation in *B2M*. 21% of all tumors have a somatic mutation in one or more genes associated with either MHC class I transcriptional regulation or the antigen loading

pathway. *NFKB*, *IRF1*/*2*, *SP1*, *NLRC5*, *RFX1*, and *USF1* are each transcription factors that have been shown to interact with the HLA promoter region in various cell types (76-78). The transcription factors that regulate MHC class I have not been described in HNSCC. *BCAP31* is necessary for shuttling the MHC class I from the ER to the GA, and each of the remaining proteins has a role in MHC class I folding and neoantigen loading within the ER (79-81). Transcriptional regulation of the MHC class I has been shown to depend on a complex network of cis and trans-acting regulation, so a comprehensive list of transcription factors with a defined role in MHC class I regulation has not been fully determined (58). (**B**–**D**) Localization of mutations detected in HNSCC patients in HLA-A (**B**), HLA-B (**C**), and HLA-C (**D**). Black dots represent a nonsense mutation; green dots represent a missense mutation. Peptide domains are marked on each MHC class I peptide. Green domain: MHC class I alpha 1 and 2 domains. Red: C1-set domain. Blue: C terminal domain. (**E**–**G**) Location of somatic missense mutations within the HLA protein detected in HNSCC tumors. The HLA protein consists of three α domains: α-1 and -2 contain the antigen binding groove, while the α-3 domain interacts with β-2 microglobulin. Purple: β-2 microglobulin. Grey: HLA heavy chain, including the α-1–3 domains. Blue: neoantigen peptide, located in the antigen binding groove. Red: amino acids that are subject to somatic missense mutation in one or more HNSCC patient. Mutations in HLA-A (**E**) are localized to the antigen binding groove, eight total missense mutations. Mutations in HLA-B (**F**) are present in each α domain, seven total missense mutations. Mutations in HLA-C (**G**) are present in the α-3 domain as well as the antigen binding groove, two total missense mutations. Mutation data retrieved from the TCGA PanCancer Atlas.

1.4 Tables

Program Name	Input Data Type	Summary	Website
CloudNeo	WES or WGS	Integrates neoantigen peptide sequence calling, HLA typing, and peptide-MHC binding affinity predictions	https://github.com/TheJacksonLaboratory/CloudNeo
INTEGRATE-neo	RNA-seq	Integrates gene fusion identification	https://github.com/ChrisMaherLab/INTEGRATE-Neo
NeoPredPipe	variant call set	Integrates putative neoantigen peptide sequence identification and MHC binding affinity prediction	https://github.com/MathOnco/NeoPredPipe
NetChop	Peptide sequences	Predicts peptide cleavage sites	http://www.cbs.dtu.dk/services/NetChop
NetMHC	Peptide sequences, MHC haplotype	Predicts neoantigen binding affinity in an MHC type-dependent manner	http://www.cbs.dtu.dk/services/NetMHC
NetMHCcons	Peptide sequences, MHC DNA sequence	Predicts antigen binding affinities of rare MHC haplotypes	http://www.cbs.dtu.dk/services/NetMHCcons
NetMHCpan	Peptide sequences, MHC haplotype	Similar to NetMHC, with more MHC types included in training data	http://www.cbs.dtu.dk/services/NetMHCpan
NetTepi	Peptide sequences, MHC haplotype	Predicts neoantigen activity by combining peptide-MHC binding affinity and stability, and T cell propensity	http://www.cbs.dtu.dk/services/NetTepi
Pcleavage	Protein sequences	Predicts peptide cleavage sites	http://crdd.osdd.net/raghava/pcleavage
PRED(TAP)	Peptide sequences	Predicts peptide-TAP binding patterns	antigen.i2r.a-star.edu.sg/predTAP (currently unavailable)
pVAC-Seq	WES or WGS and RNA-seq	Combines variant calling and RNA-seq to identify transcribed putative antigens	https://github.com/griffithlab/pVAC-Seq
ScanNeo	RNA-seq	Neoantigen sequence prediction, optimized for indel mutations	https://github.com/ylab-hi/ScanNeo
TIminer	RNA-seq, somatic mutation calling	Integrates RNA-seq and somatic mutations to predict expressed neoantigens	https://icbi.i-med.ac.at/software/timiner/timiner.shtml

Table 1-1. Bioinformatics methods available for neoantigen prediction.

Each algorithm described in this review is included in the table. Peptide sequences are typically generated from nonsynonymous coding mutations, identified from variant call sets. WES: whole exome sequencing. WGS: whole genome sequencing. TAP Transporter associated with antigen processing. Indel: Insertion-deletion mutation.

Chapter 2 Integrative Sequencing Discovers an ATF1-Motif Enriched Molecular Signature that Differentiates Hyalinizing Clear Cell Carcinoma from Mucoepidermoid Carcinoma[2]

2.1 Abstract

Objectives: Salivary gland tumors are comprised of a diverse group of malignancies with widely varying prognoses. These cancers can be difficult to differentiate, especially in cases with limited potential for immunohistochemistry (IHC)-based characterization. Here, we sought to define the molecular profile of a rare salivary gland cancer called hyalinizing clear cell carcinoma (HCCC), and identify a molecular gene signature capable of distinguishing between HCCC and the histopathologically similar disease, mucoepidermoid carcinoma (MEC).

Materials and Methods: We performed the first integrated full characterization of five independent HCCC cases.

Results: We discovered insulin-like growth factor alterations and aberrant IGF2 and/or IGF1R expression in HCCC tumors, suggesting a potential dependence on this pathway. Further, we identified a 354 gene signature that differentiated HCCC from MEC, and was significantly enriched for genes with an ATF1 binding motif in their promoters, supporting a transcriptional pathogenic mechanism of the characteristic *EWSR1-ATF1* fusion found in these tumors. Of the differentially expressed genes, *IGF1R*, *SGK1* and *SGK3* were found to be elevated in the HCCCs relative to MECs. Finally, analysis of immune checkpoints and subsequent IHC demonstrated that CXCR4 protein was elevated in several of the HCCC cases.

[2] This chapter was originally published in *Oral Oncology* in collaboration with the following authors: Molly E. Heft Neal, Apurva D. Bhangale, Aditi Kulkarni, Jingyi Zhai, Joshua Smith, Collin Brummel, Sue K. Foltin, Dafydd Thomas, Hui Jiang, Jonathan B. Mchugh, and J. Chad Brenner.

Conclusion: Collectively, our data identify an ATF1-motif enriched gene signature that may have clinical utility for molecular differentiation of HCCCs from other salivary gland tumors and discover potential actionable alterations that may benefit the clinical care of recurrent HCCC patients.

2.2 **Introduction**

Hyalinizing clear cell carcinoma (HCCC) is a rare malignancy that primarily arises in the intraoral minor salivary glands and less commonly in the base of tongue, tonsil, nasopharynx, and lung (82-89). Despite often multimodal therapy, this disease has a relatively high rate of recurrence (90). HCCC is morphologically characterized by infiltrating groups of clear cells embedded in hyalinizing stroma. Although pathologically the disease is occasionally confused with variants of related neoplasms such as mucoepidermoid carcinoma and clear cell squamous cell carcinoma (91-94), the presence of one of two different CREB family gene fusions, *EWSR1-ATF1* or *EWSR1-CREM*, provides clear molecular distinction for the majority of HCCCs (95, 96). Unfortunately, beyond the occurrence of these gene fusions, little else is known about how these fusions regulate the molecular composition of the tumor or the other disruptive molecular events that co-occur to differentiate the molecular profile of HCCC from mucoepidermoid carcinoma.

The EWSR1-ATF1 gene fusion was originally discovered in HCCC in 2011 using fluorescence in situ hybridization (FISH) to show break apart of EWSR1 probes and subsequent 3' RACE to identify ATF1 as the fusion partner in 13/14 (93%) of cases studied. Since that time, the role of this gene fusion in HCCC molecular oncogenesis has remained uncharacterized. (95) Given this gap in knowledge, we sought to further characterize the molecular composition of HCCC. We performed integrative exome and transcriptome sequencing analysis on two and five independent *EWSR1-ATF1* fusion positive HCCC tumors respectively. Because of the rarity of the disease, we postulated that this information could help define novel drivers of HCCC and, when compared to molecular information from a cohort of

mucoepidermoid carcinoma, could begin to identify a gene signature characteristic of *EWSR1-ATF1* fusion positive HCCC that may have future utility to help distinguish the two tumor types as well as understand the mechanistic role of *EWSR1-ATF1* fusions in disease pathogenesis.

2.3 **Materials and Methods**

2.3.1 ***Clinical specimens and clinical data***

Patients with HCCC were identified from the University of Michigan pathology archive using an IRB-approved protocol for next generation sequencing of DNA and RNA (HUM00080561). Clinical, histologic, and outcome data was collected from medical records and the Social Security Death Index as previously described(97). Following hematoxylin and eosin staining of sections from each block, our head and neck pathologist (J.B.M.) identified blocks with >60% tumor content for coring. DNA and RNA were then simultaneously isolated using the Qiagen AllPrep kit as described (98) and advanced for NGS if it met our previously defined quality standards defined by Qubit and Bioanalyzer analysis (4, 99).

2.3.2 ***Exome sequencing***

Total genomic DNA from each tumor (n=2; HCCC1, HCCC2) and adjacent normal specimen was submitted to the University's DNA sequencing core for library preparation and exome sequencing using the DNA TruSeq Exome Library Preparation kit (Illumina, Catalogue number: FC-150-100x). Libraries were then pooled post-capture and sequenced on a single lane of an Illumina HiSEQ4000 using 150 nucleotide paired end protocol, yielding an average depth of greater than 100x per sample (**Table 2-1)**.

2.3.3 ***Exome variant calling***

We assessed the quality of the sequencing reads using FastQC v.0.11.5 as previously described (100). Trim Galore v0.4.4 was used to remove adapters and trim reads. Reads were aligned to the hg19 reference genome using BWA v0.7.1. Mapping was followed by marking duplicates using Picard v1.79. Next, base quality score recalibration was completed using GATK v3.6. SAMtools v1.2 was then used to create the pileup files for each tumor-normal pair in the set. We then used Varscan v2.4.1 to call variants from the mpileup files using the somatic mode of the variant caller. Golden Helix Varseq v1.4.6 was used to annotate variants. All variants in the introns and intergenic regions were filtered out. Variants with more than 30 reads were considered high confidence calls and variants with a minimum of 5 reads supporting the alternate allele in the tumor samples were annotated as low confidence and considered as potential positives for validation by Sanger sequencing.

2.3.4 ***Copy number analysis***

We used the aberration Detection in Tumour Exome (ADTEx) v.2.0 to make copy number estimation calls. A copy number state from 0 to 4 was assigned by the software based on its estimated copy number at each position. State 0 corresponded to a homozygous deletion, 1 corresponds to a heterozygous deletion, normal copy number is denoted by state 2, while states 3 and 4 represent a gain and amplification, respectively. The software was also used to generate representative Manhattan plots for each chromosome of each tumor/normal pair. Genes were annotated using R script (v3.4.0) to list the gene associated with each change.

2.3.5 ***Microsatellite instability (MSI) detection***

We used the MANTIS algorithm to detect somatic MSI from the tumor-normal sample pairs as described (101). This software package assigns an MSI status to each sample pair based on an instability score calculated for each sample pair.

2.3.6 ***Transcriptome sequencing***

We submitted up to 500ng of RNA from each tumor (n=5; HCCC1, HCCC2, HCCC4, HCCC5, HCCC7) to the University of Michigan Advanced Genomics Core for library preparation and subsequent sequencing. Library preparations were completed with the Illumina TruSeq Stranded Total RNA library prep kit (Cat#: RS-122-2201/2) following the manufacturer's protocol, with only one modification to increase the number of PCR cycles to 14 in order to increase amplification of the library prior to final purification. The samples were then pooled and loaded onto a single lane of an Illumina HiSEQ4000 for paired end sequencing to 75nt. RNA sequencing yielded an average of 62M reads with >88% uniquely mapped for each tumor. A summary of sequencing quality statistics including total unique mapped reads for each sample is provided in **Table 2-2**.

2.3.7 ***Transcript quantification and clustering analysis***

The quality of RNA sequencing reads was assessed using FastQC v0.11.5. We did not identify any quality issues from either library. We used STAR v2.5.2a to generate genome index files based on hg19 version and then generated the transcript annotation file. We used the genome index files to map the sequencing reads against the human genome. Using Picard v1.79 and SAMtools v1.2 we then extracted only uniquely mapped reads from the BAM files for quantification. Fragments per kilobase million (FPKM) were then calculated with cufflinks v2.2.0. Because the RNA-seq had a very high depth of coverage, the --max-bundle-frags argument was changed to 100,000,000 from its default value which is 1,000,000. This enabled the calculation of FPKMs at loci with high depth of coverage.

2.3.8 ***Gene fusion analysis and annotation of potential viral genomes***

FusionCatcher (v1.00) was used to identify novel gene fusions as well as validate known gene fusions in our HCCC samples (n=5; HCCC1, HCCC2, HCCC4, HCCC5, HCCC7). In addition to looking for gene fusions, FusionCatcher also provides a list of viral and bacterial reads seen in the samples. This output in addition to HPVDetector (102) was used to confirm that no detectable HPV DNA was identified in the HCCC samples.

2.3.9 ***Sanger sequencing validation of candidate molecular alterations and fusion genes***

Genomic DNA was used to verify somatic alterations including single nucleotide variations. 10ng of total DNA was PCR amplified with Platinum Taq DNA Polymerase High Fidelity (Invitrogen) according to the manufacturer's instructions. Primer sequences for each target are listed in **Table 2-3**. For amplicons greater than 200bp in length, PCR products were PCR purified using the Qiagen PCR purification kit and submit for Sanger sequencing at the University's DNA sequencing core on the 3730XL DNA Sequencer (Applied Biosystems). For amplicons less than 200bp in length, PCR products were cloned out using pCR8 TOPO vector (Invitrogen) and submitted for Sanger sequencing using forward and reverse primers from the cloning vector. Sequences were aligned using the DNASTAR Lasergene software suite against the RefSEQ annotation from hg19.

2.3.10 ***Hierarchical clustering of differentially expressed genes***

Sequences were realigned to the human genome (hg19) using Rsubread v1.32.2, to generate raw read counts. Genes with an average raw read count between 100 and 10,000 were included in further analysis as described (56). Differential expression analysis was performed with DESeq2 v1.18.1, comparing HCCC (n=5; HCCC1, HCCC2, HCCC4, HCCC5, HCCC7) and MEC (n=48) expression. We used an FDR-adjusted p value cutoff of 0.01 to define differentially expressed genes. Thus, by comparing to our previously analyzed mucoepidermoid carcinoma cohort (manuscript in review), variance between the

HCCC and MEC sample read counts was then calculated using DESeq2 v1.18.1, using a mean fit type, followed by regularized logarithmic transformation of the read counts. Samples underwent hierarchical clustering based on expression of differentially expressed genes.

2.3.11 ***Identifying MEC and HCCC-specific*** *transcription factor binding*

HCCC-specific and MEC-specific gene sets were generated, with an FDR adjusted p value cutoff of 0.1. Transcription factor binding sites associated with MEC- and HCCC-specific gene sets were queried using GenomeRunner db5.00, comparing each gene set to the ENCODE clustered transcription factor binding site database. We queried overlap between transcription factor binding sites and the genes in each gene set, including the gene body and 1000bp upstream of the transcription start site. To define the background, the gene body and 1000bp upstream region of all HG19 genes was used. Transcription factors with binding sites that are significantly enriched in each gene set were identified, using a p value cutoff of 0.01. Data was visualized using Complex Heatmap as previously described (103).

2.3.12 ***Gene set enrichment analysis***

Gene set enrichment analysis was performed using the software GSEA v4.03 from the Broad Institute (http://software.broadinstitute.org/gsea/index.jsp). Pre-ranked gene lists were prepared based on the log2-fold change in FPKM and the gene sets used were selected from the Molecular Signatures Database v7.0, including hallmark gene sets, gene ontology (GO) biological process (BP) gene sets and oncogenic signatures gene sets.

2.3.13 ***Immunohistochemistry***

Immunohistochemical staining was performed (n=7; HCCC1-7) on the DAKO Autostainer (DAKO, Carpinteria, CA) using liquid streptavidin biotin (LSAB+) and diaminobenzidine (DAB) as the chromogen as previously described (16, 104). De-paraffinized sections were labeled with the CXCR4 (1:500, clone UMB2, AbCam, Cambridge, MA. cat#Ab124824) for 60 minutes at ambient temperature after incubation of the section with background sniper (BioCare Medical, Pacheco, CA) for 30 mins at ambient temperature. Subsequently 10 mM Tris HCl/1 mM EDTA pH9 epitope retrieval was performed prior to staining. Appropriate negative (no primary antibody) and positive controls (kidney) were stained in parallel. Each sample was scored as a percent (0-100%) and intensity (0-3).

2.4 **Results**

Eight patient samples were obtained from our pathology archives and both demographic and clinical variable are shown in **Table 2-4**. Patient ages ranged from 41-58 years. Primary sites included base of tongue (BOT), hard palate, nasopharynx, and mandible. Overall stage varied from stage I to stage IV disease and two of the patients in this cohort presented with nodal disease. Two of the samples were initially classified as mucoepidermoid carcinoma and later changed to clear cell carcinoma based on *EWSR1-ATF1* fusion status. Mucicarmine stains were performed in two of the samples to differentiate from mucoepidermoid carcinoma and were both negative. All patients in this group underwent surgical excision, one of which was in the salvage setting. Three patients received additional radiation or chemoradiation post-operatively. Two patients recurred (one local and one regional recurrence) and a third had persistent disease after CRT. All patients showed no evidence of disease at their last follow up and total time since therapy ranged from 1 month to 17 years with a median follow up time of 3.4 years.

Of the 8 total cases, two HCCC specimens had DNA sufficient for integrated NGS based analysis (**Table 2-5)**, thus, we sequenced the exomes of both of these HCCC cases to greater than 100x depth and identified 142 high confidence alterations (79 somatic mutations and 63 INDELs) in HCCC1 and 28 high

confidence alterations (10 somatic mutations and 18 INDELs) in HCCC2. Detailed annotations of the alterations are provided in **Tables 2-6 and 2-7**, respectively. Surprisingly, we did not identify any recurrent molecular alterations or alterations to commonly mutated and cancer associated genes (e.g. *TP53, PIK3CA, KRAS*) in either tumor, but did identify an alteration of *NOTCH2NL* in HCCC1 (NM_001278267:c.5358-31971delC), supporting a potential disruptive role of NOTCH signaling in HCCC similar to the previously reported disruptive role in HNSCC (105), as well as an A299D mutation in TNK1, tyrosine kinase, non-receptor, 1.

We then performed functional annotation and gene set enrichment analysis of the altered genes in each tumor using the database for annotation, visualization and integrated discovery (DAVID). While this analysis did not identify any strong molecular enrichments in the genomic alterations of HCCC1, in HCCC2, the analysis revealed an enrichment of alterations to the p53 signaling pathway, with disruptive splice region variants in both *CCNB1* (cyclin B1, NM_004701:c.25-3delC) and *COP1* (NM_022457:c.1142-8delT). These genes are known to regulate sustained G2 arrest and negative feedback regulation of p53, respectively. Similarly, this tumor had an enrichment of disruptive mutations to ubiquitin kinases including: *COP1, DCAF5, PDZRN3, CUL2* and *UBR2*.

Next, we analyzed genomic copy number changes in the exome data and found that overall copy number was relatively stable in each tumor (**Figure 2-1, Tables 2-8 and 2-9**). We did observe a few genes in HCCC1 that had amplifications and several with single copy number gains including *IGF2R, TERT, AKT2* and *AURKC*, while HCCC2 had no genes with amplifications and several genes with single copy gains including *IGF2, HRAS* and *RPTOR*. Together these data suggested a potential role of IGF and PI3K/mTOR signaling in these tumors. In contrast, the tumors both had single copy deletions of *ATRX* and *BRWD3*, while HCCC1 also had single copy deletions of *CHEK2, RBL1* and *BRCA1* suggesting a potential defect in either DNA damage repair and/or cell cycle regulation.

We then performed high depth transcriptome sequencing on RNA from 5 of the HCCC tumors, including both HCCC1 and HCCC2, to an average depth of >50M reads (**Table 2-10**). Fusion status was determined in all five cases using the FusionCatcher algorithm. This confirmed that HCCC1, HCCC2, HCCC5, and HCCC7 harbored the canonical *EWSR1-ATF1* gene fusion, and identified two different fusion forms (in HCCC1 and HCCC2), which we confirmed by Sanger sequencing (**Figure 2-2A** and **2B**). Surprisingly, additional clinically relevant and previously described gene fusions common to lung cancer were also identified by the FusionCatcher in HCCC5 including *IGFBP5-ALK*, and *EML4-ALK* (**Table 2-11**). (106-108) To then determine the relationship of gene expression in HCCC relative mucoepidermoid carcinomas (MEC), we performed unsupervised hierarchical clustering on the 5 HCCC cases with 48 MEC tumors that we previously sequenced (**Figure 2-3A**). Importantly, the 5 HCCC cases clustered together and could be differentiated from the MEC tumors by a 354 gene signature (189 upregulated and 165 downregulated genes, **Tables 2-12 and 2-13,** respectively). To validate the molecular signatures, we analyzed gene signatures with altered expression specifically evaluating for transcription factor enrichments, hypothesizing that genes upregulated in the HCCC tumors would be enriched for ATF1 binding sites. For this analysis, the gene body and region 1000bp upstream of the transcription start site of each gene in each respective gene list were included in the enrichment analysis, and we utilized the ENCODE clustered transcription factor binding sites to define binding enrichment. Importantly, in the HCCC upregulated gene set, ATF1 binding sites were significantly overrepresented (FDR adjusted $p=5.6 \times 10^{-28}$), compared to background expression of all genes. These data strongly support the accuracy of our gene signature and suggest that additional analysis of our gene sets may help to identify pivotal genes and/or effectors of the EWSR1-ATF1 fusion.

Of the genes differentially expressed between HCCC and MEC, *IGF2* was upregulated exclusively in HCCC1, and *IGF1R* (Insulin growth factor receptor 1, adjusted p=0.0042) was significantly upregulated in all HCCCs relative to the MECs. Accordingly, *SGK3* (Serine/Threonine Protein kinase 3, adjusted $p=1.45 \times 10^{-11}$) and *SGK1* (Serine/Threonine Protein kinase 1, adjusted $p=1.07 \times 10^{-5}$) were significantly

upregulated in the HCCCs compared to our cohort of MECs (**Figure 2-3B**). Of note, *ERBB3* (Epidermal Growth Factor Receptor 3, or HER3, adjusted $p=6.64x10^{-7}$) was significantly downregulated in the HCCCs compared to MECs (**Figure 2-3C-F**). Subsequent pathway analysis of the differential gene sets with GSEA (Broad) demonstrated that our HCCC gene signature was enriched for 29 positively associated and 44 negatively associated concepts from the broad database with FDR q-value < 0.05 (**Table 2-14**). Positively enriched concepts included HALLMARK_MYC_TARGETS_V1 and several concepts associated with protein targeting to the endoplasmic reticulum (e.g., GO_PROTEIN_LOCALIZATION_TO_ENDOPLASMIC_RETICULUM), while negatively enriched concepts were associated with decreased humoral (GO_HUMORAL_IMMUNE_RESPONSE) and lymphocytic (GO_LYMPHOCYTIC_MEDIATED_IMMUNITY) immune expression relative to the MEC cohort (**Figure 2-3G-J**).

Despite these findings and given the recent clinical success of immune checkpoint inhibitors, we evaluated the expression of known immune checkpoints in the RNAseq data (**Figure 2-4A**). This identified *HLA-G* as modestly expressed at the RNA level in 3/5 of the HCCC tumors. Finally, because we noted strong enrichment of *SGK3* in the HCCCs, which is known to be a strong positive regulator of CXCR4 protein expression (109) through prevention of ubiquitin-mediated degradation of the checkpoint, and given the recent clinical success of CXCR4 antibodies, we sought to evaluate CXCR4 protein expression in our HCCC tumors. Immunohistochemistry confirmed expression of CXCR4 in 5/7 (71%) of our evaluable HCCC tumors, with high intensity staining in 2/7 (29%) cases (**Figure 2-4B-H**), suggesting that this may be an important immune escape mechanism for some HCCCs.

Finally, using HPVDetector, we assessed both the exomes and transcriptomes for the presence of reads mapping to any HPV type, and found that both of these tumors (HCCC1 and HCCC2) were HPV negative (data not shown). Likewise, MANTIS analysis demonstrated that both of these HCCC tumors were microsatellite stable (data not shown).

Collectively, our informatic results confirm the presence of the characteristic *EWSR1-ATF1* fusion and defines a unique ATF-enriched molecular signature that significantly differentiates our HCCC cases from a cohort of mucoepidermoid carcinomas.

2.5 **Discussion**

Our study overcomes a significant gap in knowledge by providing the first detailed view of the molecular composition of *EWSR1-ATF1* fusion positive HCCC cases. Here, we identified a unique 354 gene signature enriched in *EWSR1-ATF1* fusion positive HCCC compared to a cohort of *CRTC1-MAML2* positive, *CRTC3-MAML2* positive and *CRTC1/3-MAML2* negative mucoepidermoid carcinomas. Because of the rarity of HCCC, we are not yet positioned to validate the clinical utility of this signature. Our hope is that future molecular studies will be able to use this data set as an initiation point to refine and validate our gene signature to identify the minimal gene set capable of differentiating the two diseases.

While occurrence of *EWSR1-ATF1* gene fusions have been well documented in HCCC, our study is also the first to demonstrate the presence of an *ATF1*-enriched gene signature in the disease, which confirms that this fusion drives aberrant transcriptional activation of ATF1-regulated genomic loci in this disease. This is the first study to utilize clinical samples to confirm that ATF1-regulated genes are deregulated in HCCC, and supports development of further mechanistic studies to identify how these genes may play a role in disease pathogenesis. Indeed, these data supports mechanistic analogies to similar EWSR1-transcription factor fusions found in other cancers. For example, similar studies seeking to identify functionally pivotal effectors that are transcriptionally regulated by EWSR1-FLI1 or EWSR1-ERG in Ewing's Sarcoma have success identifying targets critical for Ewing's Sarcoma pathogenesis (110, 111).

Additionally, we also identified two potentially druggable targets in our samples. The DNA aberrations in the IGF/IGF1R signaling pathway and subsequent gene set enrichment analysis that identified an upregulation of PI3K/mTOR/FOXO pathway support a potential role for IGF1R inhibitor and several small molecule inhibitors targeting this pathway have already been advanced for other cancers (112, 113). Similarly, we identified a strong overexpression of CXCR4 in a subset of the tumors also representing a clear therapeutic target for which several therapies are being clinically evaluated (114-116). Interestingly, CXCR4 is associated with metastasis, which is not a strong clinical phenotype associated with HCCC; however, in tissues with elevated expression of the CXCR4 chemokine ligand CXCL12, CXCR4 expression tends to be associated with increased angiogenesis and tumor growth (116), which may be a critical mechanism for HCCC pathogenesis. Finally, we also discover the presence of additional fusions including *EML4-ALK* which has possible clinical implications as non-small cell lung cancer tumors harboring these fusions have shown to be sensitive to ALK inhibition. (106)

HCCC is an extremely rare cancer associated with fairly good prognostic outcomes, but a disease that can be difficult to identify without additional staining in some cases. Our study addresses the challenge of using molecular sequencing to differentiate HCCC samples and identifies a gene signature that is significantly different from MEC, and that can be evaluated in future studies containing larger case series of HCCCs. Although our study was limited to just seven samples, due to the extreme rarity of the disease, this comprehensive molecular analysis represents the first step forward in understanding the pathogenesis of disease. Our hope is that in the future, the molecular information learned in this study can serve as the basis for improved molecular annotation of these rare diseases in order to differentiate those associated with aggressive clinical phenotypes from those like HCCC with less aggressive clinical phenotypes that may benefit from de-escalated therapeutic strategies.

2.6 Figures

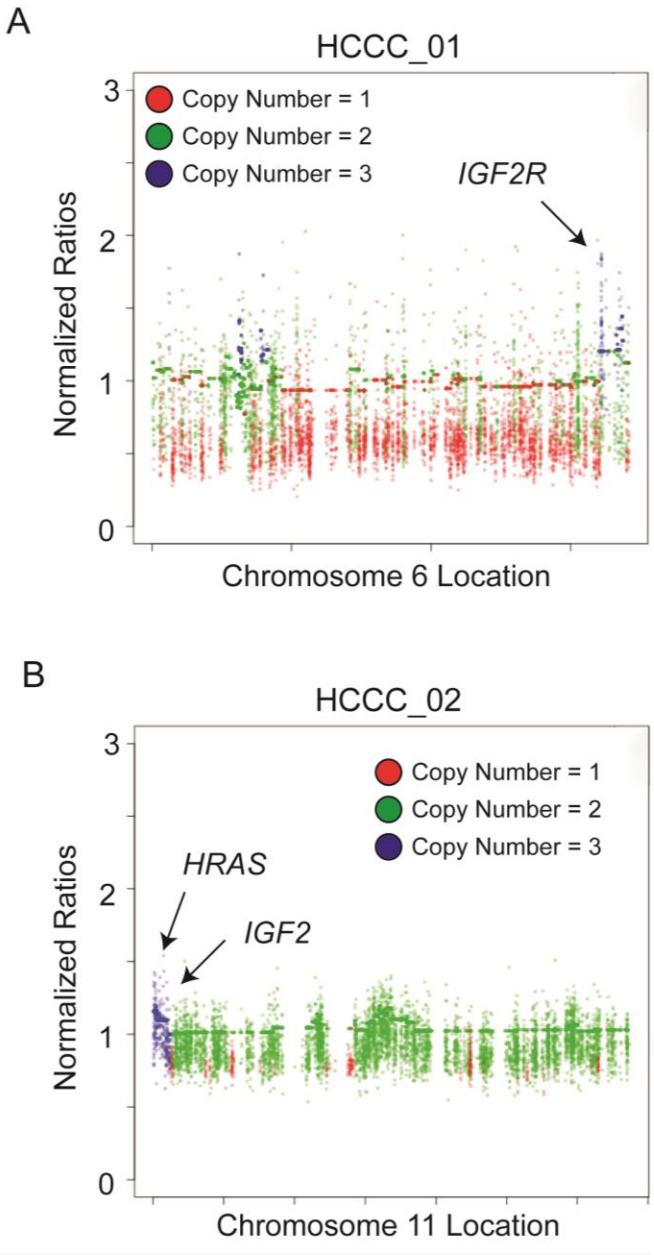


Figure 2-1. Identification of IGF2R and IGF2 copy number amplifications in two hyalinizing clear cell carcinomas.

A) Manhattan plot shows copy number amplifications or deletions across the HCCC1 exome sequencing data for Chromosome 6, with amplifications highlighted by blue dots and deletions annotated as red dots as indicated. Functionally prioritized genetic lesions are shown with respect to their genomic location on the plot as indicated. **B)** As in (A), with Manhattan plot for HCCC2 Chromosome 11.

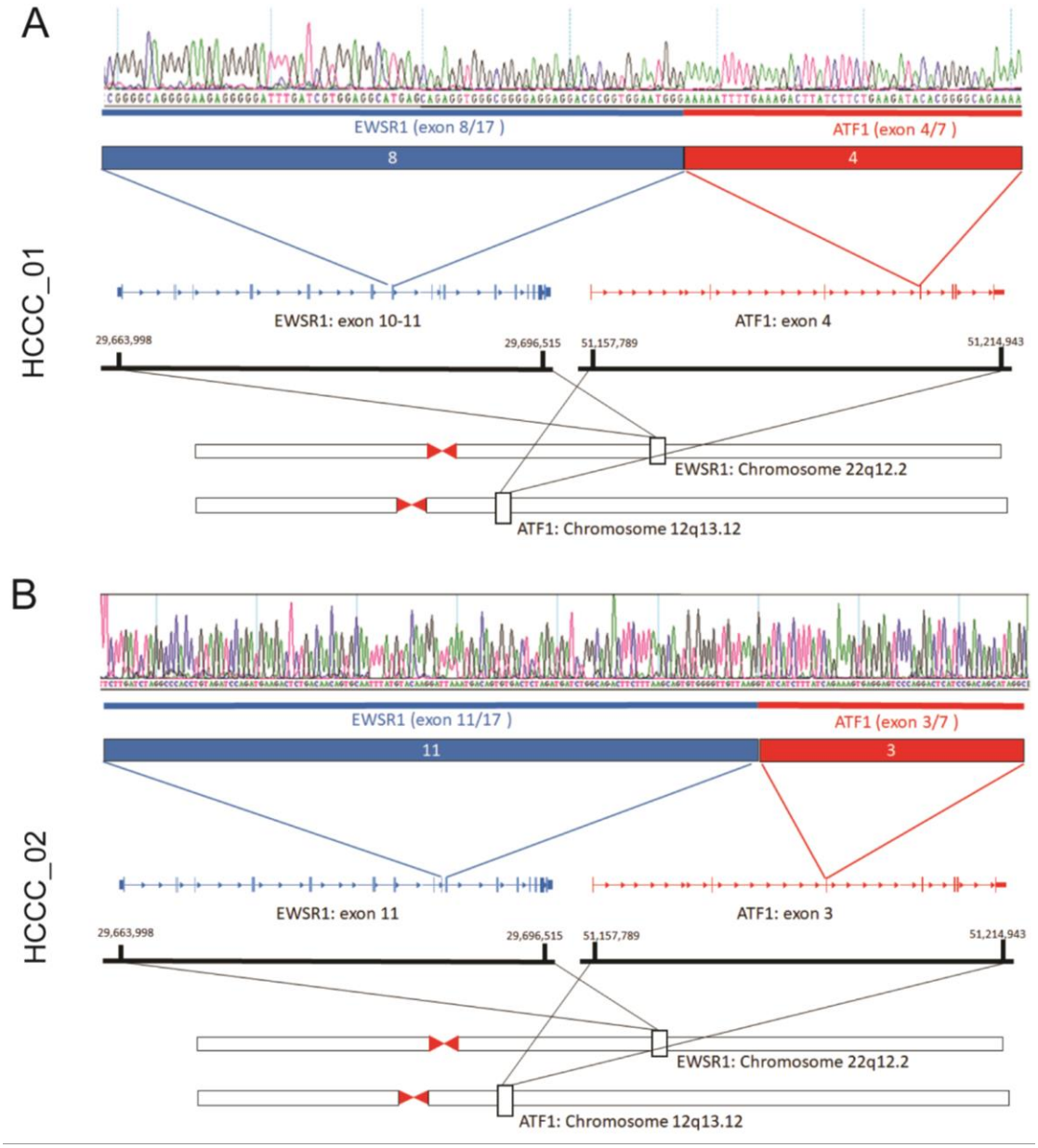


Figure 2-2. De novo gene fusion discovery confirms the EWSR1-ATF1 gene fusion structures in HCCC1 and HCCC2.

A) RNAseq data from HCCC1 was analyzed using the FusionCatcher bioinformatics algorithm, which nominated an EWSR1-ATF1 fusion gene junction in this tumor. The junction was validated by PCR amplification of tumor-derived cDNA, followed by Sanger sequencing (top panel), structure of the fusion shown below. **B)** HCCC2 also harbored a characteristic EWSR1-ATF1 fusion gene, but with a unique structure from HCCC1.

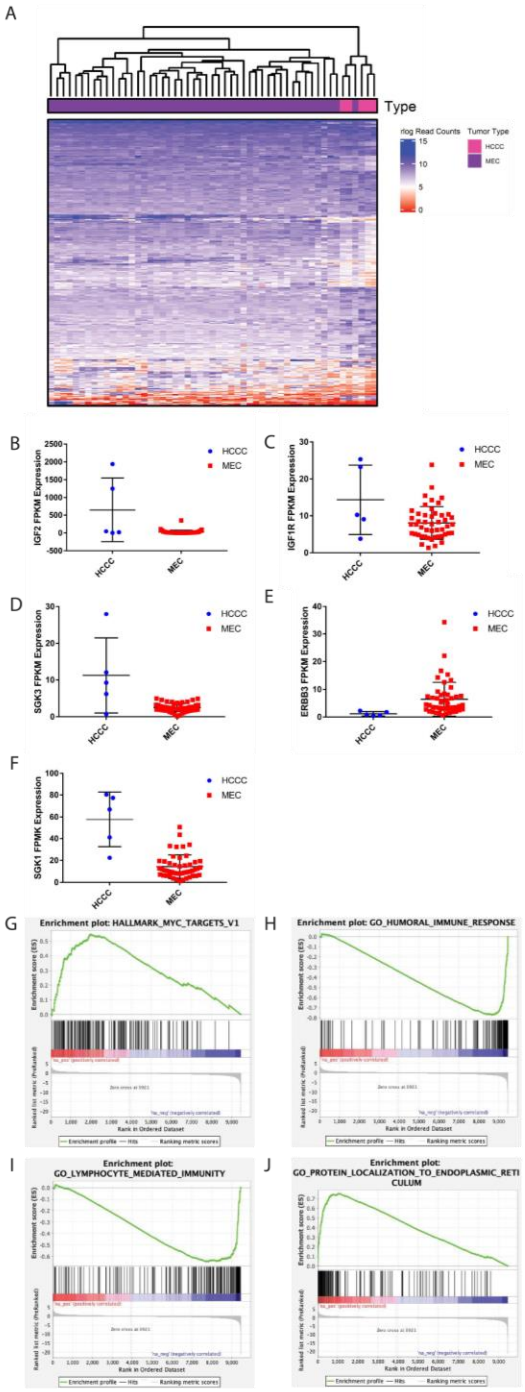


Figure 2-3. Identification of differential genes and pathways between HCCC and Mucoepidermoid Carcinoma.

A) Inverse rank normalized, log transformed RNAseq read counts from HCCC and 48 MECs were clustered using DESEQ2 to identify genes differentially expressed between the two tumor types, and complex heatmap was used to visualize the differential gene signature. **B)** Dot plot shows relative expression of *IGF2* in the HCCC and MEC cases as indicated, **C-F)** As in (B), with *IGF1R, ERBB3, SGK3,* and *SGK1* expression, **G-J)** Representative pathway enrichment analysis of gene sets differentially expressed between HCCC and MEC.

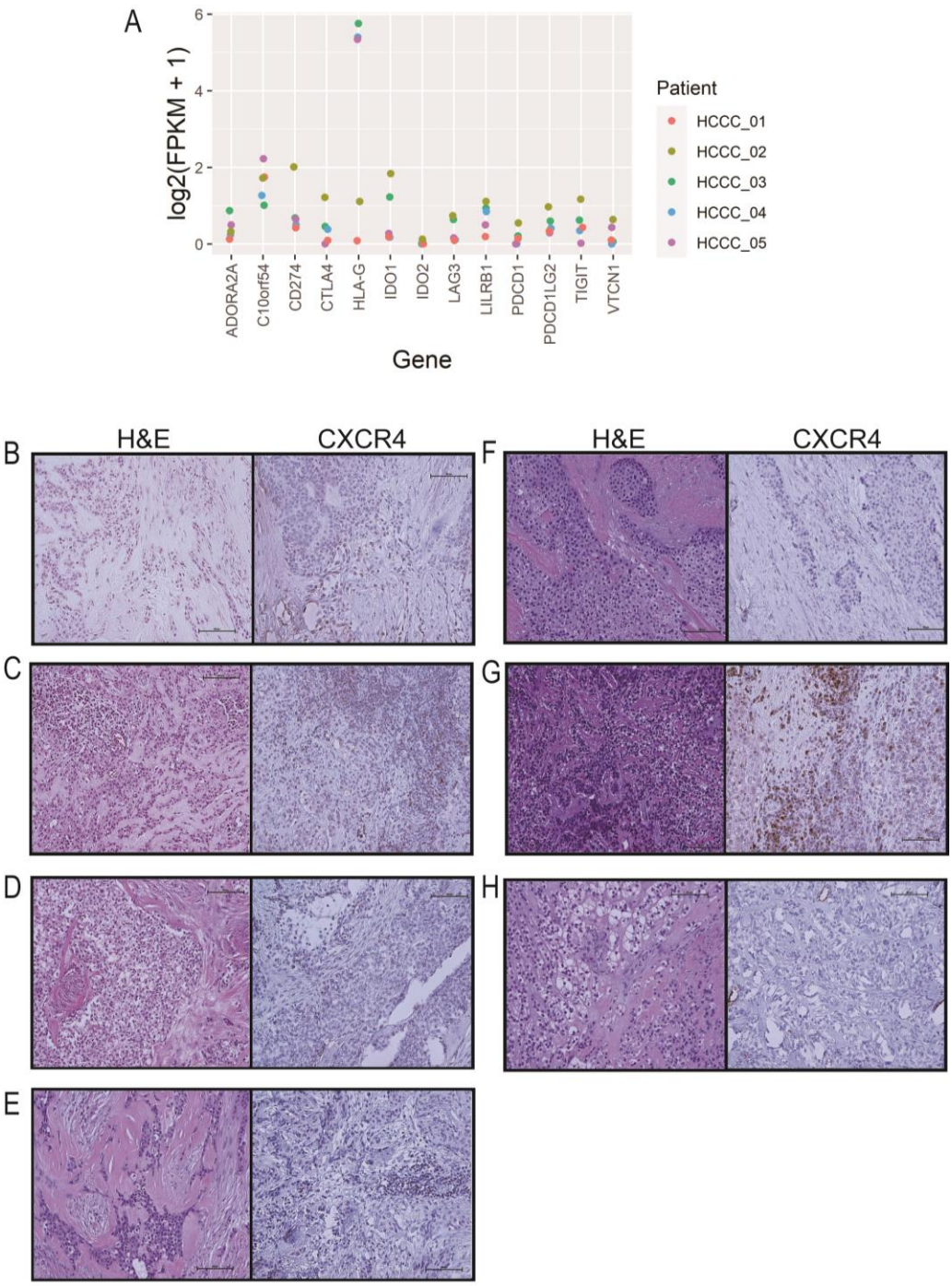


Figure 2-4. Characterization of immune checkpoint expression in HCCC.

A) RNA expression (FPKM) of immune checkpoints in our HCCC cohort. **B-H)** CXCR4 immunohistochemistry was performed on 7 different HCCC cases, and representative photographs were taken as indicated and are shown next to the representative hematoxylin and eosin stain. Each sample is scored as a percent (0-100%) and intensity (0-3). Images shown at 40x. **B)** HCCC1 - 10%, 1. **C)** HCCC2 - 50%, 3. **D)** HCCC3 - 15%, 2. **E)** HCCC4 - 20%, 1. **F)** HCCC5 - 0%, 0. **G)** HCCC6 - 50%, 3. **H)** HCCC7 - 0%, 0.

2.7 Tables

Sample ID	Sample name	Total Reads	%Mapped	%Uniquely mapped	Total unique reads
113377	HCCC_02-T	179968818	99.9	88.58	159417490
113378	HCCC_02-NL	157864500	99.94	87.87	138716439
113379	HCCC_01-T	161984416	99.88	86.94	140822724
116493	HCCC_01-NL	120995836	99.8	89.04	107730091

Table 2-1. Whole Exome Sequencing per Sample Read Depth

Name	Total Reads	% Uniquely Mapped	Non-Zero FPKMs	FPKM > 1
HCCC1	46371106	88	24936	11047
HCCC2	77887901	90.76	27469	12534
HCCC4	29283600	66.64	22374	11661
HCCC5	42814705	86.18	25095	11727
HCCC7	56828519	87.94	26069	12491

Table 2-2. RNA Sequencing per Sample Read Depth

Target		Primer	
EWSR1-ATF1 fusion (HCCC_01)		F -	CAGCAGAGTTCATTCCGACA
		R -	TTTCTGCCCCGTGTATCTTC
EWSR1-ATF1 fusion (HCCC_02)		F -	TCTTGATCTAGGCCCACCTG
		R -	GCCTATGCTGTCGGATGAGT
EBV		F -	CTGTGGTTGGGCAGGTACTGCCA
		R -	GGCTCTTGTGTCCAGGCATCCCT
HPV16	E1	F -	ATGGCTGATCCTGCAGGTACCAAT
		R -	TCATAATGTGTTAGTATTTTGTCCTGACACACA
	E2	F -	ATGGAGACTCTTTGCCAACGTTTA
		R -	TCATATAGACATAAATCCAGTAGACACTGTAAT
	E6	F-	ATGCACCAAAAGAGAACTGCAATG
		R-	TTACAGCTGGGTTTCTCTACGTGT

Table 2-3. Sanger Sequencing Primers

Age at Diagnosis	Gender	Site	Overall Stage	TNM stage	Special IHC stains	Smoking Status	Adjuvant Treatment	Recurrence	Disease Status	Time from treatment
55	F	Hard palate	IV	T4aN0M0	none	Former	no	no	AWOD	8 years
41	F	Hard palate	I	T1N0M0	none	Never	no	no	NED at last follow up	4 years
56	M	Mandible	unknown initial stage	N1 recurrence	none	Current	Second surgery for nodal recurrence	Regional	NED at last follow up	17 yrs from initial disease 5 yrs since nodal recurrence
41	F	BOT	IV	T4aN2bM0	none	Current	CRT	no	AWOD	1.8 years
58	F	Nasopharynx	III	T3N0M0	none	Never	salvage surgery after CRT	no	AWOD	1.6 years
54	M	BOT	II	T2N0M0	Positive for cytokeratin, EMA and vimentin (focally). PAS stain positive and mucicarmine negative	Never	no	Local	AWOD	17 years from initial disease 8 months local recurrence
57	F	BOT, FOM, and oral tongue	IV	T3N2bM0	none	Never	RT	no	AWOD	1 month
48	M	Nasopharynx	I	T1N0M0	Positive for p63 and focally/weakly for CAM5.2. Negative for calponin, C-kit, smooth muscle actin and S100. Mucicarmine stain is negative for intracytoplasmic mucin	Never	no	no	NED at last follow up	2.7 years

Table 2-4. HCCC Patient Demographics

HCCC Sample ID	DNA sequencing	Mutation Analysis	Copy Number and Microsatellite Instability Analyses	RNA seq.	Fusion Analysis	Immune Checkpoint Analysis	IHC
HCCC1	X	X	X	X	X	X	X
HCCC2	X	X	X	X	X	X	X
HCCC3							X
HCCC4				X	X	X	X
HCCC5				X	X	X	X
HCCC6							X
HCCC7				X	X	X	X

Table 2-5. Samples Used in Each Analysis

https://doi.org/10.1016/j.oraloncology.2021.105270

Table 2-6. Annotated SNVs from WES, Sample HCCC1

The table of SNVs is too large to include in this thesis. Please see the link above to download and view this spreadsheet, noted as **supplementary table 2.**

https://doi.org/10.1016/j.oraloncology.2021.105270

Table 2-7. Annotated SNVs from WES, Sample HCCC2

The table of SNVs is too large to include in this thesis. Please see the link above to download and view this spreadsheet, noted as **supplementary table 3.**

https://doi.org/10.1016/j.oraloncology.2021.105270

Table 2-8. Annotated CNVs from WES, Sample HCCC1

The list of CNVs is too large to include in this thesis. Please see the link above to download and view this spreadsheet, noted as **supplementary table 4.**

https://doi.org/10.1016/j.oraloncology.2021.105270

Table 2-9. Annotated CNVs from WES, Sample HCCC2

The list of CNVs is too large to include in this thesis. Please see the link above to download and view this spreadsheet, noted as **supplementary table 5.**

https://doi.org/10.1016/j.oraloncology.2021.105270

Table 2-10. RNA Expression per Sample (FPKM), HCCC Tumors

This gene expression table is too large to include in this thesis. Please see the link above to download and view this spreadsheet, noted as **supplementary table 6.**

https://doi.org/10.1016/j.oraloncology.2021.105270

Table 2-11. All Gene Fusions Called, HCCC Tumors

This table is too large to include in this thesis. Please see the link above to download and view this spreadsheet, noted as **supplementary table 7.**

Gene Name	log2FoldChange	padj
SLC6A12	7.030908	1.03E-05
CNGB1	7.005136	0.000234
CHGB	5.313711	0.004071
DLGAP2	5.082608	0.000463
CDH2	5.078345	5.86E-11
RPS27	4.575375	2.42E-28
RPS18	4.348504	2.42E-28
RPL13A	4.219707	4.10E-17
ROBO2	4.150778	7.50E-08
FREM2	4.058239	0.005547
GPR12	3.922688	0.002938
TRHDE-AS1	3.873202	0.000311
ADGRB1	3.664186	0.001037
RPL9	3.556297	5.23E-14
EDAR	3.427436	0.000222
SEMA3A	3.390962	0.000552
MT1X	3.382786	3.47E-05
SDK2	3.303375	9.07E-07
PTHLH	3.286105	0.008538
NFASC	3.239483	3.68E-07
SSBP2	3.233644	1.10E-11
RPL21	3.131889	2.27E-15
CCDC3	3.097212	0.000676
RPL12	3.060467	9.41E-14
COL18A1	3.012635	1.66E-06
CSPG4P8	2.995836	3.84E-06
SEMA6D	2.947252	0.000613
DHCR7	2.923846	0.000108
LRRN1	2.900811	1.40E-05
CHL1	2.885184	1.67E-05
SPTBN5	2.798044	3.03E-05
HLA-DRA	2.770014	0.000179
RPL7	2.765762	2.60E-10
BANP	2.754003	1.09E-12
NPHP4	2.729028	4.10E-08
CRYAB	2.721135	0.005602
AIM1	2.701454	3.23E-06
GLCE	2.658661	0.000102
DEAF1	2.587719	1.16E-08
NDUFA4L2	2.577623	0.002709
ARHGAP42	2.541528	3.63E-09
DDX39B	2.541371	9.41E-14
gcnt2	2.525286	1.58E-05

TSPAN18	2.509204	0.000282
SLC24A3	2.490779	0.000222
znf208	2.483323	0.003031
TPPP	2.470395	0.000182
ADGRG6	2.433573	0.002502
HLA-DPA1	2.419305	0.001057
SLC46A3	2.357157	0.000105
RPS10	2.35358	3.20E-06
NLGN4X	2.352234	0.003109
PLB1	2.345298	4.97E-07
RPL17	2.342729	3.81E-11
PRRC2A	2.309015	1.41E-10
rps15a	2.302841	5.43E-09
SGK3	2.277598	1.45E-11
RPS7	2.209828	2.05E-08
SGK1	2.209691	1.07E-05
ARL17A	2.204186	5.28E-06
CSPG4	2.171375	0.006795
GPR183	2.16498	8.21E-07
pcsk1	2.160904	0.006038
BAG6	2.136958	2.70E-09
PPP1R10	2.136767	1.27E-10
QSOX2	2.107446	4.28E-07
GABBR1	2.071441	3.79E-05
PIEZO2	2.07051	0.009945
RPL18A	2.056594	4.28E-07
SGTA	2.044689	0.000108
ddr1	2.041227	0.00028
TMEM80	1.979112	2.01E-06
HLA-E	1.966558	4.09E-06
DKK3	1.964416	0.002898
PLPP1	1.942068	8.42E-05
BRD2	1.8757	1.29E-14
POR	1.872547	0.000408
TENM4	1.870217	0.004648
LOC100133331	1.832227	0.001528
PDZD2	1.821373	0.003471
znf91	1.802828	0.000186
RPL7A	1.7961	0.000443
ERN1	1.792727	0.000342
KLHL42	1.764211	2.67E-08
TSPAN9	1.75239	9.07E-07
RPSA	1.723366	0.000813
HSF4	1.722943	0.007969

ALDH7A1	1.709132	1.07E-05
ARHGEF35	1.705155	0.0002
PLXNB1	1.685222	0.002993
RPS3A	1.675989	2.39E-06
SMG1P3	1.662517	7.17E-06
RPS12	1.659012	9.25E-07
RPL24	1.65687	4.24E-05
ZNF254	1.650862	0.000462
EEF1A1	1.644962	3.98E-06
FZD3	1.64126	0.001673
KAZN	1.633148	0.000221
PTPRG	1.597473	0.000148
RPS21	1.594934	0.008538
STAT5B	1.594199	0.001982
SSBP3	1.593252	0.00111
FAT1	1.591847	0.001209
NCKAP5	1.583653	0.006812
XYLT1	1.579381	0.000387
CHST3	1.572221	0.001843
FAM118A	1.564423	0.001759
NAIP	1.563461	0.000744
MCAM	1.563447	0.000741
RPL10	1.563244	2.39E-06
TRIB2	1.561379	2.86E-08
ANGPT2	1.549888	0.000626
EFNB2	1.537085	0.000111
RPLP0	1.53284	0.000374
ADCY9	1.53027	0.001936
NTN4	1.524956	0.002545
GAK	1.522989	2.98E-05
EEF1B2	1.511792	0.000107
NRBP2	1.510602	0.003344
LOC100190986	1.508517	0.002177
FKBP9	1.504083	0.000114
ATF6B	1.482452	6.15E-06
TOMM7	1.47451	0.002953
TSPYL2	1.471754	0.003854
ENOX2	1.470961	2.01E-06
TPCN2	1.458853	0.002496
LPGAT1	1.447396	0.000392
SDHA	1.432675	0.003474
IPPK	1.42894	0.004937
NPM1	1.412178	0.001735
RPS13	1.412119	0.000261

AGPAT5	1.411293	0.002329
TAPBP	1.405718	1.40E-05
LPIN1	1.379489	0.009879
RPLP1	1.379257	0.001528
gas5	1.356873	0.001721
ACSL4	1.341609	0.006094
ZFAND5	1.3404	0.004119
RPL4	1.324092	0.000172
c1orf198	1.323608	0.002789
RPL32	1.323007	4.28E-05
DOCK4	1.314953	0.004965
RPL23A	1.297498	0.000129
HSP90AB1	1.291297	0.00111
LOC100505818	1.285796	0.000688
pom121	1.285566	6.38E-07
MAST4	1.263743	0.001626
RPS25	1.252096	0.005377
CA5B	1.246561	0.007423
SH3RF1	1.246045	0.000463
RPL37	1.243132	2.73E-05
KIAA1549	1.235248	0.000429
RPS27A	1.233984	0.009195
RPL10A	1.23246	0.001148
IP6K2	1.222036	2.98E-05
PM20D2	1.218167	0.002332
ECI2	1.206823	0.003594
MAP2K1	1.181881	0.00134
BRD7	1.171813	0.000823
RPL31	1.167742	9.06E-06
FAM122A	1.165842	0.001569
ANKH	1.15247	0.006137
EIF3F	1.111862	0.001787
NIPAL3	1.086306	0.001787
RALGAPA1	1.085201	0.000903
IK	1.083492	0.00416
MAN2A1	1.077764	0.008848
AGAP1	1.062257	0.005312
RPL5	1.060409	0.000831
RPL14	1.056248	0.009945
INTS7	1.053802	0.000125
HNRNPL	1.02403	5.28E-06
RNPS1	1.021377	0.002763
IGF1R	0.951146	0.004218
HNRNPA1	0.896614	0.008574

RBBP7	0.842065	0.006898
PA2G4	0.835678	0.002221
RPS11	0.827511	0.008324
RPL23	0.820756	0.008544
HPS3	0.811087	0.00082
FBXL4	0.805321	0.004834
HNRNPK	0.804009	0.000354
NAP1L4	0.802562	0.001982
DHX9	0.789186	0.005803
ARGLU1	0.761658	0.006298
DCAKD	0.75563	0.009517
TAF1	0.722831	0.005501
SUPT16H	0.722346	0.000534
GORASP1	0.705667	0.008538

Table 2-12. Differential Expression Analysis Results. Genes with Elevated RNA Expression in HCCC, Compared to MEC

Gene Name	log2FoldChange	padj
BPIFA2	-21.61280841	8.45E-09
TCN1	-7.985873512	9.11E-09
SLPI	-5.299631918	1.16E-08
ATP6V1C2	-4.049320904	2.89E-08
SOX7	-3.308712708	1.09E-07
EHF	-3.366442649	3.68E-07
MUC7	-10.20890555	3.83E-07
DTX4	-2.438445137	6.38E-07
ERBB3	-2.910394898	6.64E-07
CPA4	-6.531035809	9.25E-07
MTSS1	-1.874347589	1.06E-06
TLE4	-1.368608413	2.14E-06
MACC1	-3.281952188	2.39E-06
PON3	-6.250982263	2.56E-06
STATH	-7.943482981	7.70E-06
PIGR	-6.332849082	1.07E-05
GPRC5A	-4.001226587	1.07E-05
ST6GALNAC1	-3.933064004	1.62E-05
LOC100132287	-2.643282563	2.20E-05
EPB41L1	-1.478886704	3.18E-05
CLIC6	-3.670533189	3.79E-05
LOC100132062	-2.336391084	3.79E-05
ARFGEF3	-3.246186464	4.02E-05
NBPF24	-1.83717648	4.28E-05
RNF144B	-1.826986227	4.28E-05
SYT12	-4.43086479	4.93E-05
GALNT12	-3.048074681	4.97E-05
BDKRB2	-3.886110142	6.98E-05
AZGP1	-6.091392732	7.47E-05
GABRE	-2.626157339	8.64E-05
KCNJ12	-4.562726329	8.66E-05
SLC6A14	-4.629258576	8.75E-05
SPDYE6	-1.663920379	9.54E-05
VTCN1	-3.941988704	0.000108
MALL	-2.815583784	0.000108
SPDYE2B	-1.695102315	0.000129
XBP1	-2.354011904	0.000132
PSCA	-5.859866392	0.000132
ICA1	-1.929703632	0.000153
CCL28	-4.207514067	0.000155
MDFIC	-1.430295321	0.000157
SNCAIP	-3.345348222	0.000176
SMR3B	-12.15618379	0.000179

fam21a	-0.941497462	0.000193
MPP7	-2.176122299	0.000216
RGPD6	-1.669631125	0.000222
KRT4	-4.794083057	0.000257
LINC00689	-6.574037983	0.00034
NUDT4P1	-2.039306403	0.000393
LOC100287119	-2.644579324	0.000393
ATP8B1	-1.15097907	0.000494
RPS6KA3	-0.832046164	0.000551
ANXA3	-3.320834008	0.000662
RGPD5	-1.135328738	0.000666
TBC1D1	-1.256975291	0.000714
GCNT3	-3.887265633	0.0008
CACUL1	-0.761026759	0.00082
SCG2	-5.808470038	0.000837
DMD	-2.605826154	0.000837
SUSD1	-1.505756214	0.000846
SMG1P1	-1.328931691	0.000861
MAL2	-3.076252009	0.0009
B4GALT5	-1.388995247	0.000902
CRLF1	-4.384883684	0.000957
AQP5	-6.677138078	0.001045
NBPF8	-2.084123662	0.001057
AGPAT3	-1.032771903	0.001063
ATP10B	-3.509072963	0.001282
KLF3	-1.019561468	0.001286
PRR5L	-2.042551756	0.001305
NBPF16	-1.491610082	0.00149
RARRES1	-3.76638162	0.001551
VWCE	-3.761416835	0.001553
CA12	-3.304715494	0.001582
RGPD8	-1.121915459	0.001582
CAPN2	-0.874603012	0.001603
SIDT1	-2.397103609	0.001631
WFDC2	-4.262128938	0.001658
fxyd2	-4.274301491	0.001749
HTN1	-8.885467795	0.001749
CCDC69	-1.967948729	0.001782
KRT13	-5.330704369	0.001787
ANKS6	-1.460686238	0.001787
SMOX	-2.431958461	0.001795
POTEE	-1.811674665	0.001936
SDIM1	-4.5996963	0.001936
SNX30	-0.979608393	0.001936

LOC100505669	-1.977722819	0.001953
RALGPS2	-1.544502388	0.00197
TAOK3	-0.762251876	0.001981
SIX1	-2.179863682	0.002288
KRT80	-2.816796437	0.00238
FXYD4	-7.330138685	0.002484
CLCA4	-4.077278814	0.002508
CPEB2	-1.448066364	0.00257
GALNT5	-1.765476085	0.002586
ATXN1	-0.762821107	0.002709
TMC5	-3.122838608	0.002924
SRGAP1	-1.04308096	0.003038
ZG16B	-4.967273756	0.003038
LINC01002	-2.458526093	0.003064
RASL10B	-4.818758002	0.003109
KCNK6	-1.735338272	0.003109
RDH10	-2.348668695	0.003109
slc5a5	-4.758381367	0.003123
MYO5C	-1.923556738	0.003257
NBPF20	-1.364478716	0.00343
CARD6	-1.408886208	0.003503
SPDYE2	-1.313409281	0.003594
CMYA5	-2.643170177	0.003951
POLR2J3	-1.21003906	0.00396
Ceacam6	-4.219568234	0.004071
SLC16A7	-1.752603311	0.004073
DDIT4L	-3.003326015	0.004114
AFF1	-0.655595854	0.004141
DMBT1	-3.784542169	0.00428
ZNF697	-1.320743564	0.004307
CD59	-1.278872747	0.0045
PLEKHB1	-3.144816221	0.004693
ALDH1L1	-3.601553748	0.004693
RASD1	-2.499323415	0.004735
SH3KBP1	-1.092730926	0.004884
TMSB4XP8	-1.849915563	0.005108
STK39	-1.650971155	0.005346
SERPINB3	-4.892415507	0.005382
CACHD1	-1.527918167	0.005437
NPIPB13	-1.611027624	0.005442
SEL1L3	-2.308227673	0.005445
SUMO1P3	-0.967478939	0.005496
PSD4	-1.830917365	0.005502
ca6	-9.638623187	0.005543

OSBP2	-2.712486862	0.005597
SAA1	-4.302655303	0.005602
EMB	-1.792099952	0.00567
SAA2	-4.265075491	0.005676
PRKAA2	-2.334821208	0.00587
CA13	-1.75414947	0.005999
AMPD3	-1.559349521	0.006014
CGN	-2.349069719	0.006038
FLJ23867	-2.169231334	0.006038
SPIRE1	-1.080640091	0.006208
LNX1	-1.163518785	0.006208
PLEKHA7	-1.489358304	0.006298
LZTS3	-2.138830171	0.006566
NPIPB3	-1.348978923	0.006795
ALPL	-2.267499452	0.006916
DOPEY2	-1.196961358	0.006995
STX3	-0.836855565	0.007234
IQGAP2	-1.675560524	0.007648
NCCRP1	-3.843047316	0.007793
CASP10	-1.4359001	0.007914
MAOA	-2.305228017	0.007917
CHST6	-2.261444247	0.007951
RGPD4	-0.846380988	0.007951
CREG1	-1.139023081	0.008574
TTC9	-2.223006433	0.008574
CST5	-7.194863167	0.008848
TTC3P1	-1.235614334	0.00889
TMEM183B	-1.32057726	0.008994
FAM3D	-5.350188928	0.009142
ighd	-4.683898803	0.009277
MEF2C	-1.685740847	0.009282
SSUH2	-5.105825596	0.009885
ERMP1	-1.087765796	0.009902
MBD2	-1.019008743	0.009945

Table 2-13. Differential Expression Analysis Results. Genes with Decreased RNA Expression in HCCC, Compared to MEC

Correlation	NAME	SIZE	ES	NES	NOM p-val	FDR q-val
Negative	GO_COMPLEMENT_ACTIVATION	72	-0.828	-2.30	0.000	0.000
Negative	GO_HUMORAL_IMMUNE_RESPONSE	132	-0.772	-2.29	0.000	0.000
Negative	GO_REGULATION_OF_HUMORAL_IMMUNE_RESPONSE	65	-0.818	-2.25	0.000	0.000
Negative	GO_HUMORAL_IMMUNE_RESPONSE_MEDIATED_BY_CIRCULATING_IMMUNOGLOBULIN	68	-0.802	-2.24	0.000	0.000
Negative	GO_DEFENSE_RESPONSE_TO_BACTERIUM	110	-0.756	-2.21	0.000	0.000
Negative	GO_ANTIMICROBIAL_HUMORAL_RESPONSE	43	-0.796	-2.11	0.000	0.000
Negative	GO_FC_RECEPTOR_MEDIATED_STIMULATORY_SIGNALING_PATHWAY	101	-0.719	-2.10	0.000	0.000
Negative	GO_B_CELL_MEDIATED_IMMUNITY	106	-0.711	-2.07	0.000	0.000
Negative	GO_B_CELL_RECEPTOR_SIGNALING_PATHWAY	57	-0.746	-2.00	0.000	0.001
Negative	GO_MEMBRANE_INVAGINATION	66	-0.716	-2.00	0.000	0.001
Negative	GO_IMMUNOGLOBULIN_PRODUCTION	81	-0.696	-1.98	0.000	0.003
Negative	KRAS.LUNG_UP.V1_DN	61	-0.724	-1.97	0.000	0.004
Negative	GO_PHAGOCYTOSIS_RECOGNITION	34	-0.777	-1.95	0.000	0.005
Negative	GO_LYMPHOCYTE_MEDIATED_IMMUNITY	160	-0.648	-1.95	0.000	0.005
Negative	GO_PHAGOCYTOSIS	203	-0.639	-1.95	0.000	0.005
Negative	GO_ADAPTIVE_IMMUNE_RESPONSE_BASED_ON_SOMATIC_RECOMBINATION_OF_IMMUNE_RECEPTORS_BUILT_FROM_IMMUNOGLOBULIN_SUPERFAMILY_DOMAINS	171	-0.635	-1.92	0.000	0.008
Negative	GO_MUSCLE_FILAMENT_SLIDING	25	-0.801	-1.91	0.000	0.011
Negative	GO_FC_EPSILON_RECEPTOR_SIGNALING_PATHWAY	117	-0.651	-1.91	0.000	0.011
Negative	GO_RECEPTOR_MEDIATED_ENDOCYTOSIS	188	-0.618	-1.91	0.000	0.011
Negative	GO_RESPONSE_TO_BACTERIUM	313	-0.612	-1.90	0.000	0.011
Negative	GO_POSITIVE_REGULATION_OF_B_CELL_ACTIVATION	64	-0.682	-1.90	0.000	0.011

Negative	GO_DRUG_TRANSPORT	63	-0.679	-1.88	0.000	0.016
Negative	GO_DIGESTION	35	-0.726	-1.86	0.000	0.024
Negative	GO_ADAPTIVE_IMMUNE_RESPONSE	216	-0.602	-1.86	0.000	0.025
Negative	GO_NEGATIVE_REGULATION_OF_PEPTIDASE_ACTIVITY	134	-0.625	-1.85	0.000	0.024
Negative	GO_LEUKOCYTE_MIGRATION	258	-0.600	-1.85	0.000	0.025
Negative	MEL18_DN.V1_DN	93	-0.637	-1.84	0.000	0.030
Negative	IL21_UP.V1_UP	65	-0.663	-1.83	0.000	0.035
Negative	GO_FC_RECEPTOR_SIGNALING_PATHWAY	169	-0.605	-1.83	0.000	0.034
Negative	KRAS.PROSTATE_UP.V1_DN	46	-0.687	-1.82	0.001	0.034
Negative	GO_MYOFIBRIL_ASSEMBLY	47	-0.689	-1.82	0.000	0.037
Negative	ALK_DN.V1_UP	45	-0.691	-1.82	0.001	0.037
Negative	BMI1_DN_MEL18_DN.V1_DN	86	-0.629	-1.81	0.000	0.041
Negative	GO_IMMUNE_RESPONSE_REGULATING_CELL_SURFACE_RECEPTOR_SIGNALING_PATHWAY	278	-0.584	-1.81	0.000	0.040
Negative	KRAS.LUNG.BREAST_UP.V1_DN	54	-0.674	-1.81	0.000	0.040
Negative	GO_DIGESTIVE_SYSTEM_PROCESS	33	-0.731	-1.80	0.000	0.040
Negative	LEF1_UP.V1_DN	102	-0.626	-1.80	0.000	0.040
Negative	GO_CELL_RECOGNITION	88	-0.635	-1.80	0.000	0.041
Negative	GO_NEGATIVE_REGULATION_OF_ENDOTHELIAL_CELL_PROLIFERATION	24	-0.763	-1.80	0.000	0.040
Negative	GO_CORNIFICATION	48	-0.681	-1.79	0.000	0.043
Negative	GO_O_GLYCAN_PROCESSING	29	-0.731	-1.79	0.000	0.043
Negative	KRAS.50_UP.V1_DN	16	-0.833	-1.79	0.001	0.044
Negative	GO_KERATINIZATION	55	-0.668	-1.78	0.000	0.048
Negative	GO_REGULATION_OF_IMMUNE_EFFECTOR_PROCESS	224	-0.581	-1.78	0.000	0.050
Positive	GO_COTRANSLATIONAL_PROTEIN_TARGETING_TO_MEMBRANE	91	0.829	3.17	0.000	0.000
Positive	GO_NUCLEAR_TRANSCRIBED_MRNA_CATABOLIC_PROCESS_NONSENSE_MEDIATED_DECAY	107	0.800	3.12	0.000	0.000
Positive	GO_ESTABLISHMENT_OF_PROTEIN_LOCALIZATION_TO_ENDOPLASMIC_RETICULUM	101	0.808	3.06	0.000	0.000
Positive	GO_PROTEIN_LOCALIZATION_TO_ENDOPLASMIC_RETICULUM	123	0.754	2.95	0.000	0.000
Positive	GO_PROTEIN_TARGETING_TO_MEMBRANE	150	0.711	2.92	0.000	0.000
Positive	GO_TRANSLATIONAL_INITIATION	162	0.714	2.89	0.000	0.000

Positive	GO_NUCLEAR_TRANSCRIBED_MRNA_CATABOLIC_PROCESS	160	0.710	2.88	0.000	0.000
Positive	GO_VIRAL_GENE_EXPRESSION	165	0.693	2.78	0.000	0.000
Positive	GO_RNA_CATABOLIC_PROCESS	289	0.599	2.77	0.000	0.000
Positive	GO_ESTABLISHMENT_OF_PROTEIN_LOCALIZATION_TO_MEMBRANE	239	0.599	2.54	0.000	0.000
Positive	GO_PROTEIN_TARGETING	299	0.545	2.46	0.000	0.000
Positive	GO_RIBOSOME_ASSEMBLY	33	0.821	2.45	0.000	0.000
Positive	GO_CYTOPLASMIC_TRANSLATION	74	0.665	2.42	0.000	0.000
Positive	GO_RIBOSOMAL_LARGE_SUBUNIT_BIOGENESIS	38	0.750	2.36	0.000	0.000
Positive	GO_RIBOSOMAL_SMALL_SUBUNIT_BIOGENESIS	40	0.706	2.26	0.000	0.000
Positive	GO_RIBOSOME_BIOGENESIS	164	0.544	2.26	0.000	0.000
Positive	GO_RIBONUCLEOPROTEIN_COMPLEX_BIOGENESIS	261	0.500	2.21	0.000	0.000
Positive	GO_CELLULAR_NITROGEN_COMPOUND_CATABOLIC_PROCESS	363	0.488	2.16	0.000	0.002
Positive	HALLMARK_MYC_TARGETS_V1	159	0.547	2.13	0.000	0.002
Positive	GO_RIBONUCLEOPROTEIN_COMPLEX_SUBUNIT_ORGANIZATION	134	0.539	2.12	0.000	0.002
Positive	GO_PEPTIDE_BIOSYNTHETIC_PROCESS	411	0.488	2.12	0.000	0.002
Positive	GO_ESTABLISHMENT_OF_PROTEIN_LOCALIZATION_TO_ORGANELLE	398	0.467	2.11	0.000	0.003
Positive	GO_AMIDE_BIOSYNTHETIC_PROCESS	489	0.428	2.01	0.000	0.012
Positive	GO_ORGANIC_CYCLIC_COMPOUND_CATABOLIC_PROCESS	382	0.437	1.96	0.000	0.026
Positive	GO_SEMI_LUNAR_VALVE_DEVELOPMENT	23	0.693	1.95	0.004	0.028
Positive	GO_RRNA_METABOLIC_PROCESS	123	0.495	1.92	0.000	0.039
Positive	GO_AORTIC_VALVE_DEVELOPMENT	20	0.702	1.91	0.000	0.039
Positive	GO_NEGATIVE_REGULATION_OF_UBIQUITIN_DEPENDENT_PROTEIN_CATABOLIC_PROCESS	37	0.588	1.91	0.004	0.041
Positive	GO_PEPTIDE_METABOLIC_PROCESS	491	0.421	1.90	0.000	0.040

Table 2-14. Gene Set Enrichment Analysis (GSEA) of the HCCC Gene Signature.

Chapter 3 Genome-wide Long Read Sequencing Resolves the Structure of Recurrent Genomic Rearrangements and Discovers TERT Driver Alterations in Mucoepidermoid Carcinoma

3.1 Abstract

Mucoepidermoid Carcinoma (MEC) is the most common malignant neoplasm of the salivary glands. Approximately 60% of MECs harbor genomic rearrangements between *CRTC1* or *CRTC3* and *MAML2* that are thought to drive disease pathogenesis; however, the precise structural mechanism driving this rearrangement has remained uncharacterized. Here, we performed multi-omics long read and linked read genome sequencing analysis and identified the rearrangement, deletion, and duplication events that form the *CRTC1-MAML2* translocation in MEC. We discover that a chain of genomic events creates not only the *CRTC1-MAML2* fusion, but also an unexpected *MAML2* to *MYBL1* rearrangement, suggesting that MYBL1 may have a larger role in salivary gland cancers than previously recognized. Further, we also discover and validate recurrent *TERT* promoter rearrangements and copy number amplifications in MEC models. Subsequent exome sequencing of a cohort MEC tumors identified a high recurrence rate of both *TERT* and *NOTCH2* alterations, and break apart FISH demonstrated that 4/4 MEC cell lines had *TERT* rearrangement and 4/4 had *TERT* copy number gain. Finally, TERT knockdown in the NCI-H292 cell line with *TERT* rearrangement reduced clonogenic cell survival supporting a critical role of the gene in MEC tumorigenesis. Overall, our data suggests that structural

variations are common in MEC, and that these events tend to occur through complex chromothripsis rearrangement mechanisms in both *CRTC1-MAML2* fusion positive and negative cases. Our data also identifies recurrent *TERT* and *NOTCH2* alterations as likely novel drivers of MEC oncogenesis, representing the first recurrent alterations in this disease beyond the canonical CRTC1/3-MAML2 translocation.

***Significance*.**

We use genome wide approaches to resolve the genetic mechanism leading to *CRTC1-MAML2* rearrangement in mucoepidermoid carcinoma (MEC), and, in the process, we also discovered both recurrent *TERT* rearrangements and alterations as well as *NOTCH2* disruptive alterations that drive malignant phenotypes in MEC models. These data suggest that MECs may be pre-disposed to genome rearrangements and are often dependent on *TERT* activation and *NOTCH2* disruption for neoplastic transformation. This study is the first to characterize recurrent alterations in *CRTC1-MAML2* negative MEC.

3.2 **Introduction**

Mucoepidermoid carcinoma (MEC) is one of the most common salivary malignancies, accounting for between 30-40% of all salivary gland malignancies in adults and 50% in children (117-119). The disease commonly occurs in the parotid gland but occasionally in the submandibular and minor salivary glands (120). MEC is characterized by grades based on histological appearance (low, intermediate, and high grade). The drivers of survival outcomes vary by study and include histologic grade, tumor location, tumor stage, nodal status, patient age, margin status, and perineural invasion (98, 118, 119, 121, 122).

There has been an increasing interest in the molecular landscape of MEC, and recent studies have identified alterations in EGFR signaling and rare *TP53* mutations (71, 123-126); however, the molecular landscape of this disease is mostly unknown. To date, the only well-described recurrent genetic alteration in MEC leads to the disruption of *MAML2* through a rearrangement between chromosomes 11 and 19 ((11;19)(q14-21; p12-13)), which results in the translocation of *CRTC1* and *MAML2* (70, 98, 122, 126-128). Initially discovered in 2003 by Tonon *et al*, *CRTC1* to *MAML2* rearrangement was found in 34-88% of MEC cases (70, 124, 126, 129), with a similar *CRTC3* to *MAML2* rearrangement occurring in a smaller subset of MEC patients (130). The resultant *CRTC1-MAML2* fusion gene is thought to play a role in tumorigenesis, and several *in vitro* studies have demonstrated that regulating fusion expression also regulates oncogenic phenotypes, possibly through a novel MYC-and/or CREB-mediated pathway (73, 74, 128, 131, 132), and potentially through an IGF1-dependent mechanism (72). Importantly, while the initial study suggested that *CRTC1-MAML2* may drive NOTCH pathway reporter activation in transfected cell lines (127), more recent studies using expression analysis of tumors have suggested that *CTRC1-MAML2* does not drive an increase in several common NOTCH signaling effectors (73, 74, 132).

While this increasing evidence for the prevalence and function of the *CTRC1-MAML2* translocation continues to advance in our field, little is known regarding the genetic mechanism that leads to the formation of this translocation. Furthermore, the importance of expanding the genomic understanding of salivary gland malignancies, especially *CRTC1/3-MAML2* negative tumors, continues to be highlighted in the current literature (133). Given these clear gaps in

knowledge in the field, we sought to use whole genome sequencing approaches to understand the genetic mechanisms that drive fusion formation in MEC and to characterize additional genetic events that contribute to the malignant process.

3.3 **Materials and Methods**

3.3.1 ***Ethics Statement***

This study was approved by two independent ethics boards. The University of Michigan's Institutional Ethical Review Board approved HUM00080561 for genetic analysis of retrospective MEC tissues and cell lines. The study was conducted in accordance with the guidelines of the Declaration of Helsinki and was approved by the Institutional Ethical Review of the University of Medical Sciences 505A/15 (06.05.2015).

3.3.2 ***Cell Lines and Reagents***

Mucoepidermoid carcinoma NCI-H292 cells were purchased from the American Type Culture Collection (ATCC) and genotyped throughout the study, as previously described, to confirm their authenticity (134). The new MEC cell line, IHG-MUC360, was derived and characterized by the following authors: V. MP, M. R., J. J., M. K., M. W., M. G., and M. J. (Patent No. P.414457, granted by the Polish Patent Office, described in the European Patent Register No. PL229507B1), and deposited in the German Collection of Microorganisms and Cell Cultures GmbH (DSMZ). The sinonasal undifferentiated carcinoma cell line MDA8788-6 was generously provided by Anderson. This cell line was generated as previously described by Takahashi *et al* (135). The oral cavity cell lines (IHG-MUC360, UMSCC-47, and UMSCC-104) used in this

study were selected from our head and neck repository (56). All cell lines were grown exponentially in DMEM (UMSCC-47, UMSCC-104, MDA8788-6) or RPMI (NCI-H292) containing 10% FBS, 7µg/mL penicillin/streptomycin and 1% non-essential amino acids in a 5% CO_2 incubator. PDX models were obtained from the NCI Patient-Derived Models Repository (PDMR) (NCI-Frederick, Frederick National Laboratory for Cancer Research, Frederick, MD, USA).

3.3.3 ***Linked Read Sequencing and Analysis***

High-molecular-weight (HMW) DNA was isolated from the NCI-H292 cell line and frozen tumors from the three NCI-PDX models. Briefly, cells were placed in lysis buffer (10 mM Tris-HCl, 400 mM NaCl, and 2 mM EDTA), 10% SDS, and proteinase K solution (1 mg/mL proteinase K, 1% SDS, and 2 mM EDTA) overnight at 37 °C. Following overnight lysis, DNA was extracted from the solution with 5 M NaCl for 1 h at 4 °C and precipitated with ice-cold ethanol for 5 hours at -20°C. High-molecular-weight DNA was eluted in TE buffer and the quality and integrity of the DNA were assessed using the Tapestation Genomic DNA ScreenTape kit (Agilent). The DNA was submitted to the University of Michigan Advanced Genomics core for 10x based linked read library generation and sequencing on an Illumina NovaSeq6000 with a 300nt paired end run. The samples were de-multiplexed and FastQ files with matched index files were generated using Long Ranger Version 2.2.2. Data were visualized using the Loupe software package, Version2.1.1 (2.4). Genome rearrangement plots were constructed using Circa Software (OMGenomics). The FastQ files for linked read sequencing were deposited in the sequencing read archives under identification number SUB8292046.

3.3.4 *Nanopore Sequencing and Analysis*

High molecular weight (HMW) DNA was isolated using the Nanobind CBB Big DNA Kit (Circulomics Inc.), according to the manufacturer's instructions, and 5ug was used a library input according to the Oxford Nanopore standard protocol at the University of Michigan Advanced Genomics Core. Raw FastQ data for nanopore sequencing were archived as described above. Alignment was performed using pipeline-structural variation (v.2.0.2), an alignment pipeline developed by Oxford Nanopore Technologies. Briefly, a QC report was generated using NanoPlot (v.1.19.0) (136) and the reads were aligned to the reference genome GRCh37/hg19 using LRA v1.3.2 (137). After alignment, we filtered the reads to include only those regions that overlapped with fusions identified by linked read sequencing and also had a split read (SA) tag. We then annotated the breakpoint position of each read using BLAT alignment to GRCh37/hg19 (138). We performed structural variant calling using two methods. First, we aligned to the GRCh37/hg19 reference genome using Minimap2 v2.17 (139). Next, the aligned reads were sorted and indexed using SAMtools v1.5 (140). Structural variant calling was then performed on aligned long reads using SVIM v2.0.0 and NanoVar v1.3.9 (141, 142). Both methods identify multiple types of structural variants including insertions, deletions, inversions, duplications, and translocations. These structural variants were then filtered to include only translocation break ends. To identify break ends called by both methods, we used the bedtools2 v2.29.2 intersect function (143). Break end pairs with a gap of less than 100 base pairs were reproducibly called using both methods. Nanopore sequencing data were deposited in the sequencing read archives under the identifier SRR17418364.

For the novel MEC model, the exact structure of translocation t(5;7)(p15;q32) was identified using informatics techniques, as previously described (144). Briefly, the library was prepared using the Mate Pair Library Preparation Kit (Illumina, San Diego, CA, USA) according to the manufacturer's instructions and pair-end sequenced (2 × 100 bp) on an Illumina HiSeq 1500. The structure of the translocation was verified by Sanger sequencing.

3.3.5 ***Fluorescence in situ hybridization (FISH) with TERT Break Apart Probe***

NCI-H292 cells were treated with 0.4ng/mL colcemid in culture media for four hours and then harvested with 0.025% trypsin. The cells were pelleted and subsequently treated with 5mL of hypotonic saline (1:1 0.075M KCL:0.9% sodium citrate) for seven minutes. Cells were again pelleted, placed in cold fixative (Carnoy's 3:1 methanol: acetic acid) for one hour on ice, and ten-minute incubation at room temperature. The cells were placed in a fresh fixative and stored at 4C overnight. Slides were prepared according to the Empire Genomics protocol. Briefly, cells were pelleted, and supernatant was removed leaving ~0.5mL above the pellet and then re-suspended to create a slightly milky solution. Slides were briefly placed in fresh cold fixative, and excess fixative was removed by tapping onto a paper towel. The slide was held slightly vertical using a glass Pasteur pipette, and the cells were dropped onto the slide and gently rotated to distribute the sample. The slide was dried at room temperature until a grainy surface appeared and then placed into a preheated slide warmer at 45C for final drying. Slides were stored at -20C for FISH. FISH was performed using an Empire Genomics *TERT* break apart probe (TERTBA-20-GROR, Empire Genomics) following the manufacturer's protocol, and mounted with DAPI (Prolong Diamond Antifade Mountant plus DAPI, Thermo Fisher Scientific, Waltham, MA, USA).

3.3.6 ***Digital droplet PCR***

Commercial digital droplet PCR (ddPCR) copy number assays were purchased from Bio-Rad Laboratories. Assay dHsaCNS779291124 was selected to measure the TERT target (*TERT*) and the assay dHsaCNS626260126 was selected to measure the control gene (*LARP1*). Each ddPCR contained 10 µL of ddPCR Supermix for Probes (No dUTP), 1 µL of 20x ddPCR assay, 7 µL of water, and 2 µL of either NCI-H292 DNA or control human genomic DNA (G3041, Promega). Droplet generation was performed using a QX200 Droplet Digital PCR System (Bio-Rad, Hercules, CA, USA). PCR was run using the following thermocycler conditions: enzyme activation for 10 min at 95°C with a 2°C/second ramp rate, 40 cycles of denaturation at 94°C for 30 s and annealing/extension at 60°C for 60 s with a 2°C/second ramp rate, enzyme deactivation for 10 min at 98°C with a 2°C/second ramp rate, followed by a hold at 4°C. Droplets were read using the QuantaSoft Software (Bio-Rad, Hercules, CA, USA).

3.3.7 ***Exome Sequencing and Variant Calling***

FFPE-derived genomic DNA was isolated from areas of high MEC tumor content identified by pathologist J.B.M. and prepared for exome sequencing using quality control metrics, as previously described by our team (145). The DNA was then prepared for sequencing with the Illumina TruSeq exome library preparation and capture kit and sequenced using 150nt PE reads as previously described (146). The quality of sequencing reads was determined using FastQC v.0.11.8. Trim galore v0.6.5, was used to remove sequencing adapters, as described previously (57). These processed reads were aligned to the hg19 reference genome using BWA v0.7.15

(147). PicardTools v2.23.2 was used to sort, deduplicate, and index the mapped reads, as previously described (148). Subsequently, base quality scores were recalibrated using GATK v4.1.0.0 (149) and the reads were finalized for variant calling. SAMtools v1.9 (140) was used to create pileup files for each tumor-normal pair. Next, we applied the somatic mode of VarScan v2.4.1, to call variants from these mpileup files, and Golden Helix VarSeq v2.2.0 (Bozeman, MT) was used to annotate these variants. Variants belonging to intronic and intergenic regions were filtered out. Furthermore, variants with greater than 5 reads supporting the alternate allele in tumor samples were considered true positives.

3.3.8 *Variant Effect Scoring*

The Cancer-Related Analysis of Variants Toolkit v5.2.4 (150) was used to perform variant impact scoring. This web-based informatics toolkit offers two analysis programs for variant interpretation: VEST and CHASM. We used VEST v4.0 to assign pathogenicity scores to the variants and CHASM v3.1 was used for somatic missense variant scoring, respectively.

3.3.9 *Copy Number Analysis*

Copy number estimation was performed using Aberration Detection in Tumor Exome (ADTEx) v.2.0 (151). The software utilizes previously processed tumor-normal BAM files to make copy number calls. Each call was then assigned a state ranging from zero to four based on the estimated copy number. State 0 corresponds to a homozygous deletion and state 1 corresponds to a heterozygous deletion. State 2 represents a normal copy number, whereas states 3 and 4 denote gain and amplification, respectively. Furthermore, representative Manhattan plots were generated

for each chromosome of each tumor/normal pair. Using copy number data, we assessed the global view of copy number variation in Circos plots with Circos v.0.69-9 (152).

Cell lines genome-wide copy number analysis was performed using the Affymetrix Oncoscan Copy Number Plus array (Affymetrix, Santa Clara, CA, USA), and analyzed using ChAS (v4.3, Thermo Fisher Scientific).

3.3.10 ***Neoantigen Prediction***

High-confidence non-synonymous coding somatic mutations were used to identify putative neoantigen peptides for each tumor. Each peptide sequence containing a coding mutation was accessed from the CustomProDB database using R package CustomProDB (v1.18.0)(153). The peptide sequence of the putative neoantigen is a 21 amino acid fragment of the mutated protein, with a nonsynonymous mutation at position 11. Putative neoantigen sequences of 11 amino acids were generated from these peptide sequences. HLA allelotypes were called at 4-digit resolution from WES data using OptiType (v1.3.3)(154). The binding affinity between each putative neoantigen and patient-specific HLA allele was predicted using NetMHCpan (v4.0)(155). Neoantigens with a binding affinity of >500nM were considered to strongly bind to MHC class I molecules. The number of strongly binding neoantigens was used to calculate neoantigen load.

3.3.11 *Mutational Signature Profiling*

The mutation characteristics were summarized using the R package MutationalPatterns (v2.0.0)(156). All high-confidence somatic SNVs for each sample were summarized based on mutation type and trinucleotide mutation context. The genetic signatures were downloaded from

the COSMIC database (v2). The pairwise cosine similarity score between each tumor somatic mutation profile and each COSMIC signature was calculated. The relative contribution of each COSMIC signature to each tumor somatic mutation profile was calculated using non-negative least-squares constraint optimization.

3.3.12 Microsatellite Instability (MSI) Detection

We used the MSIsensor algorithm (v0.6) to annotate the percentage of somatic MSI loci from tumor-normal sample pairs (157). This software package assigns microsatellite instability status based on a threshold of more or less than 3.5% of microsatellites with alterations. We present this score as well as the overall percentage of microsatellite alterations for each tumor-normal pair.

3.3.13 ***Sanger Sequencing of rearrangements and TERT promotor status***

PCR was performed on the isolated source DNA from the cell lines or PDX model using the AccuPrime GC-rich kit (Thermo Scientific), according to the manufacturer's instructions. Primers used in the analysis were those described for the TERT promoter (158). Amplicons were confirmed on a 2% agarose gel, extracted using the Qiagen gel extraction kit, and subjected to Sanger sequencing at the University of Michigan's Advanced Genomics core. Sanger trace files were then analyzed using DNAstar (Laser Gene Suite, version 17) against the hg38 reference genome. For exome data, IGV/2.8.10 was used to check the TERT promoter status of MEC tumors. *Igvtools count* was used to obtain the coverage track information (occurrences of A, T,

G, C, N, and INDELS) for each sample at the chr5:1,295,228 (hg19) and chr5:1,295,250 (hg19) positions.

3.3.14 Fusion Calling by RNA sequencing

Total cell line RNA underwent library preparation and sequencing by the University of Michigan Advanced Genomics Core. 500ng of RNA was used for library preparations with the Illumina Stranded mRNA library prep kit, according to the manufacturer's recommendations. Samples were paired-end sequenced to 150nt on an Illumina NovaSeq 6000.

RNA sequences were trimmed using TrimGalore (v0.6.5), then aligned to a GRCh37 reference genome, including multimapping chimeras, with STAR (v3.7.3). Fusions were then detected from aligned reads using Arriba, using inbuilt GRCh37 reference information (v.2.4.0) (159). Aligned reads were then sorted and indexed with SAMtools (v.1.13), and each fusion was visualized using Arriba's draw_fusions.R script.

3.3.15 ***Telomere Length Analysis***

Genomic DNA was harvested from the cell lines using the Wizard Genomic DNA Purification Kit (Promega, Madison, WI, USA). Briefly, the cell pellets were lysed with 600µL of Nuclei Lysis Solution (Promega catalog # A7941) and incubated at 55C for one hour. Subsequently, 200µL µL protein precipitation solution (Promega, catalog # A795A) was added, mixed thoroughly by vortexing, and placed on ice for five minutes. The samples were then centrifuged

at 13,000 RPM for 3 minutes (at 4C). The supernatant was then removed and added to a new Eppendorf tube containing 600µL of fresh isopropanol and incubated at room temperature for 10 min to allow the DNA to spiral out of the solution. The DNA was then pelleted, washed with 70% ethanol, and resuspended in 35-70µL of nuclease free H20. Samples were incubated overnight at 4C to allow the DNA to completely dissolve into the solution. Quantification was performed using a Nanodrop ND-1000. Telomere length was assessed using the absolute human telomere length quantification qPCR assay kit (ScienCell Catalog #8918) according to the manufacturer's protocol.

3.3.16 ***Clonogenic Cell Survival***

UM-SCC-104 and NCI-H292 cells were transduced with lentivirus prepared using the pLKO.1 backbone with scrambled control or *TERT* targeting shRNA constructs, respectively, purchased from Sigma Aldrich. Following transduction, cells were treated with 2ug/mL puromycin for 48hours and surviving cells were seeded at densities of 250, 500, and 1000 cells in 6-well plates in triplicate to assess clonogenic cell survival. Excess cells from each condition were immediately lysed and used for western blotting analysis. After 10 days, with colonies of approximately 50 cells, we fixed the plates and stained them with crystal violet for quantification and imaging.

3.3.17 ***Western Blot Analysis***

Protein lysates (RIPA buffer) were quantified as previously described (160) and used as inputs for western blot analysis. The primary antibodies used were TERT (1:1000) and HSP90 (1:1000)

(Cell Signaling Technology) and incubated overnight in 5% goat serum in TBS-Tween 20 at 4°C. Membranes were washed with TBS-T and incubated with the appropriate HRP-linked secondary antibodies (Thermo Fisher) for 1 hour. Blots were activated using enhanced chemiluminescence (ECL) and imaged.

3.3.18 Statistical Analysis

Telomere length was compared using Student's t-test with a p-value <0.05, which was considered statistically significant. Statistical analyses were performed using Prism v8 (GraphPad Software, San Diego, CA, USA).

3.4 **Results**

3.4.1 ***Nanopore Long Read Sequencing and Linked Read Sequencing Identifies Novel Rearrangements:***

We first sought to characterize the sequence of the *CRTC1-MAML2* breakpoint by Nanopore long read sequencing on high molecular weight DNA from MEC-derived cell lines NCI-H292, UM-HMC-1, UM-HMC-3A, and UM-HMC-3B (**Figure 3-1**). These generated 617,700-4,130,516 reads with an average read length of 6,122.4-8,705.1 nucleotides (**Table 3-1 and Figure 3-2**). For each cell line sequenced, we identified multiple reads that directly span the *CRTC1* and *MAML2* fusion (**Figure 3-1A**), and determined the location and sequence of the *CRTC1-MAML2* breakpoint for each cell line (**Figure 3-1B-D**). As expected, the translocation breakpoint varied for each cell line, though it occurred in the first intron of *CRTC1* and *MAML2* for each cell line tested. As UM-HMC-3A is derived from a local recurrence, and UM-HMC-3B

is derived from a recurrent metastasis obtained by surgical resection from the same patient, both cell lines contained the same *CRTC1-MAML2* breakpoint sequence (161).

Next, we sought to detect novel rearrangements in these cell lines. We first performed NCI-H292 *de novo* genomic rearrangement discovery using the SVIM and NanoVar pipelines, which identified 392 and 592 potential rearrangements, respectively (**Figure 3-3A, Table 3-2 and 3-3**). Surprisingly, only eight of the rearrangements were common between the two pipelines (**Figure 3-3B**), and of our validated rearrangements with confirmatory Sanger sequencing shown above, only one was identified by the SVIM pipeline. As the NanoVar pipeline is optimized for low-depth structural variant calling, we then performed NanoVar genomic rearrangement calling on UM-HMC-1, 3A, and 3B. This identified 1,241, 1,118, and 2,090 potential breakpoints respectively; 858, 370, and 1,523 of which are associated with previously identified cancer driver mutations or cancer-associated gene fusions (**Figure 3-3C**). We identified thousands of mutations, many of which include genes that contain oncogenic driver mutations in other cancer types. Therefore, either improving calling accuracy with higher-depth long-read sequencing or developing a high-throughput orthogonal validation strategy are necessary to identify novel breakpoint mutations that affect MEC tumor phenotypes.

Due to the difficulty reproducibly identifying *de novo* rearrangements with Nanopore sequencing, we performed linked read genome-wide sequencing of the *CRTC1-MAML2* fusion-positive NCI-H292 cell line. Following high molecular weight DNA isolation and 10x-based linked read barcoding, we generated 913,409,176 sequencing reads from the NCI-H292 library on a NovaSEQ6000, of which 95.90% mapped the reference genome, leading to a mean genome

depth of 39.8X (**Table 3-4**). From the library, 93.4% of the sequenced molecules were greater than 20kb in length and 18.8% were greater than 100kb in length. From this dataset, 99.5% of SNPs were phased across the genome, with phase blocks up to 39,295,952 bp in length, supporting a high-quality linked read genome dataset. In an in-depth analysis of the genome, 139 large structural variations were identified. The majority of these alterations were either deletions or duplications ranging in length from 20kb to 1-2Mb (**Table 3-4**), from which we identified nine high-confidence chromosomal rearrangements, including *CRCT1-MAML2* translocation (**Figure 3-4A, B; Table 3-5**). While the linked read data identified the *CRTC1-MAML2* rearrangement, we did not identify any reads that contained the exact *CRTC1-MAML2* translocation junction, potentially due to the high GC content and low read representation of *CRTC1* intron 1. Moreover, we also successfully identified translocation events associated with *CRTC1-MAML2* formation.

In NCI-H292, the *CRTC1-MAML2* translocation occurred through a series of five genomic events consisting of four linked translocations and one genomic deletion. Unique reads were used to map the breakpoints to Chr19:18,794,730 in *CRTC1* and Chr11:95,834,323 in *MAML2* (genome build hg19). Of these genomic events, *CRTC1* Chr19:18,795,369 was translocated adjacent to *SGK3* Chr8:67,777,508, and the 3' end of the breakpoint at Chr8:67,777,554 was translocated to Chr6:75,131,786 in *RP11554D151*. Furthermore, the 5' end of *MAML2* was translocated from Chr11:95,834,411 to Chr8:67,527,412, which resides between *VCPIP1* and *MYBL1*, and was also associated with a deletion between Chr8:46,880,000 and Chr8:67,530,000, which contains the *MYBL1* gene (**Figures 3-4A and 3-4C**). Nanopore sequencing further confirmed the breakpoint structures between *MYBL1* and *MAML2*, which contained three

nucleotide microhomology at the breakpoint, as well as *SGK3* and *CRTC1*, which had five nucleotide microhomology at the breakpoint (**Figure 3-4D**).

Sanger sequencing was then used to confirm the novel translocation junctions between *MYBL1* and *MAML2* as well as *SGK3* and *CRTC1* (**Figure 3-5A and B, Table 3-6**). Sanger sequencing also validates *CRTC1* and *MAML2*, joined by a 3 nucleotide microhomology (**Figure 3-5C**). In addition, we validate the *CRTC1-MAML2* translocation in UM-HMC-1 (**Figure 3-5D**).

We performed fusion detection on RNA sequencing data from NCI-H292, UM-HMC-1, UM-HMC-3A, and UM-HMC-3B (**Table 3-7**). We validate the CRTC1-MAML2 fusion in each cell line, with a read depth of 21, 6, 6, and 5 reads directly spanning the junction, respectively. While the DNA sequence for the *CRTC1-MAML2* translocation differs between cell lines, the sequence of the resulting RNA fusion site remains consistent (**Figure 3-6**).

Collectively, these data support chromothripsis as a mechanism of rearrangement (162) mediated by a microhomology-driven repair process that leads to the formation of canonical *CRTC1-MAML2* rearrangement in this disease.

Subsequent analysis of additional chromosomal rearrangements in NCI-H292 linked read data revealed a second rearrangement containing over eight linked events, including a promoter rearrangement upstream of *TERT* (**Figure 3-7A, B**). We did not identify supporting nanopore reads for this translocation; therefore, Sanger sequencing was used as an additional approach to validate rearrangement junctions, and the *PPP2R1B-TERT* junction was found to occur with no

additional insertion nucleotides at the break point and no obvious homology at the junction, suggesting that the rearrangement may have occurred through a different end-joining-based mechanism (**Figure 3-7C**).

The *PPP2R1B-TERT* rearrangement moves the promoter and 5' region of *PPP2R1B* from Chr11:111,600,986 to Chr5:1,296,648, which sits immediately prior to the *TERT* gene. This rearrangement is associated with several additional structural variations in Chr11, including multiple deletions and duplications in *UBASH3B, ARHGAP32,* and the *BARX2* promoter. Finally, this chained structural variation also included a rearrangement translocation from *ARHGAP3* Chr11:128,920,000 to a region downstream of *SHFM1* at Chr7:134,940,000. The coordinates of the associated rearrangements and a schematic representation are shown in **Figure 3-7C and 3-7E**.

A second and newly derived *CRTC1-MAML2* fusion-positive MEC model, IHG-MUC360, was also sequenced to narrow the breakpoints of translocation t(5;7)(p15;q34), revealed by cytogenetic analysis, potentially involving the *TERT* gene. For this sample, 32,567,275 read pairs were generated and mapped to the reference genome (hg19). Based on 16 discordant reads (the first read of the pair maps to chromosome 5 and the second maps to chromosome 7), we identified the genomic localization of the translocation, which is presented according to the current nomenclature (163). For der(5): t(5;7)(p15;q32)dn.seq[GRCh37/hg19]t(5;7)(7qter→7q32(138,444,233)::5p15(1,302,991)→5qter).

For der(7):

t(5;7)(p15;q32)dn.seq[GRCh37/hg19]t(5;7)(7pter→7q32(138,444,235)::5p15(1,302,889)
→5pter). The breakpoint on chromosome 5 is located in non-coding DNA approximately 8 kb upstream of the *TERT* gene, whereas the breakpoint on chromosome 7 disrupts *ATP6V0A4* gene (MIM: 605239) in intron 8. (**Figure 3-7D, Table 3-7**) These data show that translocations involving *TERT* are a novel recurrent potential mechanism of tumorigenesis in MEC.

Given the potential functional importance of this translocation, we next sought to confirm *TERT* promoter rearrangement using a break-apart probe targeting the 5' and 3' regions adjacent to *TERT*. We then quantified the *TERT* copy number of each cell, counting the 3'-binding probe, 5'- binding probe, and both overlapping. This revealed a high degree of heterogeneity between cells from the same cell line, though the prevalence of each genotype varied between cell lines (summarized in **Figure 3-8A**). In NCI-H292, 77% of all cells counted demonstrated the presence of a copy gain of the 3'-binding probe relative to the 5'- binding probe, consistent with the presence of genomic rearrangement (**Figure 3-8A**). Cells with a *TERT* copy number greater than 2 were highly enriched in UM-HMC-1 and UM-HMC-3B, but not UM-HMC-3A (**Figure 3-8A**). This finding was validated by copy number alterations detected by Oncoscan copy number array, which called copy number alterations throughout the genome of each cell line (**Figure 3-8B**). Using this method, UM-HMC-1 and UM-HMC-3B had an average of 3 copies of the *TERT* gene per cell, while UM-HMC-3A contained 2 copies. We further confirmed the presence of *TERT* copy gain by digital droplet PCR in NCI-H292 cells relative to the normal human reference genomic DNA (**Figure 3-9**). Taken together, these copy number analyses support *TERT*

amplification in 3 of 4 cell lines tested, while demonstrating a translocation of the 3' region characteristic of NCI-H292.

Given the high number of chromosomal rearrangements observed in the NCI-H292 cell lines, we sought to analyze a third mucoepidermoid model, NCI-PDX-062 (with unknown CRTC1/3-MAML2 translocation status) and obtained from the NCI PDX biorepository. To understand the relative rate of chromosomal rearrangement in the PDX setting, we also sequenced two HNSCC PDX models from the biorepository, NCI-PDX-296 and NCI-PDX-356. From the three genome-wide linked read libraries, we generated over 3 billion sequencing reads from the library on a NovaSEQ6000, of which 71% mapped the reference genome (consistent with other studies of tumors grown in mice), leading to a mean human genome depth of 28.1X, 39.5X and 28X, respectively (164) (**Table 3-4**). This read depth led to high SNP phasing across genomes (**Figure 3-10A**). This analysis identified 11 different chromosomal rearrangements in the NCI-PDX-062 model, only one rearrangement in HNSCC PDX-356, and no rearrangements in HNSCC PDX-294 genomes (**Figure 3-10B, Table 3-8**). The 11 translocations identified in the MEC PDX-062 genome included a rearrangement that translocates the *AGO3* 3'-UTR and an independent rearrangement that translocates the *ZNF808* promoter (**Figure 3-10C, D and E**). However, we did not detect any *MAML2*, *TERT* or *MYBL1* rearrangements in this model. Therefore, these data suggest that this model may better represent *CRTC1/3-MAML2* fusion-negative MECs, given the high rate of translocations observed in the sample, suggesting that pathogenesis of *CRTC1/3-MAML2* fusion-negative MECs may occur through genomic rearrangement mechanisms to alternate driver/suppressor genes.

3.4.2 *Exome sequencing of MEC tumor samples identifies recurrent TERT alterations*

To further explore the molecular landscape of mucoepidermoid carcinoma and prevalence of *TERT* alterations in the cell lines, we assembled a cohort of tumor samples from patients treated at the University of Michigan (**Figure 3-11A**). Eleven samples underwent whole-exome sequencing, with an average of 124,214,959 uniquely mapped reads (83.8%) (**Table 3-9**). Using these data, we first characterized the general genetic characteristics of MEC tumors. We identified a median mutational burden of 36, with only one tumor, MEC001, harboring an exceptionally high number of 1006 mutations (**Figure 3-13A**). Analysis of microsatellite instability demonstrated that MEC001 was unstable, consistent with the high level of single-nucleotide alterations observed in this tumor (**Table 3-10**). Next, we used the data to determine the HLA type in each tumor (**Table 3-11**) and calculated the neoantigen load using the NetMHCpan pipeline. This analysis demonstrated a median predicted neoantigen load of 2.5 per tumor. Overall, 11 samples showed a relatively low total mutational burden across the genome, suggesting that the underlying mechanisms leading to enhanced genomic rearrangements may be active in this disease, given the large number of structural rearrangements detected in the linked read sequencing data.

To test this hypothesis, we performed molecular signature analysis using the Mutational Patterns pipeline (156), which identified a strong enrichment of COSMIC mutational signature 3 (**Figure 3-11B, C**), which is associated with homologous recombination double-strand break repair deficiencies in breast, ovarian, and pancreatic cancers (165, 166). The other highlighted signatures (5 and 12) had unknown significance at this time (165). Collectively, the exome sequencing data support the concept that MEC tumors may have an impaired ability to repair

DNA double-strand breaks, which is consistent with the enrichment in genomic rearrangements and structural variations observed in the linked read data.

We then analyzed specific high-confidence single nucleotide variants and indels, as depicted in **Figure 3-13A** (SNV and INDELs are summarized in **Tables 3-11 and 3-12**, respectively). Notable findings include recurrent alterations in *MUC12, MUC3A, MUC4,* and *MUC5B,* a gene family that is implicated in disease pathogenesis for tumors of the aerodigestive tract (167, 168) and has more recently been shown to have potential prognostic significance in MEC (169, 170). Recurrent *NOTCH2* mutational alterations were also found (4/11 tumors), including a Cys877Phe in MEC023, predicted to be functionally significant by the VEST functional annotation pipeline ($P < 0.1$, **Table 3-14**), a 5'-UTR 3 nucleotide deletion and a frameshift Pro6Argfs*27 in two different tumors (**Figure 3-13A**). Furthermore, we identified two additional tumors with a single copy deletion of the *NOTCH2* allele for 6/11 (54%) MECs with altered *NOTCH2* (**Table 3-15**). This *NOTCH* gene is known to play a role in tumorigenesis in adenoid cystic carcinoma but has not previously been described in MEC tumors (171, 172). Importantly, we also discovered a *TERT* His412Tyr missense mutation in one of the 11 tumors (MEC 1), which had a strong functional prediction score using functional annotation pipelines (**Table 3-14**). Given the prevalence of *TERT* promoter mutations in other cancers, we also evaluated *TERT* promoter status in our cohort; however, no evidence of *TERT* promoter mutations was noted in any of the tumors (data not shown). We then explored copy number alterations of the highlighted genes in each tumor and identified copy number changes in several of the recurrently mutated genes (**Figure 3-13B, Table 3-15**), as well as a recurrent focal amplification of *TERT* region in several tumors (**Figure 3-13C**). In total, 5/11 (45%) of the MEC tumors that we profiled

contained a *TERT* mutation or copy number amplification. As with the *TERT* amplifications, copy number alterations of some of the recurrently altered genes identified in this tumor cohort, including *MUC5B and MAML1* were also found in MEC-derived cell lines (**Figure 3-14**).

3.4.3 Functional validation of TERT in Mucoepidermoid cell line model

Given the unexpected recurrence of *TERT* alterations in MEC, we sought to determine the functional importance of *TERT* in NCI-H292 cells. Therefore, we first evaluated *TERT* promoter mutation status for an additional 19 head and neck cancer models (**Figure 3-15A**). The most common SNVs included G229A and G229T, found in 6/19 and 3/19 of the models, with the remaining models containing WT *TERT* promoter status. Telomere length was then obtained for NCI-H292 and compared to two HNSCC cell lines UMSCC-47, and UMSCC-104, as well as to a cell line from another rare head and neck cancer called sinonasal undifferentiated carcinoma (MDA8788-6). These results demonstrated that NCI-H292 cells displayed significantly longer telomeres than UMSCC-47, UMSCC-104, and MDA8788-6 cells, independent of passage number or patient age, at the time of cell line creation ($p=0.0005$) (**Figure 3-15B and Table 3-16**). Together, these data suggest a potential role for *TERT* in tumorigenesis in MEC and should be evaluated further.

To test the functional significance of TERT in MEC, we leveraged publicly available DepMap CRISPR screen data (173), in which genome-wide CRISPR knockout screens were performed to define essential genes for survival in a panel of 342 cancer cell lines. We evaluated the importance of *TERT* relative to *MAML2* in NCI-H292 cell survival (174). Previously published data showed that *CRTC1-MAML2* expression is essential for NCI-H292 cell growth in vitro (123,

175). We confirmed that *MAML2* gRNAs were identified in these depletion screens using data recently released by the consortium (CRISPR (Avana) Public 19Q3) (173). Importantly, *MAML2* was highly essential for NCI-H292 survival (CERESscore = -0.83), but not for the survival of most other cancer models, confirming a specific role for *MAML2* in mucoepidermoid cells (**Figure 3-15C**). Given that these data demonstrated that dropout CRISPR screening was an effective strategy to identify model-specific drivers, we then explored the sensitivity to knockouts of the additional genes with structural alterations identified above. While the majority of these genetically altered genes were not essential for NCI-H292 survival, the data demonstrated that *TERT* knockouts caused a significant effect on the survival of NCI-H292 cells (CERESscore = -0.62), which again was highly specific for the model, as the response was in the top 2% of all profiled models. To validate these results independently, we transduced NCI-H292 and UM-SCC-104 cells with shRNA constructs targeting TERT. Consistent with the DepMap data, knockdown of TERT caused a significant reduction in clonogenic cell survival in the NCI-H292 cell line, but not in the control cells (**Figure 3-15D-F**), confirming that *TERT* is an essential driver of the *TERT* rearrangement cell line model.

3.5 **Discussion**

Multiple studies have shown the importance of *CRTC1-MAML2* fusion in MEC tumors and delineated the mechanism by which this fusion gene promotes tumorigenesis (70, 126-128), and until now, the molecular mechanism driving fusion formation has remained unresolved. Here, we provide the first evidence for a chain of five linked structural events that result in the known (11;19)(q14-21; p12-13) or *CRTC1* to *MAML2* translocation. Our data demonstrate the involvement of *RP11554D151, VCPIP1* and *MYBL1* with the resultant rearrangement of *MYBL1*.

Because *MYBL1* translocation is known to play a role in the tumorigenesis of acinic cell carcinomas (126, 176), these data suggest that *MYBL1* may also play a previously undiscovered role in MEC and/or other salivary gland cancers. However, our exome cohort did not reveal additional *MYBL1* alterations, so we chose not to pursue additional studies of the gene in MEC at this time. In contrast, our data revealed recurrent alterations that were predicted to activate *TERT* as well as alterations predicted to disrupt *NOTCH2* in MEC.

In head and neck malignancies in general, disruption of NOTCH signaling appears to be an important mechanism of pathogenesis (177), and with the exception of some developmental studies suggesting that NOTCH signaling may play a role in the terminal fate decisions of salivary gland cells, the NOTCH receptors in mucoepidermoid carcinomas remain largely unknown (178, 179). Furthermore, as discussed previously, prior studies have shown that the *CRTC1-MAML2* fusion does not reproducibly alter NOTCH effector signaling. Notably, our data showed that the *NOTCH2* receptor was disrupted in both *CRTC1-MAML2* fusion-positive and-negative diseases, which supports the hypothesis NOTCH effector signaling acts independently from *MAML2* translocation in this cell type. Although we were limited by a relatively small sample size due to the scarcity of this disease, the high recurrence rate of *NOTCH2* alterations in our cohort suggests that the loss of *NOTCH2* is an important event in MEC progression, which may be critical for restricting cell fate decisions.

The role of *TERT* in mucoepidermoid carcinoma has not been previously reported, and evaluation of *TERT* and telomere length in salivary gland malignancies has been limited to a few small studies. However, there is some evidence for the role of *TERT* in adenoid cystic

carcinoma; a study by Morris *et al*. found that 14% (5/36) of adenoid cystic carcinoma samples harbored mutations in the *TERT* promoter region (180). In contrast, a study by Kim *et al*. evaluated 36 benign and malignant salivary gland tumors and found that none of the samples contained mutations in the *TERT* promoter region (181). Further investigation into telomere length in parotid pleomorphic and carcinoma ex-pleomorphic adenoma samples revealed an increase in relative telomere length in carcinoma ex-pleomorphic adenoma samples compared to pleomorphic adenomas or control samples. This was suggested to be secondary to alternative lengthening of telomeres (ALT) driven by aberrant methylation of p14(ARF) and p16(INK4A), rather than increased telomerase activity (182). Our data revealed two novel rearrangements involving the *TERT* promoter region, revealing a strong dependence of TERT activity in NCI-H292 cells, demonstrating increased telomere length compared to other head and neck cancer cell lines, and representing the first evidence for a possible role of *TERT* in MEC pathogenesis.

Finally, our data are the first to characterize full MEC genomes and support a high prevalence of structural rearrangements in this disease, especially compared to the HNSCC models that we concurrently profiled. In the future, larger cohort studies that leverage these genome sequencing approaches may lead to a better understanding of the relative number of rearrangement events in this disease as well as the genetic mechanisms by which these rearrangements are formed. This is supported by our exome sequencing data from tumors, which suggests that MEC tumors have a defect in homologous recombination-based DNA damage repair. Most importantly, not only does this data answer the longstanding question regarding the genetic mechanism of *CRTC1-MAML2* translocation, it also identifies *TERT* as a new driver of MEC. Given the ongoing efforts to build TERT-targeting therapeutic approaches for tumors with *TERT* aberrations, our discovery

may have a significant impact on our basic understanding of MEC molecular oncogenesis, as well as on the advancement of some of the first targeted therapy options for patients with this disease.

Acknowledgements

DNA sequencing of IHG-MUC360 was carried out using the CePT Medical University of Warsaw infrastructure (Innovative economy 2007–13, Agreement POIG.02.02.00-14-024/08-00). The VMP is a beneficiary of the START Program (2021) of the Foundation for Polish Science. Thank you for data curation to Molly Heft Neal, Apurva Bhangale, Nicole Michmerhuizen, Dana Al Majid, Clifford Chang, Erin Sandford, Jiayu Wang, Małgorzata Jarmuż-Szymczak, Malgorzata Rydzanicz, Joanna Janiszewska, Magdalena Kostrzewska-Poczekaj, Malgorzata Wierzbicka, and Maciej Giefing. Thank you for assistance with data analysis to Molly Heft Neal, Apurva Bhangale, Muneesh Tewari, Maciej Giefing, Jonathan McHugh, Małgorzata Jarmuż-Szymczak, Victor Murcia Pienkowski, Joanna Janiszewska, Magdalena Kostrzewska-Poczekaj, Matthew Spector, and Chad Brenner.

3.6 Figures

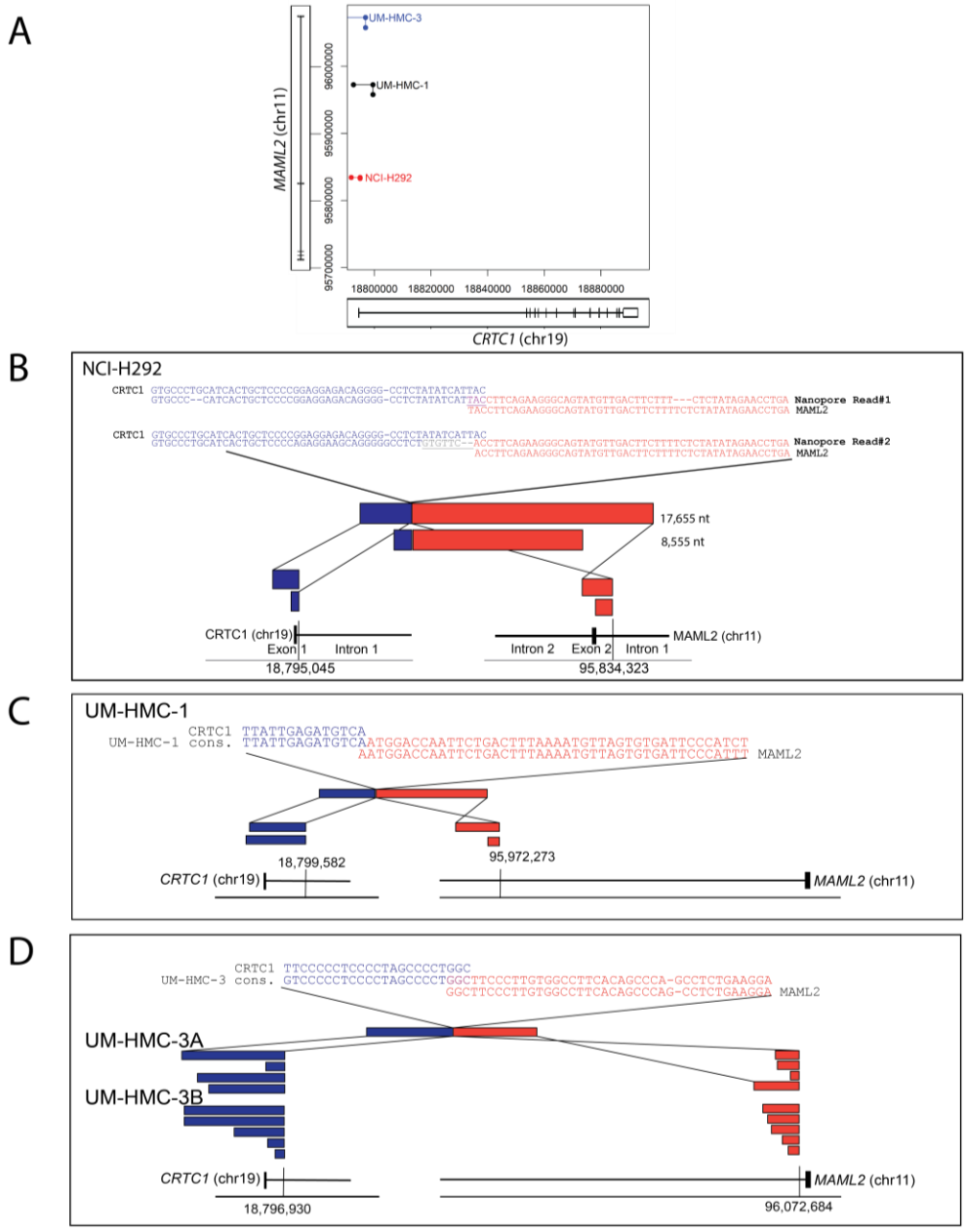


Figure 3-1. *CRTC1-MAML2* breakpoint is directly sequenced by long read Nanopore DNA sequencing.

A) Coordinates of reads split between both CRTC1 and MAML2 for each cell line sequenced. UM-HMC-3A and UM-HMC-3B share a genetic background, so are recorded as UM-HMC-3. **B)** Schematic represents independent reads that identify the CRTC1-MAML2 breakpoint in NCI-H292, **C)** UM-HMC-1, **D)** UM-HMC-3A, and UM-HMC-3B. Consensus sequences of each breakpoint are shown.

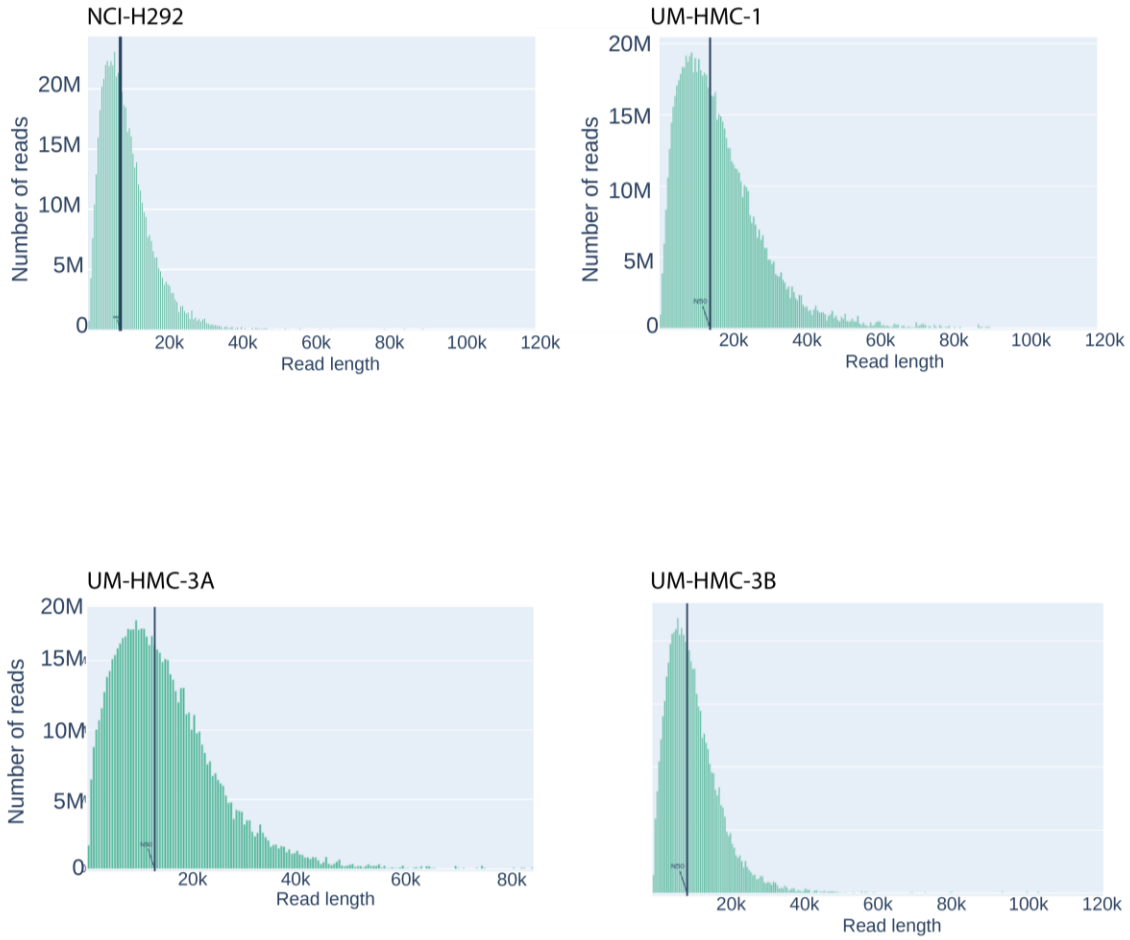
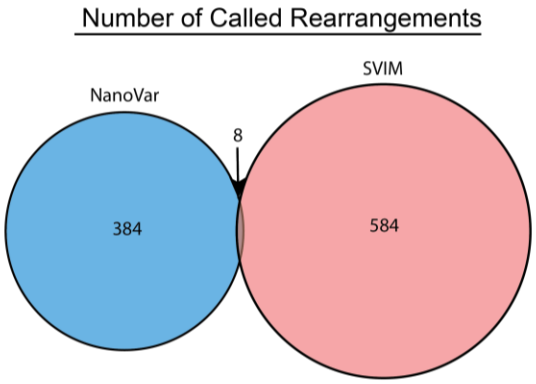


Figure 3-2. Nanopore Quality Control Data for Genome Sequencing Data.

Distribution of read lengths from each sequencing run is plotted for **A)** NCI-H292, **B)** UM-HMC-1, **C)** UM-HMC-3A, and **D)** UM-HMC-3B. Each line represents the NC50.

A



B

NanoVar								SVIM							
Breakpoint position 1				Breakpoint Position 2				Breakpoint position 1				Breakpoint Position 2			
chrom	position	Ensembl ID	Gene Nam	chrom	position	Ensembl ID	Gene Nam	chrom	position	Ensembl ID	Gene Name	chrom	position	Ensembl ID	Gene Name
chr15	20788149	-	-	chr15	23454162	-	-	chr15	20788149	-	-	chr15	23454165	-	-
chr15	21097491	-	-	chr15	22108615	-	-	chr15	21097491	-	-	chr15	22108618	-	-
chr15	22108615	-	-	chr15	21097491	-	-	chr15	22108618	-	-	chr15	22108618	-	-
chr15	23454162	-	-	chr15	20788149	-	-	chr15	23454165	-	-	chr15	23454165	-	-
chr8	43837102	-	-	chr8	46842640	-	-	chr8	43837075	-	-	chr8	46845222	-	-
chr8	46842640	-	-	chr8	43837102	-	-	chr8	46842667	-	-	chr8	46842667	-	-
chr9	82314040	ENSG0000	TLE4	chrM	2640	-	-	chr9	82314040	ENSG0000	TLE4	chrM	2539	-	-
chrX	61712928	-	-	chrX	58571792	-	-	chrX	58579352	-	-	chrX	61712921	-	-

C

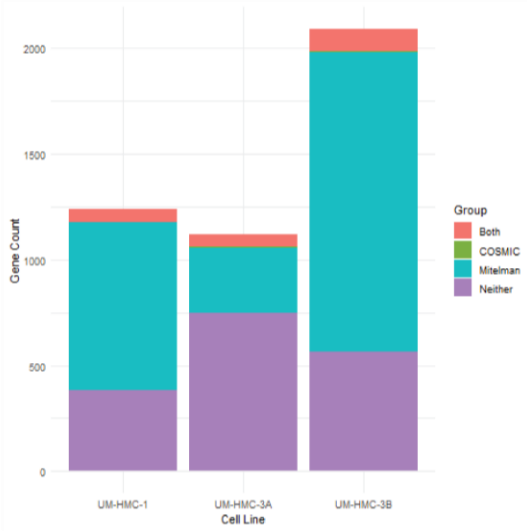
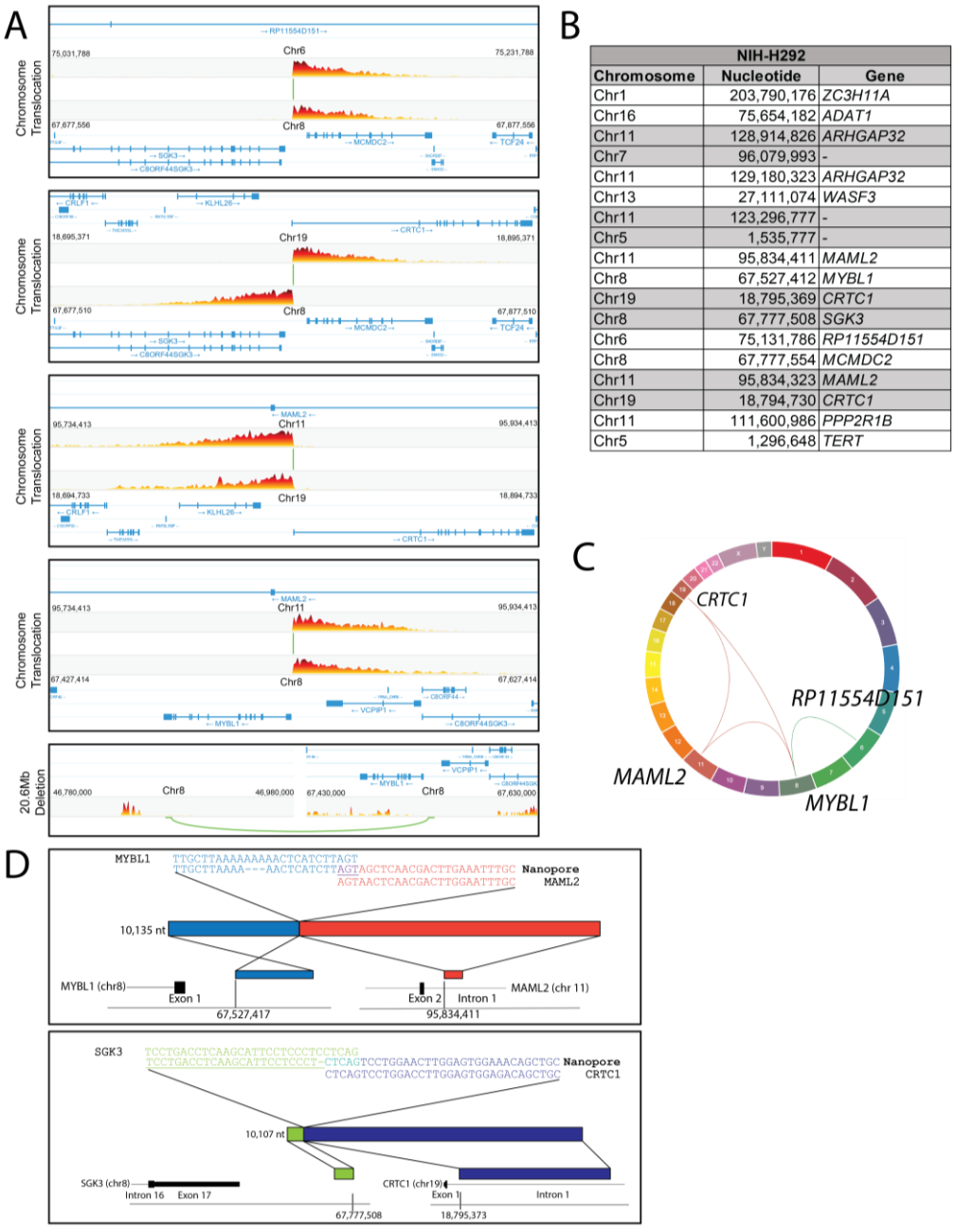


Figure 3-3. Analysis of Additional Genomic Rearrangements Based on Nanopore Data.

A) NanoVar and SVIM were used to call structural rearrangements, and overlapping alterations between the two outputs are shown. **B)** Common structural rearrangements called by both algorithms. **C)** Each gene containing a translocation in UM-HMC-1, 3A, and 3B, called by NanoVar. These genes are associated with common cancer driver mutations (COSMIC database, green), known cancer-associated fusion genes (Mitelman database, cyan), both types of genes (red), or no currently known association (purple).



NIH-H292		
Chromosome	**Nucleotide**	**Gene**
Chr1	203,790,176	*ZC3H11A*
Chr16	75,654,182	*ADAT1*
Chr11	128,914,826	*ARHGAP32*
Chr7	96,079,993	-
Chr11	129,180,323	*ARHGAP32*
Chr13	27,111,074	*WASF3*
Chr11	123,296,777	-
Chr5	1,535,777	-
Chr11	95,834,411	*MAML2*
Chr8	67,527,412	*MYBL1*
Chr19	18,795,369	*CRTC1*
Chr8	67,777,508	*SGK3*
Chr6	75,131,786	*RP11554D151*
Chr8	67,777,554	*MCMDC2*
Chr11	95,834,323	*MAML2*
Chr19	18,794,730	*CRTC1*
Chr11	111,600,986	*PPP2R1B*
Chr5	1,296,648	*TERT*

Figure 3-4. Linked Read Sequencing Resolves the Genetic Mechanism Driving *CRTC1-MAML2* Rearrangement in the NCI-H292 Mucoepidermoid Carcinoma Cell Line.

A) Five structural variations were discovered in NCI-H292, associated with the *CRTC1* to *MAML2* genomic rearrangement. Structures of individual events are shown. **B)** Schematic representation shows independent reads used to identify the *MYBL1* to *MAML2* translocation, and the *SGK3* to *CRTC1* translocation. GRCh37/hg19 coordinates are shown, with sequences spanning the breakpoint annotated at the top of the figure. **C)** High confidence chromosomal rearrangements and estimated genomic breakpoints discovered in NCI-H292. **D)** Genome-wide view of the chained structural events leading to the *CRTC1* to *MAML2* translocation.

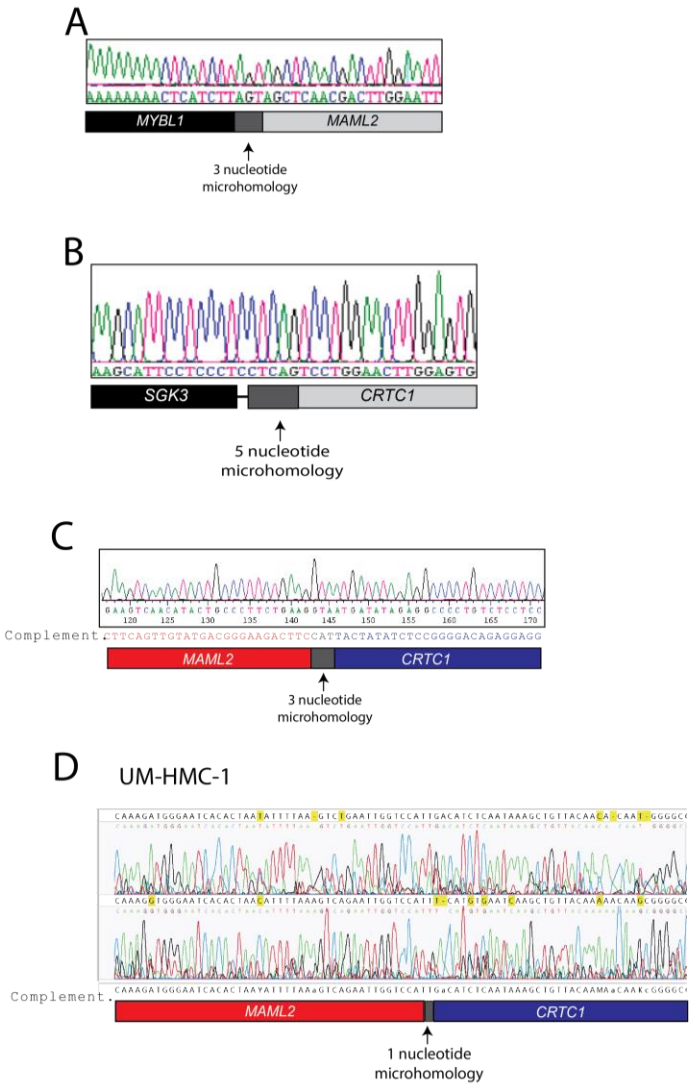


Figure 3-5. Validation of Select Genomic Rearrangement Breakpoints in MEC cell lines.

A-C) Sanger sequencing traces of PCR amplicons from genomic DNA isolated from NCI-H292. **D)** Sanger sequencing traces of PCR amplicons derived from UM-HMC-1.

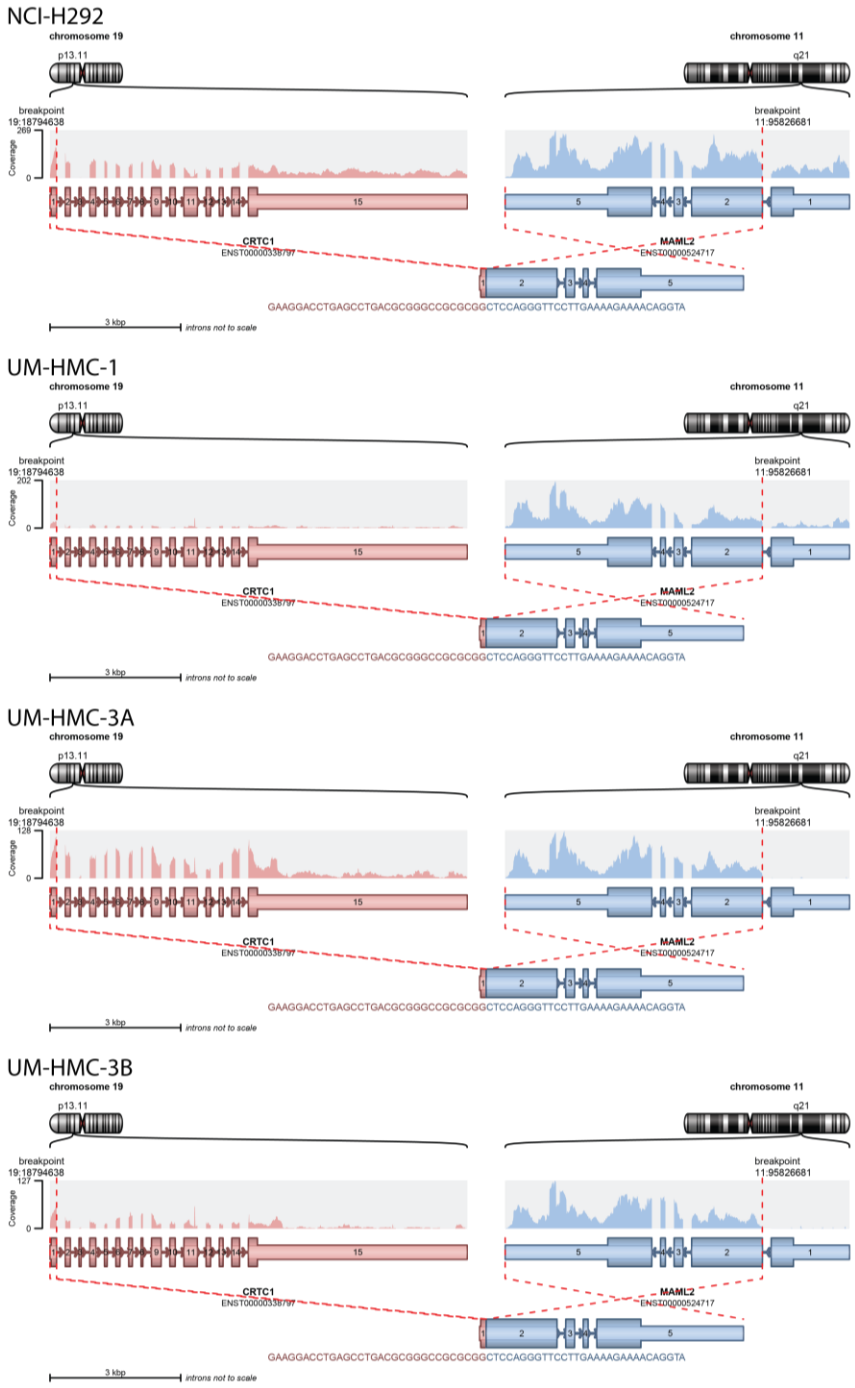
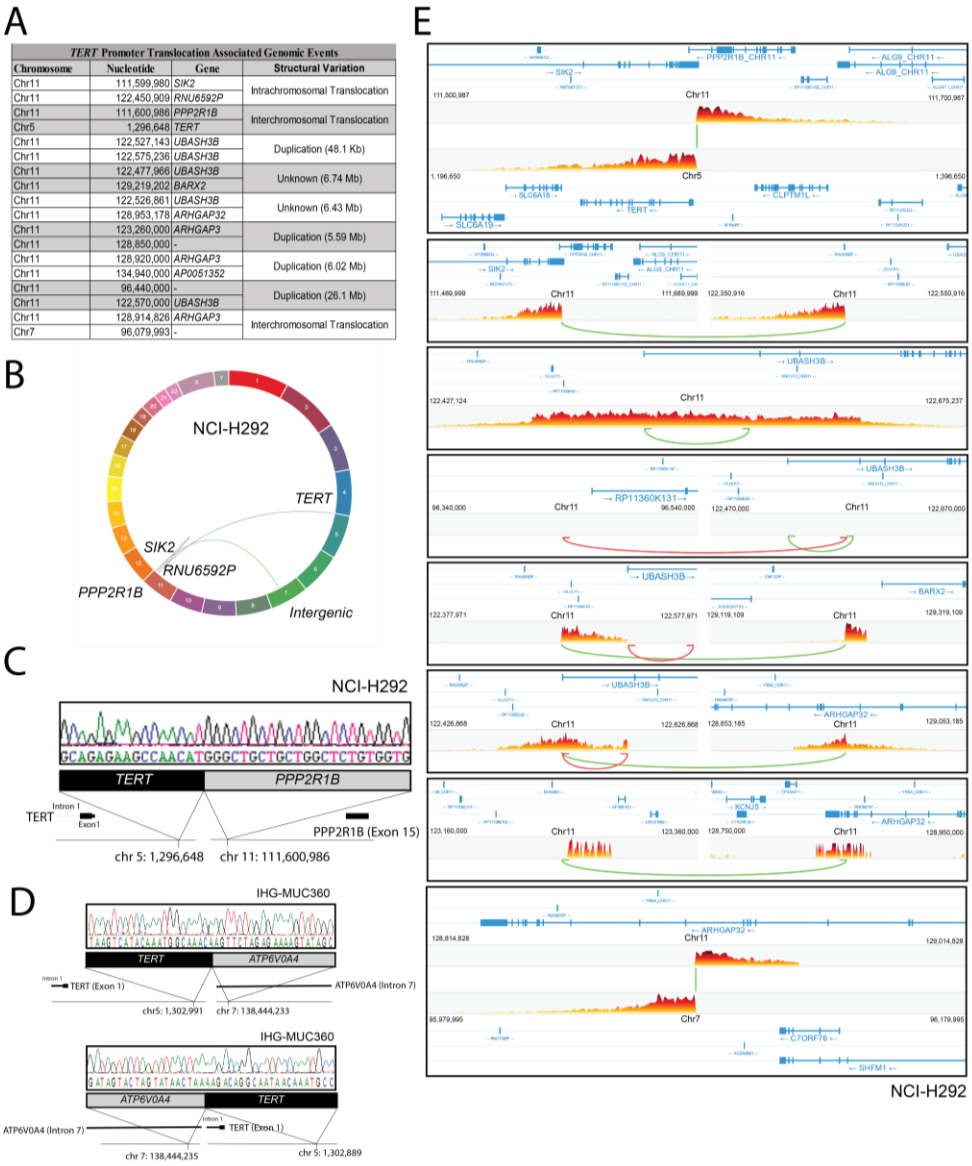


Figure 3-6. Gene fusion transcripts detected by RNA sequencing.



TERT Promoter Translocation Associated Genomic Events			
Chromosome	**Nucleotide**	**Gene**	**Structural Variation**
Chr11	111,599,980	*SIK2*	Intrachromosomal Translocation
Chr11	122,450,909	*RNU6592P*	
Chr11	111,600,986	*PPP2R1B*	Interchromosomal Translocation
Chr5	1,296,648	*TERT*	
Chr11	122,527,143	*UBASH3B*	Duplication (48.1 Kb)
Chr11	122,575,236	*UBASH3B*	
Chr11	122,477,966	*UBASH3B*	Unknown (6.74 Mb)
Chr11	129,219,202	*BARX2*	
Chr11	122,526,861	*UBASH3B*	Unknown (6.43 Mb)
Chr11	128,953,178	*ARHGAP32*	
Chr11	123,260,000	*ARHGAP3*	Duplication (5.59 Mb)
Chr11	128,850,000	-	
Chr11	128,920,000	*ARHGAP3*	Duplication (6.02 Mb)
Chr11	134,940,000	*AP0051352*	
Chr11	96,440,000	-	Duplication (26.1 Mb)
Chr11	122,570,000	*UBASH3B*	
Chr11	128,914,826	*ARHGAP3*	Interchromosomal Translocation
Chr7	96,079,993	-	

Figure 3-7. Discovery and Validation of Recurrent Structural Variation at the *TERT* Promoter in MEC.

A) The coordinates for the nine linked high and low confidence structural events associated with a structural rearrangement in which the 5' region of the *PPP2R1B* gene is rearranged to replace the *TERT* promoter in NCI-H292 are listed. Additional structural variations to Chr11 include potential deletions and duplications of *UBASH3B*, *ARHGAP32*, and the *BARX2* genes as well as translocation of *ARHGAP3* and *SIK2*. Linked read estimated breakpoints are shown. **B)** Genome-wide view of the chained structural events leading to the *PPP2R1B* to *TERT* translocation, **C)** Sanger sequencing was used to validate the *PPP2R1B* to *TERT* rearrangement breakpoint in NCI-H292. **D)** Sanger sequencing presenting the chromosome der(5) (left) and chromosome der(7) (right) junction at single base pair resolution of the TERT rearrangement identified in the IHG-360 MEC cell line. **E**) Structure of individual breakpoint events in NCI-H292, identified by linked read sequencing.

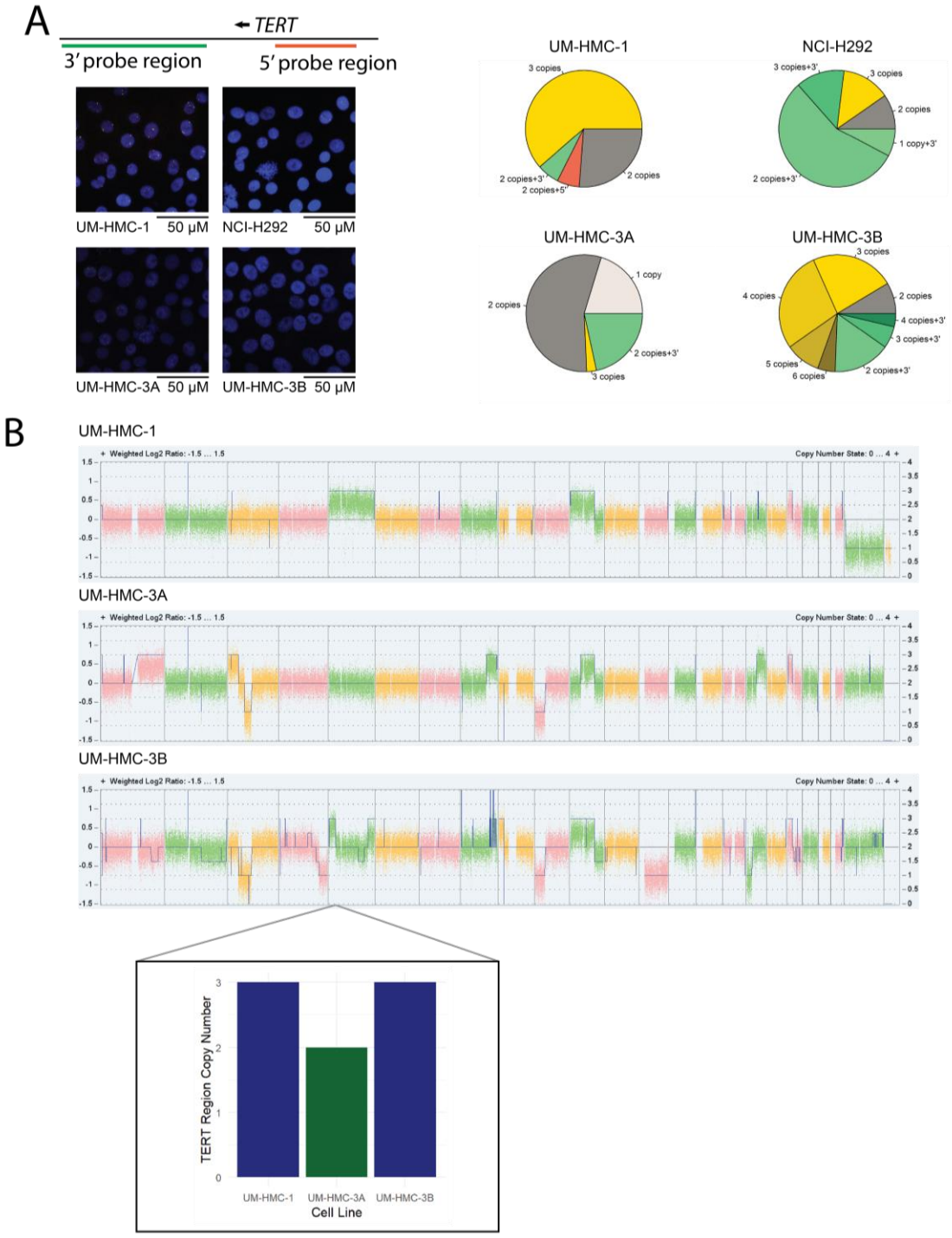


Figure 3-8. *TERT* locus copy number changes in MEC cell lines.

A) *TERT* locus visualized by FISH with *TERT* break apart probe with orange (5-TAMRA) fluorophore adjacent to the 5' region of *TERT* and green (5-Fluorescene) fluorophore adjacent to the 3' region (schematic top panel). Nucleus is blue (DAPI). Right panel shows quantification of *TERT* locus FISH, N>50 cells per cell line. **B)** Genome-wide copy number array reveals 3 *TERT* locus copies in UM-HMC-1 and UM-HMC-3B and 2 copies in UM-HMC-3B.

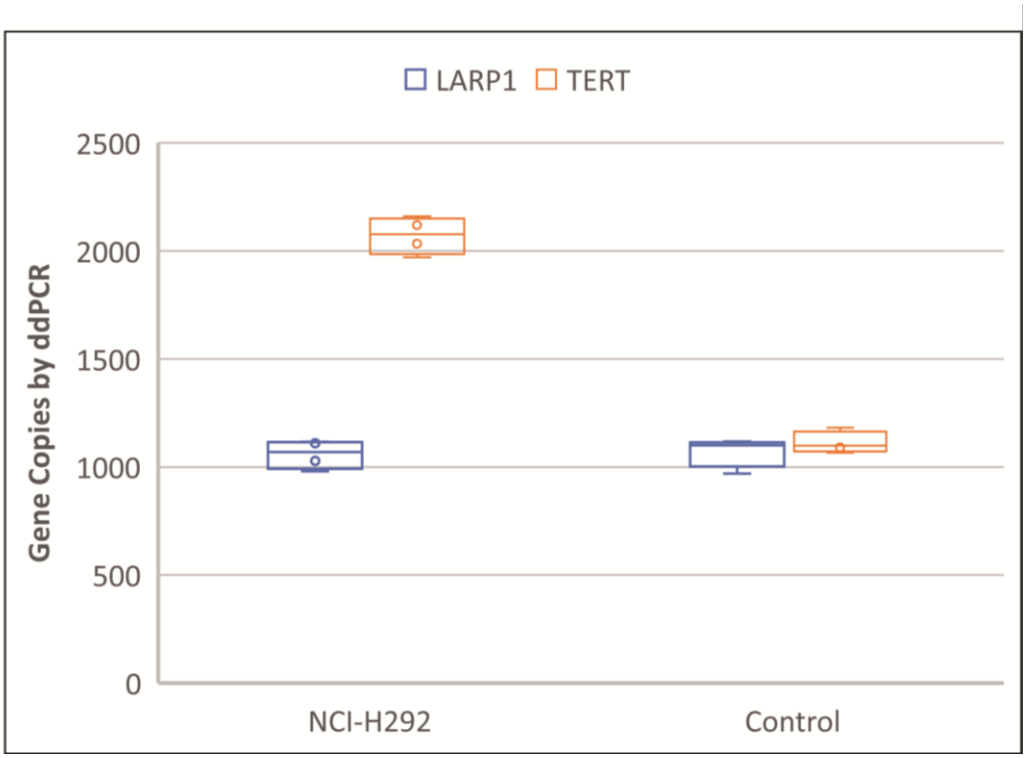
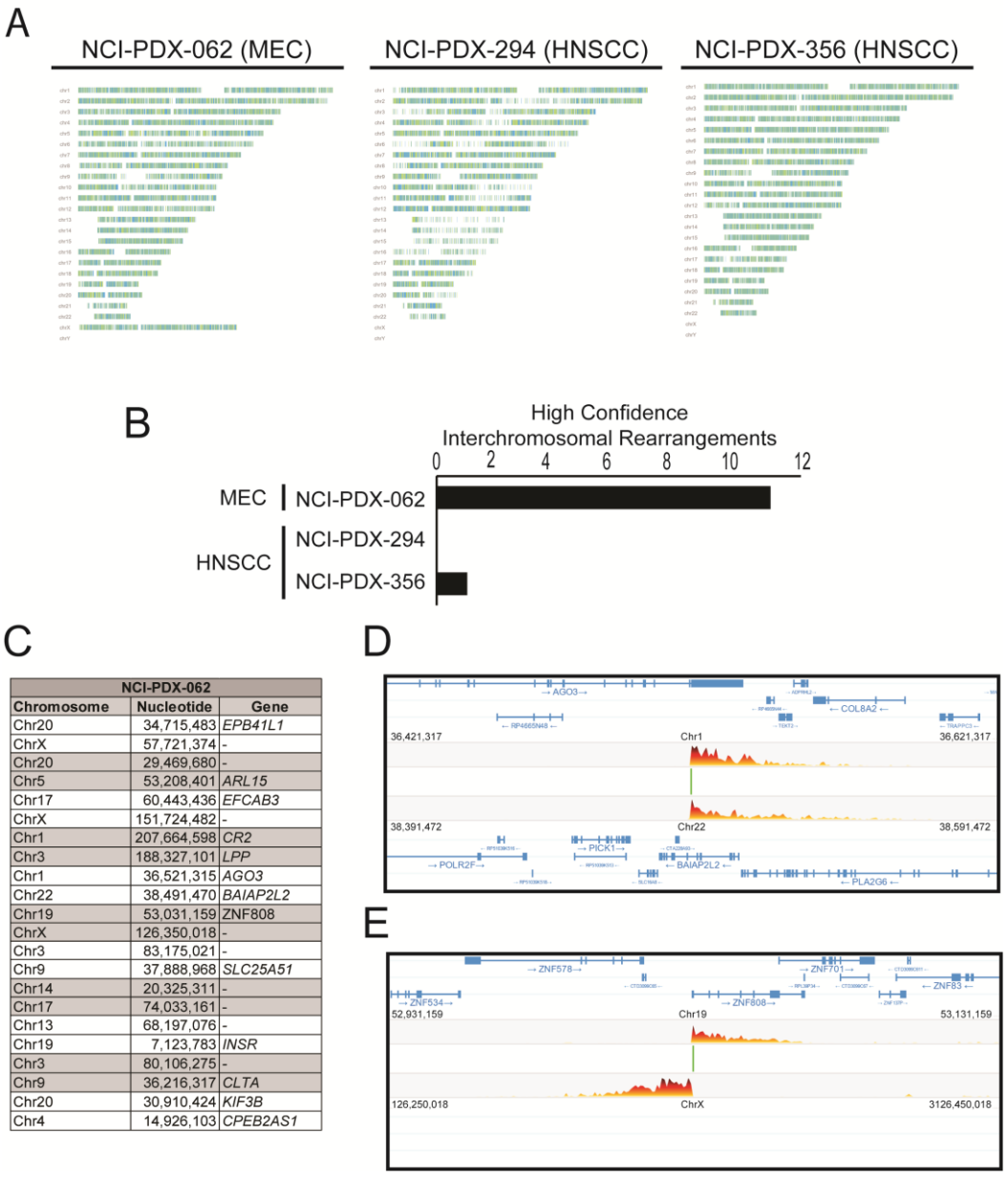


Figure 3-9. Confirmation of *TERT* Amplification in NCI-H292.

Digital droplet PCR using *TERT* probes was compared to a reference control gene, *LARP1*, and the relative ratio of positive droplets was calculated for genomic DNA isolated from NCI-H292 cells or normal reference human DNA as indicated.



NCI-PDX-062		
Chromosome	**Nucleotide**	**Gene**
Chr20	34,715,483	*EPB41L1*
ChrX	57,721,374	-
Chr20	29,469,680	-
Chr5	53,208,401	*ARL15*
Chr17	60,443,436	*EFCAB3*
ChrX	151,724,482	-
Chr1	207,664,598	*CR2*
Chr3	188,327,101	*LPP*
Chr1	36,521,315	*AGO3*
Chr22	38,491,470	*BAIAP2L2*
Chr19	53,031,159	ZNF808
ChrX	126,350,018	-
Chr3	83,175,021	-
Chr9	37,888,968	*SLC25A51*
Chr14	20,325,311	-
Chr17	74,033,161	-
Chr13	68,197,076	-
Chr19	7,123,783	*INSR*
Chr3	80,106,275	-
Chr9	36,216,317	*CLTA*
Chr20	30,910,424	*KIF3B*
Chr4	14,926,103	*CPEB2AS1*

Figure 3-10. Linked Read Sequencing Discovers Novel Chromosome Rearrangements in NCI-PDX-062 Mucoepidermoid Carcinoma Model.

A) Phase map summarizing data for a MEC PDX, NCI-PDX-062, as well as two HNSCC PDX models, NCI-PDX-294 and NCI-PDX-356. **B)** The number of high confidence chromosomal rearrangements identified in each model is shown. **C)** Estimated breakpoint details for each of the high confidence chromosomal rearrangements discovered in NCI-PDX-062. **D)** Representative image depicts the structural variation map chromosomal rearrangement leading to the *AGO3* final exon loss (the exon contains the *AGO3* 3-UTR region) in NCI-PDX-062. **E)** Representative image depicts chromosomal rearrangement leading to the *ZNF808* promoter translocation.

A

Cohort Clinical Details		
Age (Years)		52.4
Gender	M	10
	F	1
Smoking Status	Former	5
	Never	5
	Unk	1
Site	Parotid	4
	Other	7
Grade	High	1
	Intermediate	7
	Low	3

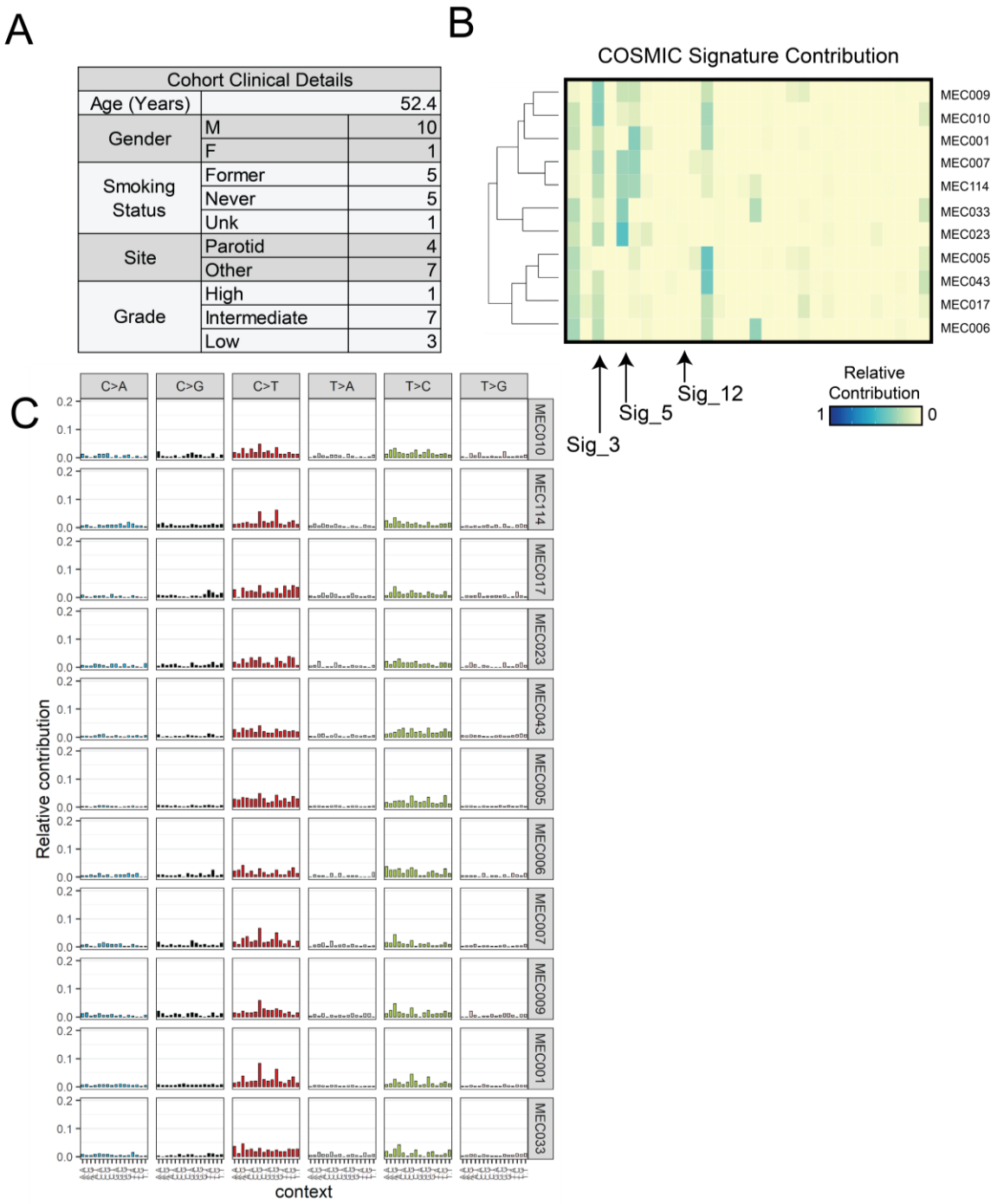


Figure 3-11. Mutational Signature Deconvolution.

A) Clinical variables of the MEC tumors in our exome sequencing cohort. **B)** Molecular signature analysis. Heat map shows the relative contribution of each COSMIC molecular signature to the overall mutational burden of each tumor. Prevalent signatures are highlighted with an arrow **C)** Relative contribution of each possible somatic SNV pattern across each MEC tumor. Each of the six types of DNA base substitutions was described for each sample separately. Within each chart, variation in the relative contribution of each mutation type based on the 5' and 3' contexts is displayed for a total of 96 trinucleotide mutation patterns.

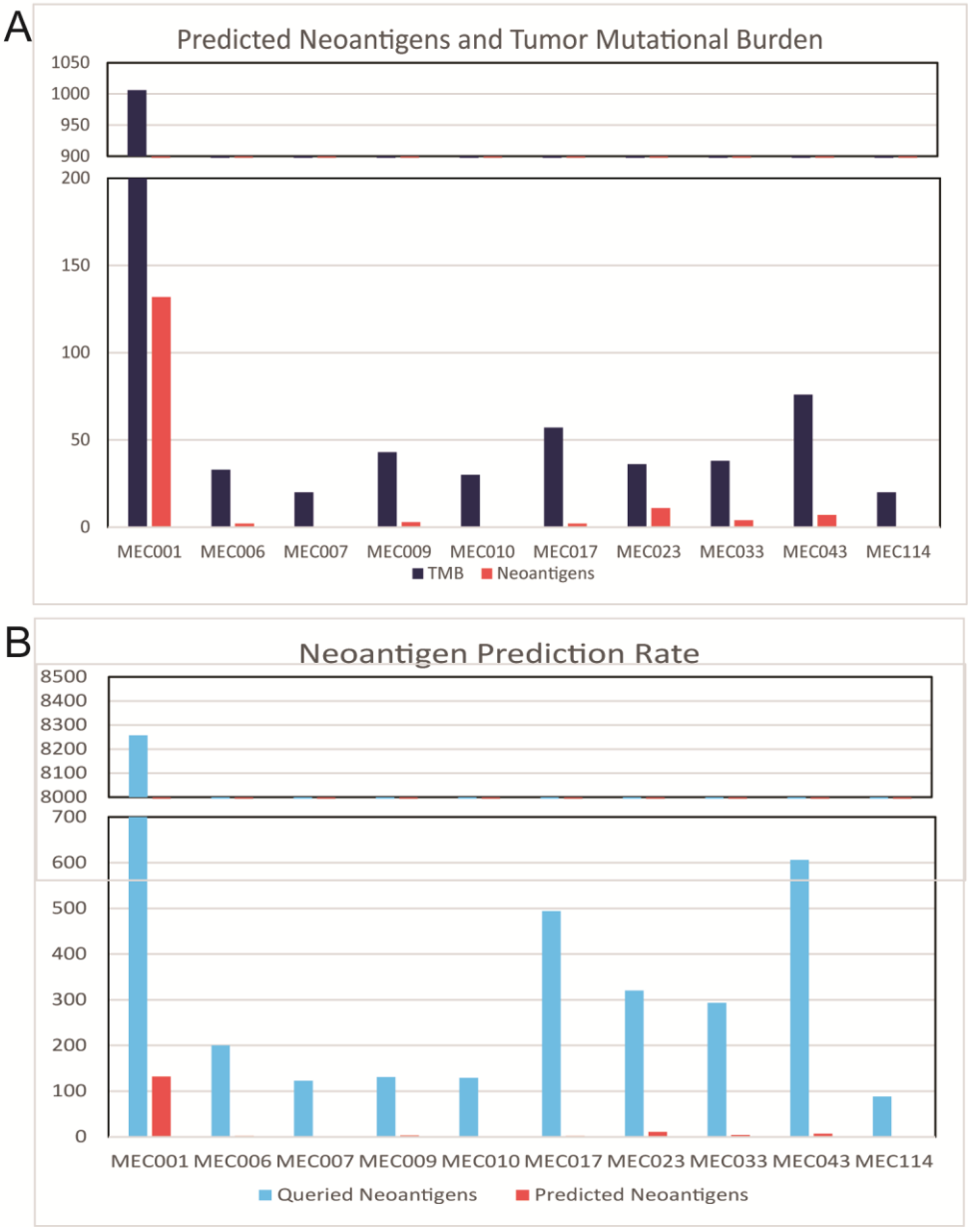


Figure 3-12. Total Mutational Burden and Predicted Neoantigen Load.

A) Predicted neoantigens (red) and total exome tumor mutational burden (dark blue) in each sample. Predicted neoantigens were predicted to bind to the individualized HLA haplotype at an affinity of >500nM, as predicted using the software NetMHCpan. **B)** Predicted high-affinity neoantigens (red) compared with the number of queried neoantigens (light blue). Queried neoantigens were determined from all amino acid changes caused by non-synonymous coding mutations in each sample.

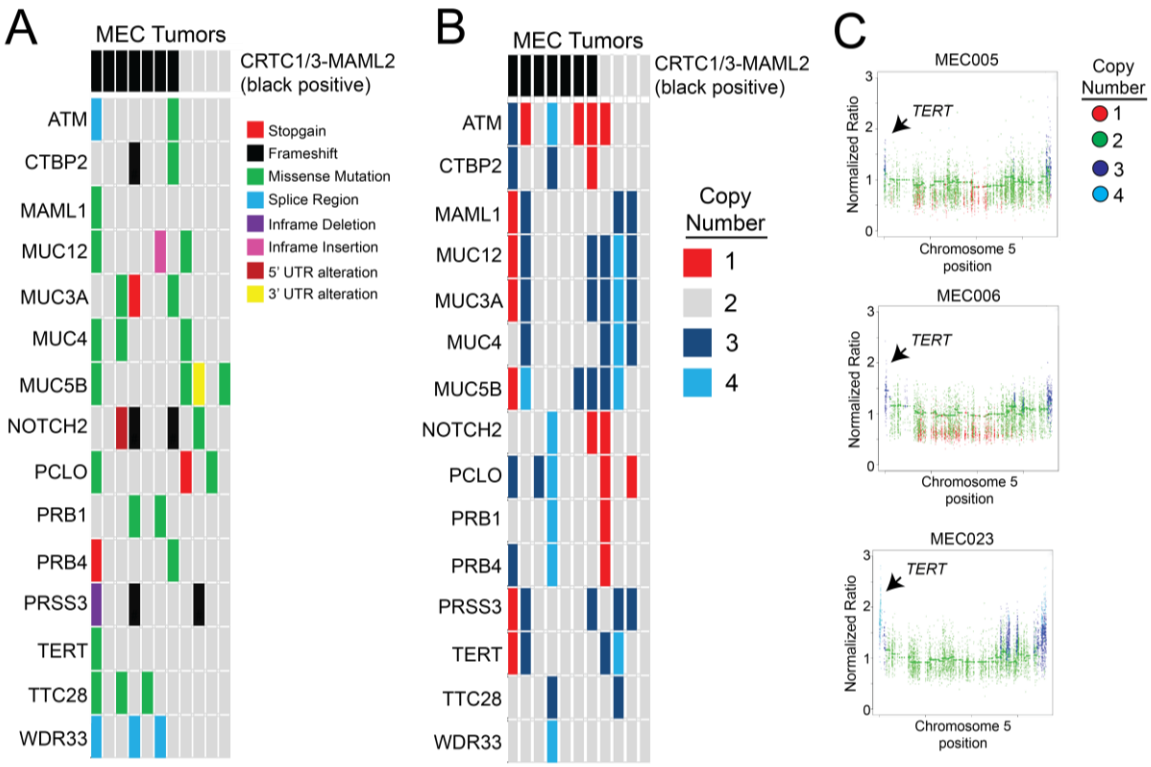


Figure 3-13. Exome Sequencing of MEC Tumors Identifies Recurrent NOTCH2 and TERT Alterations as well as a Double Strand Break-Associated Molecular Signature

A) Oncoplot highlighting recurrent single nucleotide variants and insertions/deletions (INDELs) in the cohort. *CRTC1/3-MAML2* fusion status of each tumor is shown in the top bar. **B)** Copy number annotation of each of the highlighted genes from panel C, along with *CRTC1/3-MAML2* fusion status. **C)** Representative Manhattan plots of Chromosome 5 from the three tumors with called focal *TERT* amplification.

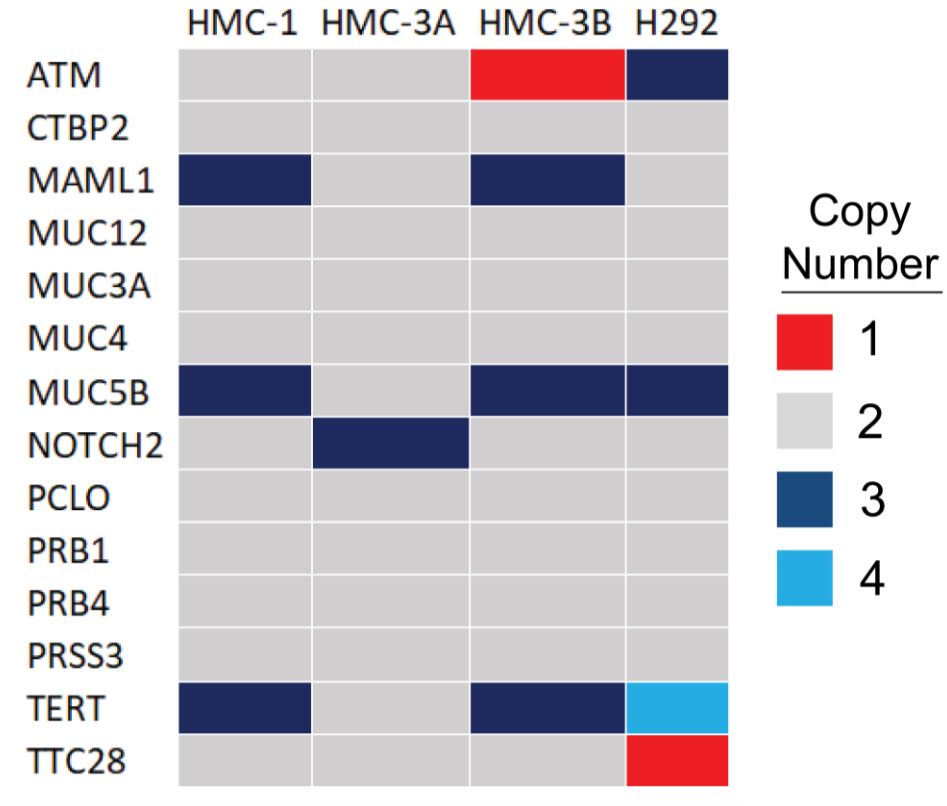
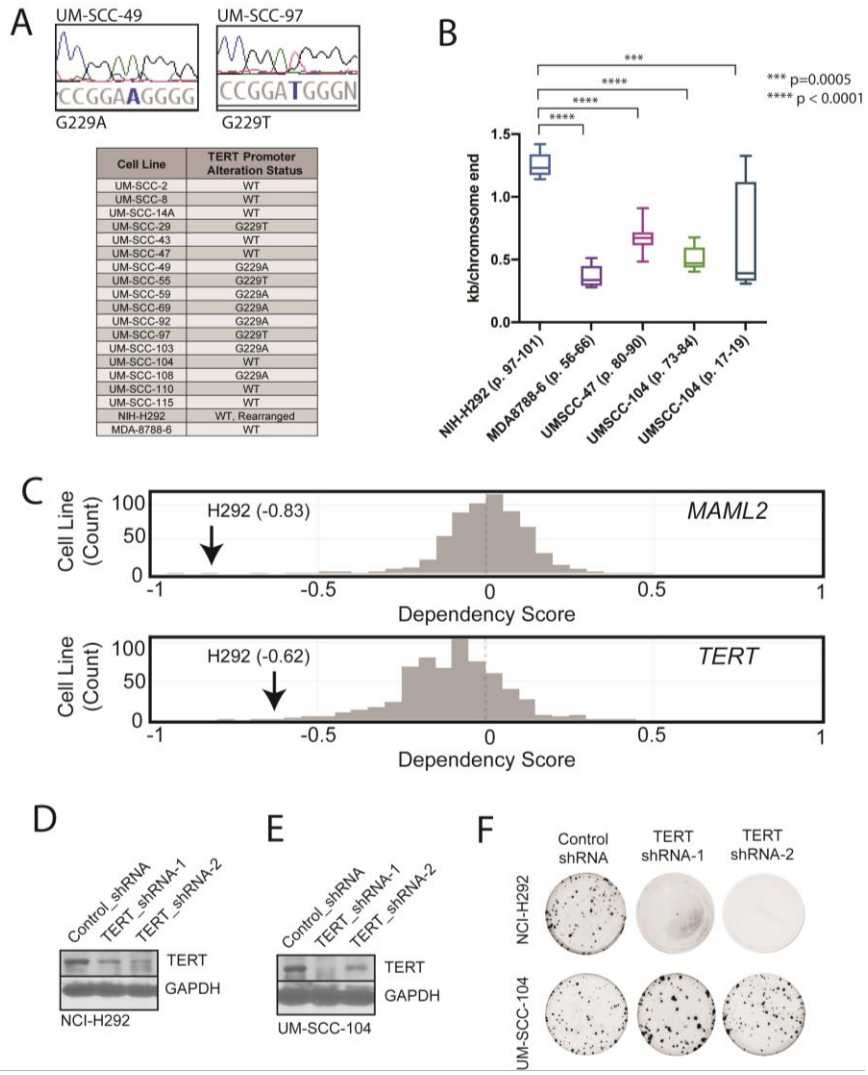


Figure 3-14. Copy Number Alterations of MEC Cell Lines.

Genes pictured are those included in figure 3-10, panel B. NCI-H292 copy number alterations are publicly available from the Cancer Dependency Map (DepMap Public 22Q4) (183)



Cell Line	TERT Promoter Alteration Status
UM-SCC-2	WT
UM-SCC-8	WT
UM-SCC-14A	WT
UM-SCC-29	G229T
UM-SCC-43	WT
UM-SCC-47	WT
UM-SCC-49	G229A
UM-SCC-55	G229T
UM-SCC-59	G229A
UM-SCC-69	G229A
UM-SCC-92	G229A
UM-SCC-97	G229T
UM-SCC-103	G229A
UM-SCC-104	WT
UM-SCC-108	G229A
UM-SCC-110	WT
UM-SCC-115	WT
NIH-H292	WT, Rearranged
MDA-8788-6	WT

Figure 3-15. Functional Evaluation of the Role of TERT in NCI-H292.

A) *TERT* promotor mutation status was characterized in multiple models. Sanger sequencing confirmation of G229A (UMSCC-49) and G229T (UMSCC-97) is illustrated. Results from a series of HNSCC and rare cancer models are summarized in the table. **B)** Telomere length in NCI-H292 compared to other patient derived cell lines (UMSCC-47, UMSCC-104, and MDA8788-6). Telomere length in NCI-H292 was significantly longer compared to all other tested cell lines. Data represents three independent experiments each run in triplicate. **C)** Dependency Score. The CERESdepletion score is based on data from a depletion assay. A lower CERESscore indicates a higher likelihood that the gene is essential in a given cell line. A score of zero indicates that a gene is not essential and a score of -1 is comparable to the median of all pan-essential genes. Dropout CRISPR screening data was downloaded from in the Avana 19Q3 public release. *MAML2* and *TERT* are essential for survival in NCI-H292. **D and E)** *TERT* shRNA were used to infect NCI-H292 or UM-SCC-104 cells and Western blot analysis used to check changes in TERT protein expression. **F)** Cells from D&E were re-plated into Clonogenic cell survival assays after 48 hours and grown in culture in parallel for >10 days at which point plates were imaged to show relative colony formation for each condition.

3.7 Tables

Quality Metric	NCI-H292	UM-HMC-1	UM-HMC-3A	UM-HMC-3B
Fraction of bases aligned:	0.9	0.9	0.9	0.9
Mean read length:	6,191.7	8,705.1	7,507.9	6,122.4
Median read length:	4,803	6,172	5,154	4,688
Number of reads:	617,700	3,437,834	4,130,516	4,052,710
Read length (N50):	9,050	14,077	13,003	9,460
STDEV read length:	5,201.8	8,218.7	7,488.6	5,466.5
Total bases:	3,824,585,725	29,926,831,896	31,011,601,676	24,812,207,056
Total bases aligned:	3,519,913,355	26,828,113,868	27,870,768,046	22,523,319,498

Table 3-1. Quality Control Variables for Nanopore Sequencing Data.

Total and mapped reads from long-read genome sequencing data are shown.

Breakpoint position 1				Breakpoint Position 2			
chrom	position	Ensembl ID	Gene Name	chrom	position	Ensembl ID	Gene Name
chr1	108022	00000238009	-	chr1	253248	00000241670	-
chr1	253248	00000241670	-	chr1	253248	00000241670	-
chr1	111776	00000238009	-	chr1	258075	00000241670	-
chr1	258075	00000241670	-	chr1	258075	00000241670	-
chr1	12934135	-	-	chr1	13042574	00000270601	Pramef5
chr1	13042574	00000270601	Pramef5	chr1	13042574	00000270601	Pramef5
chr1	13493764	-	-	chr1	13714034	-	-
chr1	13714034	-	-	chr1	13714034	-	-
chr1	13523168	00000204486	-	chr1	13742308	00000204478	PRAMEF20
chr1	13742308	00000204478	PRAMEF20	chr1	13742308	00000204478	PRAMEF20
chr1	16835911	-	-	chr1	16985717	-	-
chr1	16985717	-	-	chr1	16985717	-	-
chr1	16906808	00000219481	NBPF1	chr1	146248321	00000232637	-
chr1	146248321	00000232637	-	chr1	146248321	00000232637	-
chr1	58329533	00000226759	DAB1-AS1	chr7	318850	-	-
chr7	318850	-	-	chr7	318850	-	-
chr1	61769177	00000162599	NFIA	chr20	23657143	-	-
chr20	23657143	-	-	chr20	23657143	-	-
chr1	66439644	00000184588	PDE4B	chr18	25977795	-	-
chr18	25977795	-	-	chr18	25977795	-	-
chr1	69905372	-	-	chr5	173660674	00000170091	HMP19
chr5	173660674	00000170091	HMP19	chr5	173660674	00000170091	HMP19
chr1	71758631	00000229956	ZRANB2-AS2	chr8	104607848	00000176406	RIMS2
chr8	104607848	00000176406	RIMS2	chr8	104607848	00000176406	RIMS2
chr1	73420127	-	-	chr19	35434990	00000168661	ZNF30
chr19	35434990	00000168661	ZNF30	chr19	35434990	00000168661	ZNF30
chr1	94125348	00000137936	BCAR3	chr5	44630283	-	-
chr5	44630283	-	-	chr5	44630283	-	-
chr1	95643771	00000152078	TMEM56	chr11	99566939	00000149972	CNTN5
chr11	99566939	00000149972	CNTN5	chr11	99566939	00000149972	CNTN5
chr1	96531861	-	-	chr6	26275495	-	-
chr6	26275495	-	-	chr6	26275495	-	-
chr1	103309182	00000230864	-	chr2	102326192	00000071054	map4k4
chr2	102326192	00000071054	map4k4	chr2	102326192	00000071054	map4k4
chr1	105602406	-	-	chr3	23613435	00000182247	UBE2E2
chr3	23613435	00000182247	UBE2E2	chr3	23613435	00000182247	UBE2E2
chr1	109269433	00000143107	FNDC7	chr11	78841299	00000149256	TENM4
chr11	78841299	00000149256	TENM4	chr11	78841299	00000149256	TENM4
chr1	115179641	00000175984	DENND2C	chr7	152073127	00000055609	KMT2C
chr7	152073127	00000055609	KMT2C	chr7	152073127	00000055609	KMT2C

chr1	120881744	00000227193	-	chr1	206177497	-	-
chr1	206177497	-	-	chr1	206177497	-	-
chr1	143886880	-	-	chr1	149440255	00000203817	-
chr1	149440255	00000203817	-	chr1	149440255	00000203817	-
chr1	143892565	-	-	chr1	149445752	00000203817	-
chr1	149445752	00000203817	-	chr1	149445752	00000203817	-
chr1	144029600	00000196369	SRGAP2B	chr1	206492130	-	-
chr1	206492130	-	-	chr1	206492130	-	-
chr1	144035119	00000196369	SRGAP2B	chr1	206497740	-	-
chr1	206497740	-	-	chr1	206497740	-	-
chr1	144045935	00000196369	SRGAP2B	chr1	206508559	-	-
chr1	206508559	-	-	chr1	206508559	-	-
chr1	144055346	00000196369	SRGAP2B	chr1	206516990	00000163486	-
chr1	206516990	00000163486	-	chr1	206516990	00000163486	-
chr1	144321363	00000235398	-	chr1	144501598	00000236943	-
chr1	144501598	00000236943	-	chr1	144501598	00000236943	-
chr1	144517467	00000236943	-	chr1	147925536	00000215863	-
chr1	147925536	00000215863	-	chr1	147925536	00000215863	-
chr1	144552238	-	-	chr1	148511801	-	-
chr1	148511801	-	-	chr1	148511801	-	-
chr1	144821035	00000168614	-	chr1	145308827	00000163386	-
chr1	145308827	00000163386	-	chr1	145308827	00000163386	-
chr1	144826617	00000168614	-	chr1	145312176	00000163386	-
chr1	145312176	00000163386	-	chr1	145312176	00000163386	-
chr1	145376701	00000233396	LOC101928979	chr1	146473881	00000234225	-
chr1	146473881	00000234225	-	chr1	146473881	00000234225	-
chr1	146161405	00000232637	-	chr1	147702314	-	-
chr1	147702314	-	-	chr1	147702314	-	-
chr1	148574507	00000243452	-	chr1	148736515	-	-
chr1	148736515	-	-	chr1	148736515	-	-
chr1	148600202	-	-	chr1	148762091	-	-
chr1	148762091	-	-	chr1	148762091	-	-
chr1	148612244	-	-	chr1	148773838	-	-
chr1	148773726	-	-	chr1	148773838	-	-
chr1	149293283	-	-	chr1	149678793	00000234232	-
chr1	149678793	00000234232	-	chr1	149678793	00000234232	-
chr1	149336385	-	-	chr1	149711999	00000234232	-
chr1	149711999	00000234232	-	chr1	149711999	00000234232	-
chr1	170949717	00000117501	MROH9	chr7	40509273	00000175600	SUGCT
chr7	40509273	00000175600	SUGCT	chr7	40509273	00000175600	SUGCT
chr1	174604300	00000152061	RABGAP1L	chr5	89555227	00000248555	-
chr5	89555227	00000248555	-	chr5	89555227	00000248555	-
chr1	200572468	00000118193	KIF14	chr11	105381914	00000254998	-

chr11	105381914	00000254998	-	chr11	105381914	00000254998	-
chr1	215724614	-	-	chr4	92130799	00000184305	CCSER1
chr4	92130799	00000184305	CCSER1	chr4	92130799	00000184305	CCSER1
chr1	231561370	-	-	chr19	14132039	-	-
chr19	14132039	-	-	chr19	14132039	-	-
chr10	10199821	-	-	chr19	30592000	-	-
chr19	30592000	-	-	chr19	30592000	-	-
chr10	17892508	00000183748	-	chr10	18135899	00000120586	-
chr10	18135899	00000120586	-	chr10	18135899	00000120586	-
chr10	32882669	00000150076	CCDC7	chr15	26863639	00000166206	GABRB3
chr15	26863639	00000166206	GABRB3	chr15	26863639	00000166206	GABRB3
chr10	46760393	00000204177	BMS1P5	chr10	48949944	00000204164	-
chr10	48949944	00000204164	-	chr10	48949944	00000204164	-
chr10	46814188	-	-	chr10	49003903	-	-
chr10	49003903	-	-	chr10	49003903	-	-
chr10	47311566	-	-	chr10	49020044	-	-
chr10	49020044	-	-	chr10	49020044	-	-
chr10	47584349	-	-	chr10	51458733	-	-
chr10	51458733	-	-	chr10	51458733	-	-
chr10	47952589	-	-	chr10	51893762	-	-
chr10	51893762	-	-	chr10	51893762	-	-
chr10	51378229	00000204152	TIMM23B	chr10	51621319	00000138297	-
chr10	51621319	00000138297	-	chr10	51621319	00000138297	-
chr10	63708674	00000150347	ARID5B	chr10	110726382	-	-
chr10	110726382	-	-	chr10	110726382	-	-
chr10	65138941	00000171988	JMJD1C	chr3	150885432	00000144893	MED12L
chr3	150885432	00000144893	MED12L	chr3	150885432	00000144893	MED12L
chr10	68238607	00000183230	CTNNA3	chr13	33231320	00000083642	PDS5B
chr13	33231320	00000083642	PDS5B	chr13	33231320	00000083642	PDS5B
chr10	69413472	00000183230	CTNNA3	chr2	51183557	00000179915	NRXN1
chr2	51183557	00000179915	NRXN1	chr2	51183557	00000179915	NRXN1
chr10	81507118	00000226381	-	chr10	81611812	-	-
chr10	81611812	-	-	chr10	81611812	-	-
chr10	87568830	00000182771	GRID1	chr5	85737639	-	-
chr5	85737639	-	-	chr5	85737639	-	-
chr10	90167813	00000184719	RNLS	chr7	130349696	00000213265	TSGA13
chr7	130349696	00000213265	TSGA13	chr7	130349696	00000213265	TSGA13
chr10	91473242	00000138182	KIF20B	chr21	9864370	-	-
chr21	9864370	-	-	chr21	9864370	-	-
chr10	110939430	-	-	chr11	19406301	00000255270	-
chr11	19406301	00000255270	-	chr11	19406301	00000255270	-
chr10	119710595	-	-	chr14	25851606	-	-
chr14	25851606	-	-	chr14	25851606	-	-
chr10	131886361	00000237489	-	chr10	132277770	00000236303	-

chr10	132277770	00000236303	-	chr10	132277770	00000236303	-
chr11	12959050	00000187079	TEAD1	chr5	84834164	-	-
chr5	84834164	-	-	chr5	84834164	-	-
chr11	26406038	00000134343	ANO3	chr6	50667228	-	-
chr6	50667228	-	-	chr6	50667228	-	-
chr11	30516305	00000066382	MPPED2	chr17	4732007	-	-
chr17	4732007	-	-	chr17	4732007	-	-
chr11	36352638	00000135362	PRR5L	chr15	41764467	00000137815	rtf1
chr15	41764467	00000137815	rtf1	chr15	41764467	00000137815	rtf1
chr11	36697833	-	-	chr6	146642381	00000152822	GRM1
chr6	146642381	00000152822	GRM1	chr6	146642381	00000152822	GRM1
chr11	37838603	-	-	chr4	40801654	00000179299	NSUN7
chr4	40801654	00000179299	NSUN7	chr4	40801654	00000179299	NSUN7
chr11	40460692	00000148948	LRRC4C	chr3	99501436	00000144810	COL8A1
chr3	99501436	00000144810	COL8A1	chr3	99501436	00000144810	COL8A1
chr11	58216660	-	-	chr11	82981999	00000137500	CCDC90B
chr11	82981999	00000137500	CCDC90B	chr11	82981999	00000137500	CCDC90B
chr11	78660848	00000149256	TENM4	chrX	121468659	-	-
chrX	121468659	-	-	chrX	121468659	-	-
chr11	80747851	-	-	chr14	49090813	-	-
chr14	49090813	-	-	chr14	49090813	-	-
chr11	86948296	00000166575	TMEM135	chr5	32149610	00000113384	GOLPH3
chr5	32149610	00000113384	GOLPH3	chr5	32149610	00000113384	GOLPH3
chr11	91941746	-	-	chr7	70179989	00000158321	AUTS2
chr7	70179989	00000158321	AUTS2	chr7	70179989	00000158321	AUTS2
chr11	101655306	00000137672	TRPC6	chr2	166375095	00000178662	CSRNP3
chr2	166375095	00000178662	CSRNP3	chr2	166375095	00000178662	CSRNP3
chr11	103467858	-	-	chr17	41606830	00000067596	DHX8
chr17	41606830	00000067596	DHX8	chr17	41606830	00000067596	DHX8
chr11	105688554	00000263371	-	chr9	104928064	-	-
chr9	104928064	-	-	chr9	104928064	-	-
chr11	107356917	-	-	chr9	80039427	00000156049	GNA14
chr9	80039427	00000156049	GNA14	chr9	80039427	00000156049	GNA14
chr12	4175726	-	-	chr2	42490994	00000143924	EML4
chr2	42490994	00000143924	EML4	chr2	42490994	00000143924	EML4
chr12	8470522	00000226091	LINC00937	chr15	50102819	-	-
chr15	50102819	-	-	chr15	50102819	-	-
chr12	20399557	-	-	chr13	38495388	-	-
chr13	38495388	-	-	chr13	38495388	-	-
chr12	56895579	-	-	chr2	157120003	-	-
chr2	157120003	-	-	chr2	157120003	-	-
chr12	62915344	00000061987	MON2	chr5	158665875	00000253256	-
chr5	158665875	00000253256	-	chr5	158665875	00000253256	-
chr12	110783845	00000174437	ATP2A2	chr6	169974996	00000184465	WDR27

chr6	169974996	00000184465	WDR27	chr6	169974996	00000184465	WDR27
chr12	121407000	00000241388	HNF1A-AS1	chr4	169633928	00000129116	PALLD
chr4	169633928	00000129116	PALLD	chr4	169633928	00000129116	PALLD
chr13	35068813	00000271850	-	chr9	73888983	00000083067	TRPM3
chr9	73888983	00000083067	TRPM3	chr9	73888983	00000083067	TRPM3
chr13	39008960	-	-	chrX	101780165	-	-
chrX	101780165	-	-	chrX	101780165	-	-
chr13	53043261	00000136108	CKAP2	chr7	10077537	-	-
chr7	10077537	-	-	chr7	10077537	-	-
chr14	23724665	00000215277	RNF212B	chr15	26155327	00000206187	LOC100128714
chr15	26155327	00000206187	LOC100128714	chr15	26155327	00000206187	LOC100128714
chr14	53724969	00000237356	-	chr6	125861260	-	-
chr6	125861260	-	-	chr6	125861260	-	-
chr14	66207222	00000033170	FUT8	chr7	92378806	00000105810	CDK6
chr7	92378806	00000105810	CDK6	chr7	92378806	00000105810	CDK6
chr14	68314812	00000182185	RAD51B	chrX	93148186	-	-
chrX	93148186	-	-	chrX	93148186	-	-
chr14	72379814	-	-	chr9	100629581	-	-
chr9	100629581	-	-	chr9	100629581	-	-
chr14	85402443	-	-	chr5	94951841	-	-
chr5	94951841	-	-	chr5	94951841	-	-
chr14	93456476	00000100605	ITPK1	chr8	137406655	-	-
chr8	137406655	-	-	chr8	137406655	-	-
chr14	106053746	00000211890	-	chr14	106173746	00000211895	-
chr14	106173746	00000211895	-	chr14	106173746	00000211895	-
chr15	20224049	-	-	chr15	21229941	-	-
chr15	21229941	-	-	chr15	21229941	-	-
chr15	20253407	-	-	chr15	21260033	-	-
chr15	21260033	-	-	chr15	21260033	-	-
chr15	20274631	-	-	chr15	21278101	-	-
chr15	21278101	-	-	chr15	21278101	-	-
chr15	20375578	-	-	chr15	21381894	-	-
chr15	21381894	-	-	chr15	21381894	-	-
chr15	20716977	-	-	chr15	23383547	-	-
chr15	23383547	-	-	chr15	23383547	-	-
chr15	20729522	-	-	chr15	23396006	-	-
chr15	23396006	-	-	chr15	23396006	-	-
chr15	20753599	00000258797	-	chr15	28609534	00000261497	-
chr15	28609534	00000261497	-	chr15	28609534	00000261497	-
chr15	20759179	-	-	chr15	23426188	-	-
chr15	23426188	-	-	chr15	23426188	-	-
chr15	20763484	-	-	chr15	23428152	-	-
chr15	23428152	-	-	chr15	23428152	-	-

chr15	20774795	00000181984	GOLGA8CP	chr15	23434870	-	-
chr15	23434870	-	-	chr15	23434870	-	-
chr15	20788149	-	-	chr15	23454165	-	-
chr15	23454165	-	-	chr15	23454165	-	-
chr15	20945221	00000260409	-	chr15	21950638	00000247765	-
chr15	21950638	00000247765	-	chr15	21950638	00000247765	-
chr15	20954899	00000260409	-	chr15	21960928	00000247765	-
chr15	21960928	00000247765	-	chr15	21960928	00000247765	-
chr15	20966903	00000259383	-	chr15	21973299	-	-
chr15	21973299	-	-	chr15	21973299	-	-
chr15	20993098	-	-	chr15	21999754	-	-
chr15	21999754	-	-	chr15	21999754	-	-
chr15	21024640	-	-	chr15	22030285	-	-
chr15	22030285	-	-	chr15	22030285	-	-
chr15	21097491	-	-	chr15	22108618	-	-
chr15	22108618	-	-	chr15	22108618	-	-
chr15	21105261	-	-	chr15	22116354	-	-
chr15	22116354	-	-	chr15	22116354	-	-
chr15	21160714	00000258710	LINC01193	chr15	22170838	-	-
chr15	22170838	-	-	chr15	22170838	-	-
chr15	22688772	-	-	chr15	28748862	-	-
chr15	28748862	-	-	chr15	28748862	-	-
chr15	22707536	00000221000	-	chr15	28768332	00000238607	-
chr15	28768332	00000238607	-	chr15	28768332	00000238607	-
chr15	23433040	-	-	chr15	23597719	-	-
chr15	23597719	-	-	chr15	23597719	-	-
chr15	23579416	-	-	chr15	28601876	00000237850	-
chr15	28601876	00000237850	-	chr15	28601876	00000237850	-
chr15	23714349	-	-	chr3	147300488	-	-
chr3	147300488	-	-	chr3	147300488	-	-
chr15	50325535	00000104043	ATP8B4	chr8	80728597	00000249328	LOC101927040
chr8	80728597	00000249328	LOC101927040	chr8	80728597	00000249328	LOC101927040
chr15	51120007	-	-	chr3	69284073	00000114541	FRMD4B
chr3	69284073	00000114541	FRMD4B	chr3	69284073	00000114541	FRMD4B
chr15	54973410	-	-	chr15	86210867	00000259367	-
chr15	86210867	00000259367	-	chr15	86210867	00000259367	-
chr15	75022668	-	-	chr9	85409628	-	-
chr9	85409628	-	-	chr9	85409628	-	-
chr15	78249631	00000261403	-	chr15	97692591	-	-
chr15	97692591	-	-	chr15	97692591	-	-
chr15	82991195	-	-	chr15	84964216	00000235370	-
chr15	84964216	00000235370	-	chr15	84964216	00000235370	-

chr15	84877004	00000225151	GOLGA2P7	chr15	85755697	00000259295	LOC440300
chr15	85755697	00000259295	LOC440300	chr15	85755697	00000259295	LOC440300
chr15	90233088	00000166821	PEX11A	chrX	35736574	-	-
chrX	35736574	-	-	chrX	35736574	-	-
chr16	14184890	00000186260	MKL2	chr7	26822028	00000005020	SKAP2
chr7	26822028	00000005020	SKAP2	chr7	26822028	00000005020	SKAP2
chr16	15302608	-	-	chr16	18214121	00000259929	-
chr16	18214121	00000259929	-	chr16	18214121	00000259929	-
chr16	21478353	00000258186	-	chr16	21910007	00000185710	-
chr16	21910007	00000185710	-	chr16	21910007	00000185710	-
chr16	21481709	00000258186	-	chr16	21912843	00000185710	-
chr16	21912843	00000185710	-	chr16	21912843	00000185710	-
chr16	21768401	00000155719	OTOA	chr16	22583527	00000257838	LOC653786
chr16	22583527	00000257838	LOC653786	chr16	22583527	00000257838	LOC653786
chr16	21800005	-	-	chr16	22615543	-	-
chr16	22615543	-	-	chr16	22615543	-	-
chr16	25937664	00000182601	HS3ST4	chr4	49600834	-	-
chr4	49600834	-	-	chr4	49600834	-	-
chr16	29446163	00000261740	BOLA2-SMG1P6	chr16	30287111	00000183604	SMG1P5
chr16	30287111	00000183604	SMG1P5	chr16	30287111	00000183604	SMG1P5
chr16	29448004	00000261740	BOLA2-SMG1P6	chr16	30289028	00000183604	SMG1P5
chr16	30289028	00000183604	SMG1P5	chr16	30289028	00000183604	SMG1P5
chr16	29448889	00000238639	-	chr16	30290507	00000183604	SMG1P5
chr16	30290507	00000183604	SMG1P5	chr16	30290507	00000183604	SMG1P5
chr16	29498024	00000169203	-	chr16	30234265	00000183604	SMG1P5
chr16	30234265	00000183604	SMG1P5	chr16	30234265	00000183604	SMG1P5
chr16	31766266	00000197302	LOC107983990	chr2	195229681	00000225539	LOC101927406
chr2	195229681	00000225539	LOC101927406	chr2	195229681	00000225539	LOC101927406
chr16	32101482	-	-	chr16	33044653	-	-
chr16	33044653	-	-	chr16	33044653	-	-
chr16	32143991	00000230267	-	chr16	33086283	00000261599	-
chr16	33086283	00000261599	-	chr16	33086283	00000261599	-
chr16	32203806	00000260847	-	chr16	33144776	00000261682	-
chr16	33144776	00000261682	-	chr16	33144776	00000261682	-
chr16	32296218	-	-	chr16	33236752	00000205457	TP53TG3C
chr16	33236752	00000205457	TP53TG3C	chr16	33236752	00000205457	TP53TG3C
chr16	33705186	-	-	chr2	48479892	-	-
chr2	48479892	-	-	chr2	48479892	-	-
chr16	34914446	-	-	chr2	20414317	00000115884	sdc1
chr2	20414317	00000115884	sdc1	chr2	20414317	00000115884	sdc1

chr17	15276412	-	-	chrX	73017153	00000270641	Tsix
chrX	73017153	00000270641	Tsix	chrX	73017153	00000270641	Tsix
chr17	30824304	00000266718	-	chr17	58540061	00000062725	APPBP2
chr17	58540061	00000062725	APPBP2	chr17	58540061	00000062725	APPBP2
chr17	33256996	00000132141	CCT6B	chr5	136890481	00000152377	SPOCK1
chr5	136890481	00000152377	SPOCK1	chr5	136890481	00000152377	SPOCK1
chr17	34528161	-	-	chr17	34629582	00000161583	-
chr17	34629582	00000161583	-	chr17	34629582	00000161583	-
chr17	34574287	-	-	chr17	36330770	-	-
chr17	36330770	-	-	chr17	36330770	-	-
chr17	34574520	-	-	chr17	34738546	00000161583	-
chr17	34738546	00000161583	-	chr17	34738546	00000161583	-
chr17	34726927	00000161583	-	chr17	36301471	-	-
chr17	36301484	-	-	chr17	36301471	-	-
chr17	56510905	00000176160	hsf5	chr3	125991688	00000250934	-
chr3	125991688	00000250934	-	chr3	125991688	00000250934	-
chr18	1801538	-	-	chr4	144438656	00000153147	SMARCA5
chr4	144438656	00000153147	SMARCA5	chr4	144438656	00000153147	SMARCA5
chr18	7264725	-	-	chr5	57395312	-	-
chr5	57395312	-	-	chr5	57395312	-	-
chr18	21231157	00000154065	ANKRD29	chr5	144956410	00000186314	PRELID2
chr5	144956410	00000186314	PRELID2	chr5	144956410	00000186314	PRELID2
chr18	26068632	-	-	chr18	27951132	-	-
chr18	27951132	-	-	chr18	27951132	-	-
chr18	46943808	00000141627	DYM	chr8	118784436	-	-
chr8	118784436	-	-	chr8	118784436	-	-
chr18	47942168	-	-	chr4	171427452	-	-
chr4	171427452	-	-	chr4	171427452	-	-
chr18	55764370	00000049759	NEDD4L	chr19	52187013	-	-
chr19	52187013	-	-	chr19	52187013	-	-
chr18	69650710	-	-	chr2	131565771	-	-
chr2	131565771	-	-	chr2	131565771	-	-
chr18	73504242	-	-	chr4	151110291	00000170390	DCLK2
chr4	151110291	00000170390	DCLK2	chr4	151110291	00000170390	DCLK2
chr19	12029697	-	-	chr5	33546280	00000151388	adamts12
chr5	33546280	00000151388	adamts12	chr5	33546280	00000151388	adamts12
chr19	12415122	-	-	chr5	172063045	-	-
chr5	172063045	-	-	chr5	172063045	-	-
chr19	18795372	00000105662	CRTC1	chr8	67777507	-	-
chr8	67777507	-	-	chr8	67777507	-	-
chr19	23812025	00000197372	ZNF675	chr5	29821391	-	-
chr5	29821391	-	-	chr5	29821391	-	-
chr19	50609920	-	-	chrX	38496408	00000156298	TSPAN7

chrX	38496408	00000156298	TSPAN7	chrX	38496408	00000156298	TSPAN7
chr2	11922070	00000134324	LPIN1	chr9	16003190	00000164989	CCDC171
chr9	16003190	00000164989	CCDC171	chr9	16003190	00000164989	CCDC171
chr2	16534572	-	-	chr6	134611577	00000118515	SGK1
chr6	134611577	00000118515	SGK1	chr6	134611577	00000118515	SGK1
chr2	57724529	-	-	chr2	97399628	00000114988	LMAN2L
chr2	97399628	00000114988	LMAN2L	chr2	97399628	00000114988	LMAN2L
chr2	87912907	-	-	chr2	112099262	00000172965	MIR4435-2HG
chr2	112099262	00000172965	MIR4435-2HG	chr2	112099262	00000172965	MIR4435-2HG
chr2	90417775	-	-	chr2	91638781	-	-
chr2	91638781	-	-	chr2	91638781	-	-
chr2	90455101	-	-	chr2	91675509	-	-
chr2	91675509	-	-	chr2	91675509	-	-
chr2	90464877	-	-	chr2	91679526	00000219041	-
chr2	91679526	00000219041	-	chr2	91679526	00000219041	-
chr2	90485501	-	-	chr2	91707645	-	-
chr2	91707645	-	-	chr2	91707645	-	-
chr2	91633753	-	-	chr2	156192075	-	-
chr2	156192075	-	-	chr2	156192075	-	-
chr2	92325517	-	-	chr5	27260879	-	-
chr5	27260879	-	-	chr5	27260879	-	-
chr2	143838223	-	-	chr3	9713842	00000163719	MTMR14
chr3	9713842	00000163719	MTMR14	chr3	9713842	00000163719	MTMR14
chr2	145607374	00000226674	TEX41	chr9	114010661	00000227531	-
chr9	114010661	00000227531	-	chr9	114010661	00000227531	-
chr2	169507773	00000172292	CERS6	chr4	7883620	00000196526	AFAP1
chr4	7883620	00000196526	AFAP1	chr4	7883620	00000196526	AFAP1
chr2	180836562	00000163510	CWC22	chr2	186499434	-	-
chr2	186499434	-	-	chr2	186499434	-	-
chr2	191809745	00000115419	GLS	chr5	14147752	00000038382	TRIO
chr5	14147752	00000038382	TRIO	chr5	14147752	00000038382	TRIO
chr2	194037476	-	-	chr4	65780503	00000250125	LOC401134
chr4	65780503	00000250125	LOC401134	chr4	65780503	00000250125	LOC401134
chr2	195882358	00000235056	-	chr7	32316755	00000154678	PDE1C
chr7	32316755	00000154678	PDE1C	chr7	32316755	00000154678	PDE1C
chr21	10021329	-	-	chr4_gl000194_rand	167853	-	-
chr4_gl000194_rand	167853	-	-	chr4_gl000194_rand	167853	-	-
chr21	27972754	-	-	chrX	29688720	00000169306	IL1RAPL1
chrX	29688720	00000169306	IL1RAPL1	chrX	29688720	00000169306	IL1RAPL1
chr21	36570008	00000159216	LOC100506403	chr5	118771064	-	-

chr5	118771064	-	-	chr5	118771064	-	-
chr22	18663014	00000187979	-	chr22	21657364	00000206142	LOC100996335
chr22	21657364	00000206142	LOC100996335	chr22	21657364	00000206142	LOC100996335
chr22	18669176	00000187979	-	chr22	21663162	00000206142	LOC100996335
chr22	21663162	00000206142	LOC100996335	chr22	21663162	00000206142	LOC100996335
chr22	18799287	00000239989	-	chr22	21597017	-	-
chr22	21597017	-	-	chr22	21597017	-	-
chr22	18808872	00000239989	-	chr22	21609469	-	-
chr22	21609469	-	-	chr22	21609469	-	-
chr22	20356312	-	-	chr22	21551468	-	-
chr22	21551468	-	-	chr22	21551468	-	-
chr22	20357027	-	-	chr22	21551904	-	-
chr22	21551904	-	-	chr22	21551904	-	-
chr22	20489771	-	-	chr22	20656660	00000223579	-
chr22	20656660	00000223579	-	chr22	20656660	00000223579	-
chr3	9398065	00000206573	Thumpd3-as1	chr5	21269885	-	-
chr5	21269885	-	-	chr5	21269885	-	-
chr3	27556956	-	-	chr3	121706950	00000145103	ILDR1
chr3	121706950	00000145103	ILDR1	chr3	121706950	00000145103	ILDR1
chr3	44166352	-	-	chr8	142068934	00000253266	-
chr8	142068934	00000253266	-	chr8	142068934	00000253266	-
chr3	176974627	-	-	chrX	82733651	-	-
chrX	82733651	-	-	chrX	82733651	-	-
chr3	181544083	00000242808	SOX2-OT	chr4	163757977	-	-
chr4	163757977	-	-	chr4	163757977	-	-
chr3	189397739	00000073282	TP63	chr6	141792191	-	-
chr6	141792191	-	-	chr6	141792191	-	-
chr4	9245209	-	-	chr4	9363467	-	-
chr4	9363467	-	-	chr4	9363467	-	-
chr4	25934377	-	-	chrX	96524210	00000147202	DIAPH2
chrX	96524210	00000147202	DIAPH2	chrX	96524210	00000147202	DIAPH2
chr4	133369024	-	-	chr7	29422631	00000106069	LOC101928168
chr7	29422631	00000106069	LOC101928168	chr7	29422631	00000106069	LOC101928168
chr4	142606850	00000164136	IL15	chrX	75237794	-	-
chrX	75237794	-	-	chrX	75237794	-	-
chr4	172612207	-	-	chrX	150544694	-	-
chrX	150544694	-	-	chrX	150544694	-	-
chr4	181568227	-	-	chr6	118574987	00000196376	SLC35F1
chr6	118574987	00000196376	SLC35F1	chr6	118574987	00000196376	SLC35F1
chr5	8087558	-	-	chr9	69909824	-	-
chr9	69909824	-	-	chr9	69909824	-	-

chr5	21263916	-	-	chr5	36736289	-	-
chr5	36736289	-	-	chr5	36736289	-	-
chr5	50960623	-	-	chr6	77754221	-	-
chr6	77754221	-	-	chr6	77754221	-	-
chr5	60872349	-	-	chr8	131335003	00000153317	ASAP1
chr8	131335003	00000153317	ASAP1	chr8	131335003	00000153317	ASAP1
chr5	68867238	00000183474	GTF2H2C	chr5	69720244	00000226259	-
chr5	69720244	00000226259	-	chr5	69720244	00000226259	-
chr5	68901107	00000253203	GUSBP3	chr5	69752181	-	-
chr5	69752181	-	-	chr5	69752181	-	-
chr5	68909692	00000248477	-	chr5	69758342	00000250687	-
chr5	69758342	00000250687	-	chr5	69758342	00000250687	-
chr5	69064962	-	-	chr5	70643045	-	-
chr5	70643045	-	-	chr5	70643045	-	-
chr5	69093577	-	-	chr5	69967977	-	-
chr5	69967977	-	-	chr5	69967977	-	-
chr5	69113545	-	-	chr5	69982366	-	-
chr5	69982366	-	-	chr5	69982366	-	-
chr5	69202691	00000198237	-	chr5	70068029	00000253366	-
chr5	70068029	00000253366	-	chr5	70068029	00000253366	-
chr5	69252143	00000198237	-	chr5	69457686	00000254701	-
chr5	69457686	00000254701	-	chr5	69457686	00000254701	-
chr5	69298391	-	-	chr5	70170997	-	-
chr5	70170997	-	-	chr5	70170997	-	-
chr5	69346861	00000205571	smn2	chr5	70221487	00000172062	SMN1
chr5	70221487	00000172062	SMN1	chr5	70221487	00000172062	SMN1
chr5	69443493	00000254701	-	chr5	70443687	00000215630	-
chr5	70443687	00000215630	-	chr5	70443687	00000215630	-
chr5	69456980	00000254701	-	chr5	70457172	00000215630	-
chr5	70457172	00000215630	-	chr5	70457172	00000215630	-
chr5	69510945	00000251007	-	chr5	70510509	00000215630	-
chr5	70510509	00000215630	-	chr5	70510509	00000215630	-
chr5	102412707	-	-	chr6	147687245	00000164506	STXBP5
chr6	147687245	00000164506	STXBP5	chr6	147687245	00000164506	STXBP5
chr5	130556158	-	-	chr8	16282186	00000038945	MSR1
chr8	16282186	00000038945	MSR1	chr8	16282186	00000038945	MSR1
chr5	166673944	-	-	chr7	35030913	00000173852	DPY19L1
chr7	35030913	00000173852	DPY19L1	chr7	35030913	00000173852	DPY19L1
chr5	175515935	00000182230	LOC100507387	chr5	177461538	00000204677	FAM153C
chr5	177461538	00000204677	FAM153C	chr5	177461538	00000204677	FAM153C
chr6	8354378	-	-	chr6	159703675	-	-
chr6	159703675	-	-	chr6	159703675	-	-
chr6	26686507	00000261584	-	chr8	48773616	00000253729	PRKDC
chr8	48773616	00000253729	PRKDC	chr8	48773616	00000253729	PRKDC

chr6	57637678	-	-	chr7	61673547	-	-
chr7	61673547	-	-	chr7	61673547	-	-
chr6	76460099	00000196586	MYO6	chr8	144765103	-	-
chr8	144765103	-	-	chr8	144765103	-	-
chr6	153748543	-	-	chr6	153908166	-	-
chr6	153908166	-	-	chr6	153908166	-	-
chr7	8593783	00000122584	NXPH1	chrX	140293018	-	-
chrX	140293018	-	-	chrX	140293018	-	-
chr7	61288658	-	-	chr7	61414461	-	-
chr7	61414461	-	-	chr7	61414461	-	-
chr7	64673132	00000164669	INTS4P1	chr7	65183494	-	-
chr7	65183494	-	-	chr7	65183494	-	-
chr7	72653420	-	-	chr7	74206646	00000160828	-
chr7	74206646	00000160828	-	chr7	74206646	00000160828	-
chr7	75088512	00000242073	-	chr7	76270216	00000241350	-
chr7	76270216	00000241350	-	chr7	76270216	00000241350	-
chr7	76103688	00000091073	DTX2	chr7	76599296	00000241350	-
chr7	76599296	00000241350	-	chr7	76599296	00000241350	-
chr7	76112983	00000091073	DTX2	chr7	76609856	00000241350	-
chr7	76609856	00000241350	-	chr7	76609856	00000241350	-
chr7	84114378	00000075213	SEMA3A	chrX	100537397	00000102387	TAF7L
chrX	100537397	00000102387	TAF7L	chrX	100537397	00000102387	TAF7L
chr7	103446737	00000189056	RELN	chr7	139742540	00000059378	PARP12
chr7	139742540	00000059378	PARP12	chr7	139742540	00000059378	PARP12
chr7	112616560	00000234520	HRAT17	chr7	113133965	-	-
chr7	113133965	-	-	chr7	113133965	-	-
chr7	122594603	-	-	chrX	58482587	-	-
chrX	58482587	-	-	chrX	58482587	-	-
chr8	7464414	-	-	chr8	7912309	-	-
chr8	7912309	-	-	chr8	7912309	-	-
chr8	43822685	-	-	chr8	46845434	-	-
chr8	46845434	-	-	chr8	46845434	-	-
chr8	43823285	-	-	chr8	46841986	-	-
chr8	46841986	-	-	chr8	46841986	-	-
chr8	43826045	-	-	chr8	46850498	-	-
chr8	46850498	-	-	chr8	46850498	-	-
chr8	43829540	-	-	chr8	46842667	-	-
chr8	46842667	-	-	chr8	46842667	-	-
chr8	43834497	-	-	chr8	46848660	-	-
chr8	46848660	-	-	chr8	46848660	-	-
chr8	43836746	-	-	chr8	46850010	-	-
chr8	46849954	-	-	chr8	46850010	-	-
chr8	43836979	-	-	chr8	46854474	-	-
chr8	46854474	-	-	chr8	46854474	-	-

chr8	43837075	-	-	chr8	46845222	-	-
chr8	46845222	-	-	chr8	46845222	-	-
chr8	86569284	00000205176	REXO1L1P	chr8	86785509	-	-
chr8	86785509	-	-	chr8	86785509	-	-
chr8	86571192	00000205176	REXO1L1P	chr8	86788289	-	-
chr8	86788289	-	-	chr8	86788289	-	-
chr8	91757952	00000180694	TMEM64	chrX	145304886	-	-
chrX	145304886	-	-	chrX	145304886	-	-
chr8	112826486	-	-	chr8	126318343	00000156831	NSMCE2
chr8	126318343	00000156831	NSMCE2	chr8	126318343	00000156831	NSMCE2
chr8	143152990	-	-	chrX	150410436	-	-
chrX	150410436	-	-	chrX	150410436	-	-
chr9	38831076	00000228467	-	chr9	45191606	00000231838	-
chr9	45191606	00000231838	-	chr9	45191606	00000231838	-
chr9	39297539	-	-	chr9	41442648	-	-
chr9	41442648	-	-	chr9	41442648	-	-
chr9	40118851	-	-	chr9	65748777	-	-
chr9	65748777	-	-	chr9	65748777	-	-
chr9	40185847	-	-	chr9	65814689	-	-
chr9	65814689	-	-	chr9	65814689	-	-
chr9	41235569	-	-	chr9	43536514	-	-
chr9	43536514	-	-	chr9	43536514	-	-
chr9	41240372	-	-	chr9	43543668	-	-
chr9	43543668	-	-	chr9	43543668	-	-
chr9	41985146	00000204837	-	chr9	46807493	-	-
chr9	46807493	-	-	chr9	46807493	-	-
chr9	42199793	-	-	chr9	70594075	-	-
chr9	70594075	-	-	chr9	70594075	-	-
chr9	42227430	-	-	chr9	70622213	-	-
chr9	70622213	-	-	chr9	70622213	-	-
chr9	42280663	00000228136	-	chr9	70675837	-	-
chr9	70675837	-	-	chr9	70675837	-	-
chr9	42383036	00000223839	FAM95B1	chr9	67939455	00000196774	-
chr9	67939455	00000196774	-	chr9	67939455	00000196774	-
chr9	42476533	-	-	chr9	69481919	-	-
chr9	69481919	-	-	chr9	69481919	-	-
chr9	42526804	-	-	chr9	69534197	-	-
chr9	69534197	-	-	chr9	69534197	-	-
chr9	42556176	-	-	chr9	69561341	-	-
chr9	69561341	-	-	chr9	69561341	-	-
chr9	42596148	-	-	chr9	67240538	-	-
chr9	67240538	-	-	chr9	67240538	-	-
chr9	43858912	00000154529	CNTNAP3B	chr9	65599089	00000237792	-

chr9	65599089	00000237792	-	chr9	65599089	00000237792	-
chr9	44852214	-	-	chr9	46364486	00000231481	-
chr9	46364486	00000231481	-	chr9	46364486	00000231481	-
chr9	44990124	-	-	chr9	45726911	-	-
chr9	45726911	-	-	chr9	45726911	-	-
chr9	45693830	-	-	chr9	46073876	-	-
chr9	46073876	-	-	chr9	46073876	-	-
chr9	46435876	-	-	chr9	67836033	-	-
chr9	67836033	-	-	chr9	67836033	-	-
chr9	67297691	00000228522	-	chr9	69659389	00000176134	-
chr9	69659389	00000176134	-	chr9	69659389	00000176134	-
chr9	67348144	-	-	chr9	70096306	-	-
chr9	70096306	-	-	chr9	70096306	-	-
chr9	67358489	-	-	chr9	70104526	-	-
chr9	70104526	-	-	chr9	70104526	-	-
chr9	67957956	00000196774	-	chr9	69412744	00000172014	ANKRD20A4
chr9	69412744	00000172014	ANKRD20A4	chr9	69412744	00000172014	ANKRD20A4
chr9	68424758	-	-	chr9	69075556	00000233961	-
chr9	69075556	00000233961	-	chr9	69075556	00000233961	-
chr9	69019329	-	-	chrX	132163628	00000134588	USP26
chrX	132163628	00000134588	USP26	chrX	132163628	00000134588	USP26
chrX	58568809	-	-	chrX	61692295	-	-
chrX	61692295	-	-	chrX	61692295	-	-
chrX	58568895	-	-	chrX	61693384	-	-
chrX	61693384	-	-	chrX	61693384	-	-
chrX	58567363	-	-	chrX	61700903	-	-
chrX	61700909	-	-	chrX	61700903	-	-
chrX	58568112	-	-	chrX	61697695	-	-
chrX	61697695	-	-	chrX	61697695	-	-
chrX	58579352	-	-	chrX	61712921	-	-
chrX	61712921	-	-	chrX	61712921	-	-

Table 3-2. NCI-H292 structural variation calls identified by SVIM.

Breakpoint position 1				Breakpoint Position 2			
chrom	**position**	**Ensembl ID**	**Gene Name**	**chrom**	**position**	**Ensembl ID**	**Gene Name**
chr1	142535608	-	-	chrY	13138242	-	-
chr1	60838431	-	-	chr8	108568650	-	-
chr1	155103357	00000169242	EFNA1	chr7	54199259	-	-
chr1	180630725	00000143324	XPR1	chr6	78048640	-	-
chr1	22949174	-	-	chr5	13742278	00000039139	DNAH5
chr1	249239894	-	-	chr21	48119895	-	-
chr1	198475825	-	-	chr21	16128085	-	-
chr1	142850490	-	-	chr21	10204324	-	-
chr1	142855124	-	-	chr21	10202626	-	-
chr1	148860094	00000231448	-	chr2	91916129	-	-
chr1	148944120	00000232527	LOC100996732	chr2	91834688	00000143429	LOC645166
chr1	148942452	00000232527	LOC100996732	chr2	91833708	00000143429	LOC645166
chr1	121484428	-	-	chr19	27741719	-	-
chr1	121484167	-	-	chr19	27741714	-	-
chr1	121485342	-	-	chr19	27737686	-	-
chr1	121484056	-	-	chr19	27737567	-	-
chr1	121485432	-	-	chr19	27734494	-	-
chr1	121484285	-	-	chr19	27733872	-	-
chr1	121484056	-	-	chr19	27733558	-	-
chr1	121485432	-	-	chr19	27731809	-	-
chr1	121484778	-	-	chr19	27731799	-	-
chr1	76264	-	-	chr19	117518	-	-
chr1	250306	00000241670	-	chr16	90269211	00000260923	-
chr1	23603854	-	-	chr11	45334935	-	-
chr1	121484802	-	-	chr10	42546685	-	-
chr1	121484056	-	-	chr10	42404645	-	-
chr1	144057042	00000196369	SRGAP2B	chr1	206514998	-	-
chr1	206514998	-	-	chr1	144057042	00000196369	SRGAP2B
chr10	42411058	-	-	chr3	93507472	-	-
chr10	42398500	-	-	chr19	27738962	-	-
chr10	42527030	-	-	chr19	27736765	-	-
chr10	42397783	-	-	chr19	27736304	-	-
chr10	42405137	-	-	chr19	27735916	-	-
chr10	42401176	-	-	chr19	27731809	-	-
chr10	42406088	-	-	chr19	27731809	-	-
chr10	42533994	-	-	chr19	27731799	-	-
chr10	13650683	00000239665	-	chr17	37778462	00000214546	-
chr10	42411068	-	-	chr12	38106883	-	-
chr10	91561818	-	-	chr11	41093824	00000148948	LRRC4C
chr10	42546685	-	-	chr1	121484802	-	-
chr10	42404645	-	-	chr1	121484056	-	-
chr11	48949259	-	-	chrX	61763324	-	-

chr11	57180268	00000254979	-
chr11	48340096	-	-
chr11	49662300	00000205035	LOC440040
chr11	50653596	-	-
chr11	76330371	-	-
chr11	103957136	00000170962	PDGFD
chr11	50534442	-	-
chr11	48931292	-	-
chr11	54737682	-	-
chr11	41093824	00000148948	LRRC4C
chr11	45334935	-	-
chr12	34851548	-	-
chr12	38106883	-	-
chr13	28334329	-	-
chr13	114217523	-	-
chr13	82144334	-	-
chr13	70852666	-	-
chr14	90911815	-	-
chr14	88748867	00000100433	KCNK10
chr14	41873236	-	-
chr15	29853214	00000104059	FAM189A1
chr15	40854200	00000128891	c15orf57
chr15	26138155	-	-
chr15	20501532	-	-
chr15	82748439	00000237550	-
chr15	83124918	00000230903	-
chr15	34712554	00000215252	GOLGA8B
chr15	34858659	00000215252	GOLGA8B
chr15	20788149	-	-
chr15	21115392	-	-
chr15	21097491	-	-
chr15	22126547	-	-
chr15	22108615	-	-
chr15	23454162	-	-
chr16	33971797	-	-
chr16	33913312	-	-
chr16	33986804	-	-
chr16	33972877	-	-
chr16	90194368	00000260528	-
chr16	32515146	-	-
chr16	34185727	-	-
chr16	35231938	-	-
chr16	59118770	00000245768	-
chr16	18212948	00000259929	-
chr7	66393177	00000106609	TMEM248
chr7	64476899	-	-
chr5	133776813	00000250564	LOC102546229
chr4	52683402	-	-
chr22	41690861	-	-
chr2	222752684	-	-
chr19	27811909	-	-
chr19	27740289	-	-
chr16	32524565	-	-
chr10	91561818	-	-
chr1	23603854	-	-
chr19	27737065	-	-
chr10	42411068	-	-
chr8	43827763	-	-
chr3	95328600	-	-
chr3	90451259	-	-
chr22	27473939	-	-
chrY	13718798	-	-
chr7	132296869	00000221866	PLXNA4
chr17	19018894	-	-
chr8	39509957	00000168619	ADAM18
chr6	116774818	-	-
chr22	22878385	-	-
chr16	33359316	-	-
chr15	83124918	00000230903	-
chr15	82748439	00000237550	-
chr15	34858659	00000215252	GOLGA8B
chr15	34712554	00000215252	GOLGA8B
chr15	23454162	-	-
chr15	22126547	-	-
chr15	22108615	-	-
chr15	21115392	-	-
chr15	21097491	-	-
chr15	20788149	-	-
chrY	9925426	-	-
chrY	10068519	-	-
chrY	10048325	-	-
chrY	10045023	-	-
chr9	141129429	00000233013	-
chr7	61845692	-	-
chr7	61759180	-	-
chr6	74627228	-	-
chr4	155078196	-	-
chr4	152829264	00000249818	-

chr16	35231939	-	-	chr21	10772816	-	-
chr16	33977014	-	-	chr21	10704895	-	-
chr16	33976280	-	-	chr21	10704147	-	-
chr16	33971041	-	-	chr21	10701882	-	-
chr16	46396621	-	-	chr16	46405862	-	-
chr16	46396979	-	-	chr16	46404295	-	-
chr16	46396774	-	-	chr16	46402838	-	-
chr16	46390738	-	-	chr16	46396979	-	-
chr16	46404295	-	-	chr16	46396979	-	-
chr16	46390738	-	-	chr16	46396816	-	-
chr16	46402838	-	-	chr16	46396774	-	-
chr16	46405862	-	-	chr16	46396621	-	-
chr16	46396816	-	-	chr16	46390738	-	-
chr16	46396979	-	-	chr16	46390738	-	-
chr16	29480062	00000271699	-	chr16	30219879	00000258130	LOC613038
chr16	29479675	00000271699	-	chr16	30216275	00000258130	LOC613038
chr16	30219879	00000258130	LOC613038	chr16	29480062	00000271699	-
chr16	30216275	00000258130	LOC613038	chr16	29479675	00000271699	-
chr16	21478816	00000258186	-	chr16	21909020	00000185710	-
chr16	21909020	00000185710	-	chr16	21478816	00000258186	-
chr16	33359316	-	-	chr15	20501532	-	-
chr16	32524565	-	-	chr11	54737682	-	-
chr16	90269211	00000260923	-	chr1	250306	00000241670	-
chr17	25263013	-	-	chrY	28788343	-	-
chr17	12154158	-	-	chr9	92443417	-	-
chr17	41464006	00000188825	LINC00910	chr6	142366513	-	-
chr17	21685667	-	-	chr21	10795105	-	-
chr17	41401193	-	-	chr17	41466850	-	-
chr17	41466850	-	-	chr17	41401193	-	-
chr17	19018894	-	-	chr14	41873236	-	-
chr17	37778462	00000214546	-	chr10	13650683	00000239665	-
chr18	18518840	-	-	chrY	10073694	-	-
chr18	18519857	-	-	chr6	58778428	-	-
chr18	25145575	00000264151	LOC105372038	chr5	41484850	00000182836	PLCXD3
chr18	18518055	-	-	chr19	27811909	-	-
chr19	22414379	-	-	chr7	4625241	-	-
chr19	27736598	-	-	chr5	110286693	-	-
chr19	27731799	-	-	chr4	68266276	-	-
chr19	27731809	-	-	chr4	68266272	-	-
chr19	27731809	-	-	chr4	68265603	-	-
chr19	27731809	-	-	chr4	68265260	-	-
chr19	30129739	-	-	chr20	37585278	-	-
chr19	18839629	00000105662	CRTC1	chr19	18839632	00000105662	CRTC1
chr19	18839632	00000105662	CRTC1	chr19	18839629	00000105662	CRTC1

chr19	27811909	-	-	chr18	18518055	-	-
chr19	27737065	-	-	chr12	34851548	-	-
chr19	27811909	-	-	chr11	50534442	-	-
chr19	27740289	-	-	chr11	48931292	-	-
chr19	27731799	-	-	chr10	42533994	-	-
chr19	27736765	-	-	chr10	42527030	-	-
chr19	27731809	-	-	chr10	42406088	-	-
chr19	27735916	-	-	chr10	42405137	-	-
chr19	27731809	-	-	chr10	42401176	-	-
chr19	27738962	-	-	chr10	42398500	-	-
chr19	27736304	-	-	chr10	42397783	-	-
chr19	117518	-	-	chr1	76264	-	-
chr19	27731809	-	-	chr1	121485432	-	-
chr19	27734494	-	-	chr1	121485432	-	-
chr19	27737686	-	-	chr1	121485342	-	-
chr19	27731799	-	-	chr1	121484778	-	-
chr19	27741719	-	-	chr1	121484428	-	-
chr19	27733872	-	-	chr1	121484285	-	-
chr19	27741714	-	-	chr1	121484167	-	-
chr19	27733558	-	-	chr1	121484056	-	-
chr19	27737567	-	-	chr1	121484056	-	-
chr2	210821030	00000144406	UNC80	chrX	117656437	00000147251	DOCK11
chr2	210821032	00000144406	UNC80	chrX	117656342	00000147251	DOCK11
chr2	114249604	00000136682	CBWD2	chr9	122678	00000172785	CBWD1
chr2	160091164	-	-	chr6	58776720	-	-
chr2	96902900	00000204685	-	chr3	57799675	00000163681	SLMAP
chr2	114363120	-	-	chr21	48116982	-	-
chr2	92321178	-	-	chr2	92325606	-	-
chr2	92318904	-	-	chr2	92321331	-	-
chr2	92325606	-	-	chr2	92321178	-	-
chr2	92316994	-	-	chr2	92320444	-	-
chr2	92321331	-	-	chr2	92318904	-	-
chr2	92320444	-	-	chr2	92316994	-	-
chr2	90373130	-	-	chr2	91600387	-	-
chr2	90390042	-	-	chr2	91597020	-	-
chr2	91597020	-	-	chr2	90390042	-	-
chr2	90381235	-	-	chr2	90387368	-	-
chr2	90378887	-	-	chr2	90385591	-	-
chr2	90387368	-	-	chr2	90381235	-	-
chr2	90385591	-	-	chr2	90378887	-	-
chr2	91600387	-	-	chr2	90373130	-	-
chr2	222752684	-	-	chr11	103957136	00000170962	PDGFD
chr2	91834688	00000143429	LOC645166	chr1	148944120	00000232527	LOC100996732
chr2	91833708	00000143429	LOC645166	chr1	148942452	00000232527	LOC100996732

chr2	91916129	-	-
chr20	52484365	-	-
chr20	52484365	-	-
chr20	41339599	00000196090	PTPRT
chr20	56416571	-	-
chr20	54123235	-	-
chr20	54123237	-	-
chr20	26319562	-	-
chr20	33696667	-	-
chr20	25893026	-	-
chr20	25888637	-	-
chr20	29441081	-	-
chr20	29438008	-	-
chr20	37585278	-	-
chr21	10773722	-	-
chr21	10703713	-	-
chr21	19906306	-	-
chr21	9668094	-	-
chr21	48116982	-	-
chr21	10795105	-	-
chr21	10772816	-	-
chr21	10704895	-	-
chr21	10704147	-	-
chr21	10701882	-	-
chr21	48119895	-	-
chr21	16128085	-	-
chr21	10202626	-	-
chr21	10204324	-	-
chr22	18814125	00000239989	-
chr22	21615340	-	-
chr22	22878385	-	-
chr22	27473939	-	-
chr22	41690861	-	-
chr3	197883135	00000236438	-
chr3	44536586	-	-
chr3	48533428	00000164054	SHISA5
chr3	90462856	-	-
chr3	57799675	00000163681	SLMAP
chr3	90451259	-	-
chr3	95328600	-	-
chr3	93507472	-	-
chr4	49183671	-	-
chr4	125965193	-	-
chr4	125965207	-	-
chr1	148860094	00000231448	-
chrX	978563	-	-
chrX	978563	-	-
chr9	91702583	00000148082	SHC3
chr6	74333518	00000119899	SLC17A5
chr4	125965207	-	-
chr4	125965193	-	-
chr3	90462856	-	-
chr3	48533428	00000164054	SHISA5
chr20	29441081	-	-
chr20	29438008	-	-
chr20	25893026	-	-
chr20	25888637	-	-
chr19	30129739	-	-
chrY	28808531	-	-
chrX	58476239	-	-
chr5	38710878	00000249911	-
chr4	49183671	-	-
chr2	114363120	-	-
chr17	21685667	-	-
chr16	35231939	-	-
chr16	33977014	-	-
chr16	33976280	-	-
chr16	33971041	-	-
chr1	249239894	-	-
chr1	198475825	-	-
chr1	142855124	-	-
chr1	142850490	-	-
chr22	21615340	-	-
chr22	18814125	00000239989	-
chr15	26138155	-	-
chr13	70852666	-	-
chr11	76330371	-	-
chr7	31621	-	-
chr5	113162157	-	-
chr20	33696667	-	-
chr20	26319562	-	-
chr2	96902900	00000204685	-
chr13	82144334	-	-
chr13	114217523	-	-
chr10	42411058	-	-
chr21	9668094	-	-
chr20	54123237	-	-
chr20	54123235	-	-

chr4	68265260	-	-
chr4	68265603	-	-
chr4	68266272	-	-
chr4	68266276	-	-
chr4	155078196	-	-
chr4	152829264	00000249818	-
chr4	52683402	-	-
chr5	49486608	-	-
chr5	21523118	00000183666	GUSBP1
chr5	113162157	-	-
chr5	38710878	00000249911	-
chr5	110286693	-	-
chr5	41484850	00000182836	PLCXD3
chr5	133776813	00000250564	LOC102546229
chr5	13742278	00000039139	DNAH5
chr6	74333518	00000119899	SLC17A5
chr6	58776720	-	-
chr6	58778428	-	-
chr6	142366513	-	-
chr6	74627228	-	-
chr6	116774818	-	-
chr6	78048640	-	-
chr7	56708378	00000228303	-
chr7	108222558	-	-
chr7	61969423	-	-
chr7	61968990	-	-
chr7	61970355	-	-
chr7	61970221	-	-
chr7	61054339	-	-
chr7	61757850	-	-
chr7	31621	-	-
chr7	4625241	-	-
chr7	61759180	-	-
chr7	61845692	-	-
chr7	132296869	00000221866	PLXNA4
chr7	66393177	00000106609	TMEM248
chr7	64476899	-	-
chr7	54199259	-	-
chr8	43813848	-	-
chr8	86834497	-	-
chr8	46849770	-	-
chr8	46843620	-	-
chr8	46853313	-	-
chr8	43828051	-	-
chr19	27731809	-	-
chr19	27731809	-	-
chr19	27731809	-	-
chr19	27731799	-	-
chr16	59118770	00000245768	-
chr16	18212948	00000259929	-
chr11	50653596	-	-
chrX	61822081	-	-
chrX	41975552	-	-
chr3	44536586	-	-
chr21	19906306	-	-
chr19	27736598	-	-
chr18	25145575	00000264151	LOC105372038
chr11	49662300	00000205035	LOC440040
chr1	22949174	-	-
chr20	56416571	-	-
chr2	160091164	-	-
chr18	18519857	-	-
chr17	41464006	00000188825	LINC00910
chr16	35231938	-	-
chr15	40854200	00000128891	c15orf57
chr1	180630725	00000143324	XPR1
chrX	16567150	-	-
chr8	925376	00000237647	-
chr7	61970355	-	-
chr7	61970221	-	-
chr7	61969423	-	-
chr7	61968990	-	-
chr7	61757850	-	-
chr7	61054339	-	-
chr3	197883135	00000236438	-
chr19	22414379	-	-
chr16	34185727	-	-
chr16	32515146	-	-
chr14	88748867	00000100433	KCNK10
chr11	57180268	00000254979	-
chr11	48340096	-	-
chr1	155103357	00000169242	EFNA1
chrX	61731887	-	-
chrX	127281670	-	-
chr8	46853313	-	-
chr8	46851037	-	-
chr8	46849770	-	-
chr8	46849600	-	-

chr8	43830773	-	-	chr8	46844850	-	-
chr8	46851037	-	-	chr8	46843620	-	-
chr8	43823121	-	-	chr8	46842792	-	-
chr8	43837102	-	-	chr8	46842640	-	-
chr8	46842640	-	-	chr8	43837102	-	-
chr8	46844850	-	-	chr8	43830773	-	-
chr8	46849600	-	-	chr8	43828051	-	-
chr8	46842792	-	-	chr8	43823121	-	-
chr8	925376	00000237647	-	chr7	108222558	-	-
chr8	39509957	00000168619	ADAM18	chr15	29853214	00000104059	FAM189A1
chr8	43827763	-	-	chr13	28334329	-	-
chr8	108568650	-	-	chr1	60838431	-	-
chr9	39315963	-	-	chr9	39845179	-	-
chr9	39845179	-	-	chr9	39315963	-	-
chr9	91702583	00000148082	SHC3	chr20	41339599	00000196090	PTPRT
chr9	122678	00000172785	CBWD1	chr2	114249604	00000136682	CBWD2
chr9	92443417	-	-	chr17	12154158	-	-
chr9	141129429	00000233013	-	chr16	90194368	00000260528	-
chrX	58579674	-	-	chrX	61717587	-	-
chrX	58576660	-	-	chrX	61716669	-	-
chrX	58571792	-	-	chrX	61712928	-	-
chrX	58568749	-	-	chrX	61705855	-	-
chrX	58579973	-	-	chrX	61699205	-	-
chrX	61699205	-	-	chrX	58579973	-	-
chrX	61717587	-	-	chrX	58579674	-	-
chrX	61716669	-	-	chrX	58576660	-	-
chrX	61712928	-	-	chrX	58571792	-	-
chrX	61705855	-	-	chrX	58568749	-	-
chrX	23639132	-	-	chrX	23640882	-	-
chrX	23640882	-	-	chrX	23639132	-	-
chrX	127281670	-	-	chr8	86834497	-	-
chrX	61731887	-	-	chr8	43813848	-	-
chrX	16567150	-	-	chr7	56708378	00000228303	-
chrX	61822081	-	-	chr5	49486608	-	-
chrX	41975552	-	-	chr5	21523118	00000183666	GUSBP1
chrX	58476239	-	-	chr21	10703713	-	-
chrX	978563	-	-	chr20	52484365	-	-
chrX	978563	-	-	chr20	52484365	-	-
chrX	117656342	00000147251	DOCK11	chr2	210821032	00000144406	UNC80
chrX	117656437	00000147251	DOCK11	chr2	210821030	00000144406	UNC80
chrX	61763324	-	-	chr11	48949259	-	-
chrY	13446517	-	-	chrY	58975985	-	-
chrY	58975985	-	-	chrY	13446517	-	-
chrY	28808531	-	-	chr21	10773722	-	-

chrY	10073694	-	-	chr18	18518840	-	-
chrY	28788343	-	-	chr17	25263013	-	-
chrY	10048325	-	-	chr16	33986804	-	-
chrY	10045023	-	-	chr16	33972877	-	-
chrY	9925426	-	-	chr16	33971797	-	-
chrY	10068519	-	-	chr16	33913312	-	-
chrY	13718798	-	-	chr14	90911815	-	-
chrY	13138242	-	-	chr1	142535608	-	-

Table 3-3. NCI-H292 structural variation calls identified by NanoVar.

Quality Variable	H292	PDX-062	PDX-294	PDX-356
GEMs Detected	1,438,684	1,475,469	1,497,818	1,478,697
N50 Linked-Reads per Molecule (LPM)	42	17	11	14
Mean DNA per GEM	543,940 bp	574,662 bp	727,420 bp	393,535 bp
DNA in Molecules >20kb	93.40%	76.60%	48.90%	43.60%
DNA in Molecules >100kb	18.80%	5.64%	4.23%	5.14%
Number of Reads	913,409,176	975,184,272	1,018,363,956	1,086,096,424
Median Insert Size	348 bp	362 bp	362 bp	351 bp
Mean Depth	39.8 X	28.1 X	39.5 X	28.0 X
Zero Coverage	0.85%	1.02%	0.40%	0.18%
Mapped Reads	95.90%	65.80%	87.40%	60.50%
PCR Duplication	4.06%	5.81%	5.92%	7.17%
Q30 bases, Read 1	89.80%	88.70%	88.10%	87.40%
Q30 bases, Read 2	86.10%	86.30%	85.70%	84.20%
Large Structural Variant Calls	139	206	126	622
Short Deletion Calls	5,334	7,213	4,665	7,104
SNPs Phased	99.50%	97.90%	97.80%	95.40%
Longest Phase Block	39,295,952 bp	4,338,721 bp	3,891,048 bp	2,585,581 bp
N50 Phase Block	9,641,782 bp	367,151 bp	500,018 bp	284,315 bp

Table 3-4. Quality Control Variables for Linked Read Sequencing Data

Total and mapped reads from the genome sequencing data are shown for each sample. In addition, we show the details of the fragment length, quality, and mapping depth for each library.

Region 1 Chr	Region 1 Start	Region 1 Read Length	Region 2 Chr	Region 2 start	Region 2 direction	Region 2 size	Region 3 Chr	Region 3 start	Region 3 direction	Region 3 size
chr8	67760470	2634	chr13	36745247	-	2648				
chr8	67782484	20461	chr8	67799963	-	1446	chr8	67800009	+	360
chr8	67799963	1786	chr8	67782484	+	19629	chr8	67800009	+	360
chr8	67800009	362	chr8	67782484	+	19629	chr8	67799963	-	1446
chr8	67807012	9168	chr8	67809027	-	5640				
chr8	67809027	7229	chr8	67807012	+	8984				
chr8	67832398	7304	chr8	67832413	-	6305				
chr8	67832413	7172	chr8	67832398	+	7086				
chr11	95705179	18865	chr11	95705078	+	1889				
chr11	95819456	14860	chr19	18792066	+	2746				
chr11	95826428	7888	chr19	18794120	+	566				
chr11	95834466	7056	chr8	67527435	+	2887				
chr11	95834466	2781	chr8	67527498	+	8228				
chr11	95871762	6187	chr11	95871997	+	4271				
chr11	95871997	5874	chr11	95871762	-	5591				
chr11	95884509	6457	chr11	95884562	-	17181				
chr11	95884562	17712	chr11	95884509	+	5040				
chr11	95950857	6807	chr11	95950943	-	4866				
chr11	95950943	6507	chr11	95950857	+	6655				
chr11	96049796	1818	chr11	96049892	+	727	chr11	96050426	-	283
chr11	96049892	877	chr11	96049796	-	1742	chr11	96050426	-	283
chr11	96050426	305	chr11	96049796	-	1742	chr11	96049892	+	727
chr11	96062789	3435	chr11	96062907	-	3267				
chr11	96062907	3342	chr11	96062789	+	3040				
chr11	111548750	2494	chr11	111550908	-	241				
chr11	111550908	280	chr11	111548750	+	2405				
chr11	111634473	4767	chr11	111634502	-	4640				
chr11	111634502	4757	chr11	111634473	+	4579				
chr11	122458352	3026	chr13	103914292	-	7334	chr2	40643130	+	3885
chr19	18792066	2841	chr11	95819456	-	14394				
chr19	18794120	641	chr11	95826428	-	7138				
chr19	18795379	905	chr8	67776116	+	1372				
chr19	18795379	8992	chr8	67776960	+	524				

Table 3-5. Long Reads Containing Rearranged Genes Identified in the NCI-H292 Linked Read Analysis.

Reads were extracted from the BAM file, and split reads containing multiple chromosomes corresponding to the genomic rearrangements identified by linked read sequencing were used to identify the local breakpoint sequence structure of rearrangements.

Fusion gene	Primer Name	Sequence
MAML2	*M2-MYB-F1*	TATGCCTGTCGTGTGTGTGT
MYBL1	*M2-MYB-R1*	AGGCAAATTCCAAGTCGTTG
SGK3	*SGK-CRTC-F1*	GGGACTACAGGCACACACCT
CRTC1	*SGK-CRTC-R1*	CACTGGGAGGAACATCACCT
TERT	*TERT-PP-F1*	TGGAGGTTAGCCTCGTCTTG
PPP2R1B	*TERT-PP-R1*	GCAGACAGACACCACAGAGC
MAML2	*CRTC1-MAML2-F1*	CTCCTGCCTTGATCTCCCAA
CRTC1	*CRTC1-MAML2-R1*	GACCCCACAAAATCCATACCTG
der(7)	*der(7)-F*	GAGATGTTTTCTTGCTATGTTGC
der(7)	*der(7)-R*	TGGGATTGCTGTATGGTAGC
der(5)	*der(5)- F*	AGGCAACCAAAGCAAAAACA
der(5)	*der(5) - R*	GAGACTTCCCATAGCCCTCA

Table 3-6. Primers Used for Breakpoint Validation.

5’ to 3’ sequences are listed.

Cell Line	Total Reads
UM-HMC-1	74,955,058
UM-HMC-3A	61,391,720
UM-HMC-3B	58,720,436
NCI-H292	62,750,079

Table 3-7. RNAseq read depth

NCI-PDX-356		
Chromosome	**Nucleotide**	**Gene**
7	55,437,819	LANCL2
9	38,044,371	RP11613M

Table 3-8. High Confidence Genomic Rearrangement Structural Details from NCI-PDX-356.

Positional information of the breakpoint was presented.

Sample	Sample name	Total reads	%Mapped	Total unique reads	%Uniquely mapped
60325	MEC001T	163200708	99.93	139923553	85.73709925
60326	MEC001N	148528472	99.91	118949682	80.08544113
60331	MEC004T	270412592	99.95	235786372	87.19504157
60332	MEC004N	249731298	99.95	214467123	85.8791528
105666	MEC008T	276946756	99.91	231070310	83.43492205
105667	MEC009T	323081288	99.92	267151261	82.68855886
105668	MEC009N	247248698	99.92	209730411	84.82568875
105669	MEC033T	126042318	99.87	106329282	84.35998614
107251	MEC033N	121498850	99.89	103510331	85.19449443
945-CB-1	M5T	57043506	99.81	46541611	81.58967473
945-CB-2	M5N	125709540	99.88	107419697	85.45071201
945-CB-3	M6T	148448352	99.9	126934701	85.50765252
945-CB-4	M6N	105964106	99.87	91295260	86.15677841
945-CB-5	M7T	180534508	99.9	153891737	85.24228343
945-CB-6	M7N	151356068	99.91	128558817	84.93799997
945-CB-7	M9T	183190324	99.9	157262057	85.84626828
945-CB-8	M9N	84483074	99.71	66908117	79.19706733
945-CB-9	M10T	253773482	99.85	207357044	81.70950029
945-CB-10	M10N	94349202	99.78	76778159	81.37658546
945-CB-11	M17T	74845794	99.76	62070132	82.93068813
945-CB-12	M17N	82686292	99.79	67592531	81.74575176
945-CB-13	M23T	174448948	99.84	145141453	83.19995888
945-CB-14	M23N	111884882	99.88	95447629	85.30878104
945-CB-15	M43T	74515094	99.84	61841799	82.9923116
945-CB-16	M43N	156590872	99.87	133335175	85.14875312
945-CB-17	M114T	211656676	99.78	176560221	83.41821498
945-CB-18	M114N	126263562	99.88	106368600	84.24330687

Table 3-9. Quality Control Variables for Exome Sequencing Data.

Total and mapped read details were described for each sample.

Sample ID	Total Number of Sites	Number of Somatic Sites	%
M10	12440	313	2.52
M114	13049	39	0.3
M17	11390	162	1.42
M21		0	0
M23	13082	87	0.67
M43	11995	288	2.4
M5	10462	509	4.87
M6	13538	42	0.31
M7	15909	54	0.34
M9	11244	97	0.86
MEC001	18100	3606	19.92
MEC004	20587	241	1.17
MEC008	19282	325	1.69
MEC009	20676	189	0.91
MEC033	13183	341	2.59

Table 3-10. Microsatellite Instability Scores Derived from Exome Sequencing Data.

MSIsensor was used to derive the relative instability values.

Sample ID	Seq. ID	A1	A2	B1	B2	C1	C2	Reads	Objective
MEC008_T	105666	03:01	03:01	55:04	38:01	12:03	01:02	1646	1616.362
MEC033_T	105669	30:03	66:01	07:02	07:02	07:04	17:01	419	407.667
MEC033_N	107251	34:02	34:02	40:01	40:01	17:01	07:04	322	319.092
MEC001_T	60325	03:01	03:01	81:01	81:01	07:01	07:01	1186	1186
MEC001_N	60326	03:01	03:01	07:02	07:02	07:02	07:02	1885	1885
MEC004_N	60332	32:04	32:04	55:04	38:01	01:02	12:03	1987	1951.214
M17_T	945-CB-11	01:01	01:01	54:01	54:01	01:02	07:02	629	623.339
M23_N	945-CB-14	31:01	03:01	40:10	40:10	07:02	07:02	1524	1496.568
M43_T	945-CB-15	24:02	11:01			07:02	07:02	857	833.851
M114_T	945-CB-17	24:02	11:01			07:02	07:02	857	833.851
M114_N	945-CB-18	24:02	11:01			07:02	07:02	857	833.851
M6_N	945-CB-4	01:01	01:01	27:05	27:05	07:01	02:02	498	493.518
M7_N	945-CB-6	03:01	03:01	49:01	49:01	07:01	07:01	1433	1420.093
M9_T	945-CB-7	24:02	11:01			07:02	07:02	857	833.851
M9_N	945-CB-8	30:02	03:01	35:01	35:01	07:01	06:08	1104	1084.118
M10_T	945-CB-9	24:02	11:01			07:02	07:02	857	833.851

Table 3-11. Tumor HLA Sub-Type Derived from Exome Sequencing Data.

OptiType was used to define the HLA subtypes.

	Type of Mutation			
Tumor Sample ID	**LoF**	**Missense**	**Other**	**Total**
M10	4	5	11	**20**
M114	0	5	2	**7**
M17	2	25	9	**36**
M23	0	18	3	**21**
M43	3	30	20	**53**
M5	22	211	71	**304**
M6	2	9	5	**16**
M7	1	8	2	**11**
M9	2	11	3	**16**
MEC001	13	601	214	**828**
MEC004	0	5	3	**8**
MEC008	6	32	16	**54**
MEC009	1	7	5	**13**
MEC033	0	17	7	**24**
Total	56	984	371	**1411**

Table 3-12. Summary of all high-confidence single nucleotide variants from the WES cohort.

	Variant Type											
Tumor Sample ID	**3' UTR**	**5' UTR**	**Frame shift**	**In frame del.**	**In frame ins.**	**initiator codon**	**splice acceptor**	**splice donor**	**splice region**	**stop gain**	**stop lost**	**Tot.**
M10	1	0	5	1	0	0	0	0	5	0	0	**12**
M114	7	3	0	0	0	0	1	0	6	0	0	**17**
M17	5	1	4	2	2	0	1	0	15	0	0	**30**
M23	5	2	4	0	0	0	1	0	5	0	0	**17**
M43	8	2	4	3	1	0	0	0	17	0	0	**35**
M5	5	2	5	4	1	0	0	0	16	0	0	**33**
M6	3	0	4	1	0	0	1	0	11	0	0	**20**
M7	5	4	1	0	0	0	0	0	3	1	0	**14**
M9	6	3	6	1	0	0	1	0	13	0	0	**30**
MEC001	19	42	29	33	14	3	3	1	41	0	1	**186**
MEC004	3	1	3	1	0	0	0	0	9	0	0	**17**
MEC008	6	5	5	4	0	0	1	0	14	0	0	**35**
MEC009	3	2	1	0	2	0	0	0	4	0	0	**12**
MEC033	5	0	2	1	1	0	0	0	12	0	0	**21**
Total	81	67	73	51	21	3	9	1	171	1	1	**480**

Table 3-13. Summary of all high-confidence insertion and deletion (INDEL) variants from the WES cohort.

Chr.	Position	Strand	Ref. base	Alt. base	HUGO symbol	Seq. ontology	Protein seq. change	CHASM pval	CHASM FDR	ClinVar Patho-genic	dbSNP	INHIBITOR TARGET	UTR/Intron	UTR/ Intron Gene	VEST pval	VEST FDR	CGL driver class	Sample
chr1	207648270	+	C	T	CR2	SG	Q809*								0.0286			M10
chr2	108479169	+	T	G	RGPD4	MS	M746R	0.2232			rs577188947				0.0859			M10
chr22	20130891	+	G	A	ZDHHC8	MS	D580N	0.4146							0.0908			M10
chr17	34951525	+	T	G	DHRS11	MS	V91G	0.079							0.0696			M114
chr1	11561407	+	G	C	PTCHD2	MS	D120H	0.5992	0.95						0.0091	0.1		M17
chr2	66668576	+	C	T	MEIS1	MS	H153Y	0.0144	0.25						0.0215	0.1		M17
chr2	70409002	+	C	T	C2orf42	MS	G39E	0.2204	0.65						0.0016	0.05		M17
chr3	122283380	+	C	T	DTX3L	MS	S36L	0.1216	0.5				3' UTR	PARP9	0.0901	0.2		M17
chr5	31508765	+	C	T	DROSHA	MS	R517Q	0.033	0.3						0.0191	0.1		M17
chr5	134366980	+	C	T	PITX1	MS	E130K	0.3754	0.8	+	rs121909109				0.0005	0.05		M17
chr7	127961387	+	C	G	RBM28	MS	D499H	0.7432	0.95						0.0293	0.1		M17
chr8	101727714	+	G	A	PABPC1	MS	L207F	0.1516	0.6		rs200538577				0.0323	0.1		M17
chr11	6007377	+	G	C	OR52L1	MS	H262D	0.3666	0.8						0.0253	0.1		M17
chr12	32764143	+	G	C	FGD4	MS	E559Q	0.3612	0.8						0.0676	0.2		M17
chr13	25831918	+	G	A	MTMR6	MS	R309C	0.0064	0.25						0.0226	0.1		M17
chr15	72500978	+	G	A	PKM	MS	H274Y	0.0644	0.4						0.0911	0.2		M17
chr1	37271852	+	C	T	GRIK3	MS	E723K	0.1382	0.6		rs745837449				0.0256	0.15		M23
chr1	120491159	+	C	A	NOTCH2	MS	C877F	0.0598	0.6			Notch	2k upstr.	NOTCH2	0.0080	0.05	TSG	M23
chr7	146818164	+	G	A	CNTNAP2	MS	R283H	0.4244	0.75		rs764747425				0.0065	0.05		M23
chr17	36091738	+	T	C	HNF1B	MS	N298S	0.1312	0.6						0.0055	0.05		M23
chr1	16065173	+	G	A	SLC25A34	MS	D228N	0.1488	0.5		rs765192797				0.0022	0.1		M43
chr1	169698352	+	C	T	SELE	SG	W355*								0.0176	0.1		M43
chr3	31725129	+	C	T	OSBPL10	MS	E575K	0.1802	0.55						0.0104	0.1		M43
chr3	36756961	+	T	C	DCLK3	MS	Y602C	0.0076	0.4						0.0853	0.25		M43
chr3	74418351	+	A	T	CNTN3	MS	L312H	0.1606	0.5						0.0289	0.15		M43
chr4	100348985	+	C	T	ADH7	MS	C190Y	0.3814	0.75		rs760665074				0.0210	0.1		M43
chr7	82763606	+	C	T	PCLO	MS	C1087Y	0.6984	0.9						0.0089	0.1		M43
chr8	39678679	+	G	A	ADAM2	SG	Q119*								0.0206	0.1		M43
chr11	66330516	+	C	T	ACTN3	MS	T896I	0.2868	0.7				2k upstr.	CTSF	0.0574	0.2		M43
chr15	89073873	+	C	T	DET1	MS	R366Q	0.0864	0.4		rs201101298		2k upstr.	DET1	0.0334	0.15		M43
chr19	51015707	+	G	A	ASPDH	MS	P188L	0.2714	0.7				2k downstr.	JOSD2	0.0164	0.1		M43
chrX	123526153	+	G	A	TENM1	MS	R1813C	0.097	0.4		rs779162769				0.0452	0.15		M43
chr1	36786362	+	C	G	SH3D21	MS	L700V	0.4106	0.8				2k upstr.	EVA1B	0.0611	0.25		M5
chr1	54509198	+	G	A	TMEM59	MS	L131F	0.458	0.8				3' UTR	TMEM59	0.0519	0.25		M5
chr2	25057682	+	C	G	ADCY3	MS	E596Q	0.7514	0.95						0.0994	0.35		M5
chr2	74590202	+	G	A	DCTN1	MS	R1150C	0.128	0.6						0.0761	0.3		M5

chr2	108614413	+	G	A	SLC5A7	MS	V190I	0.3712	0.75		rs541031936				0.0601	0.25		M5
chr3	11858803	+	G	A	TAMM41	MS	R191W	0.0846	0.55		rs140891099				0.0008	0.1		M5
chr3	38043085	+	G	A	VILL	SS	_441_						intron	VILL	0.0428	0.2		M5
chr3	52476253	+	G	A	SEMA3G	MS	T136I	0.1414	0.6						0.0025	0.1		M5
chr3	108392968	+	T	C	DZIP3	MS	L878P	0.0502	0.55						0.0121	0.15		M5
chr3	142521078	+	C	T	TRPC1	MS	S550F	0.1044	0.55						0.0081	0.1		M5
chr3	167254687	+	G	A	WDR49	MS	A290V	0.3712	0.75		rs540397701		intron	WDR49	0.0536	0.25		M5
chr3	179080270	+	G	A	MFN1	MS	S179N	0.033	0.55						0.0080	0.1		M5
chr3	179123148	+	G	A	GNB4	MS	T249I	0.0982	0.55						0.0035	0.1		M5
chr4	25419792	+	C	T	ANAPC4	MS	R740C	0.0898	0.55						0.0143	0.15		M5
chr4	102117085	+	C	T	PPP3CA	MS	A83T	0.0898	0.55				intron	PPP3CA	0.0235	0.15		M5
chr5	39201985	+	G	A	FYB	MS	P370S	0.0742	0.55						0.0329	0.2		M5
chr5	68682260	+	C	T	RAD17	MS	L299F	0.2004	0.6						0.0126	0.15		M5
chr5	156482455	+	A	C	HAVCR1	MS	C46G	0.0702	0.55						0.0061	0.1		M5
chr6	56347503	+	C	T	DST	MS	R6807Q	0.4206	0.8						0.0392	0.2		M5
chr6	64421204	+	C	A	PHF3	MS	P1315T	0.0076	0.55						0.0082	0.1		M5
chr6	74520719	+	C	T	CD109	MS	T1184I	0.304	0.75						0.0139	0.15		M5
chr6	90660225	+	C	T	BACH2	MS	G534R	0.1606	0.6						0.0417	0.2		M5
chr7	21892211	+	G	A	DNAH11	MS	E3682K	0.3754	0.75						0.0551	0.25		M5
chr7	36464162	+	T	C	ANLN	MS	V871A	0.0198	0.55						0.0499	0.25		M5
chr7	82544905	+	G	A	PCLO	MS	R4133C	0.0764	0.55		rs372497499				0.0218	0.15		M5
chr7	82579786	+	C	T	PCLO	SG	W3373*								0.0236	0.15		M5
chr7	97483940	+	A	C	ASNS	MS	L397R	0.0484	0.55						0.0034	0.1		M5
chr7	105121548	+	C	T	PUS7	MS	G376R	0.232	0.65						0.0035	0.1		M5
chr7	134719423	+	G	A	AGBL3	MS	A361T	0.0164	0.55						0.0111	0.15		M5
chr8	87498841	+	G	A	RMDN1	MS	R123W	0.1594	0.6		rs753992602		intron	CPNE3	0.0009	0.1		M5
chr8	101717279	+	G	A	PABPC1	MS	R565W	0.0826	0.55		rs760348960				0.0248	0.15		M5
chr9	104433283	+	C	T	GRIN3A	MS	D471N	0.2436	0.7						0.0635	0.25		M5
chr9	107367619	+	C	T	OR13C2	MS	C97Y	0.0218	0.55						0.0382	0.2		M5
chr10	20335939	+	G	A	PLXDC2	MS	A156T	0.0812	0.55		rs770709704		intron	PLXDC2	0.0079	0.1		M5
chr10	120832561	+	G	A	EIF3A	MS	L128F	0.1044	0.55						0.0219	0.15		M5
chr11	46785030	+	C	T	CKAP5	SS	_1230_						intron	CKAP5	0.0086	0.1		M5
chr11	125476349	+	G	A	STT3A	MS	V257M	0.11	0.55						0.0345	0.2		M5
chr12	101680268	+	G	A	UTP20	MS	D166N	0.6552	0.95						0.0883	0.3		M5
chr12	112492193	+	C	T	NAA25	MS	G543S	0.3512	0.75						0.0293	0.2		M5
chr12	117271714	+	G	A	RNFT2	MS	G334R	0.6254	0.9		rs774196399		intron	RNFT2	0.0056	0.1		M5
chr12	122966492	+	C	T	ZCCHC8	MS	S292N	0.168	0.6						0.0790	0.3		M5
chr13	24233354	+	G	A	TNFRSF19	SS	_204_				rs758430355		intron	TNFRSF19	0.0182	0.15		M5

chr13	50051154	+	G	A	SETDB2	MS	C295Y	0.026	0.55		rs779473157				0.0030	0.1		M5
chr14	60976528	+	C	T	SIX6	MS	R138C	0.1802	0.6						0.0686	0.3		M5
chr14	75136708	+	C	T	AREL1	MS	R577Q	0.1348	0.6		rs376144017				0.0118	0.15		M5
chr15	59373484	+	G	A	RNF111	SS	_766_						intron	RNF111	0.0182	0.15		M5
chr15	64508900	+	G	A	CSNK1G1	MS	P102L	0.0144	0.55						0.0084	0.1		M5
chr16	1706151	+	C	T	CRAMP1	MS	R465W	0.3084	0.75		rs767330217				0.0911	0.35		M5
chr16	53529062	+	G	A	AKTIP	MS	R113W	0.0846	0.55		rs769347114				0.0158	0.15		M5
chr16	56438824	+	C	T	AMFR	MS	M279I	0.3084	0.75						0.0054	0.1		M5
chr16	58077226	+	G	A	MMP15	SG	W472*								0.0065	0.1		M5
chr16	69184446	+	G	C	CIRH1A	MS	D249H	0.7838	1						0.0805	0.3		M5
chr17	6499480	+	C	T	KIAA0753	MS	E786K	0.458	0.8						0.0726	0.3		M5
chr17	45219780	+	A	T	CDC27	MS	M404K	0.0764	0.55						0.0164	0.15		M5
chr17	45219792	+	T	C	CDC27	MS	K400R	0.1474	0.6		rs78399615				0.0872	0.3		M5
chr17	45219794	+	G	C	CDC27	MS	S399R	0.0764	0.55		rs75040309				0.0117	0.15		M5
chr17	45234625	+	C	A	CDC27	MS	V201F	0.0702	0.55		rs774144042				0.0707	0.3		M5
chr17	48434425	+	C	T	XYLT2	MS	P585S	0.0372	0.55		rs770825174				0.0827	0.3		M5
chr18	756613	+	C	T	YES1	MS	G77E	0.097	0.55						0.0239	0.15		M5
chr18	66344411	+	G	A	TMX3	MS	P375L	0.023	0.55						0.0160	0.15		M5
chr20	33470747	+	G	A	ACSS2	MS	R110Q	0.088	0.55		rs757812388				0.0045	0.1		M5
chr20	51872525	+	G	A	TSHZ2	MS	G843D	0.0086	0.55						0.0045	0.1		M5
chr20	57246343	+	T	A	STX16	MS	I261N	0.053	0.55						0.0062	0.1		M5
chr21	30332838	+	C	T	LTN1	SS	_831_						intron	LTN1	0.0257	0.15		M5
chr22	19029394	+	G	A	DGCR2	MS	S362L	0.102	0.55						0.0147	0.15		M5
chrX	10188916	+	G	A	CLCN4	MS	G731R	0.0884	0.55						0.0126	0.15		M5
chrX	15349908	+	C	T	PIGA	MS	V49M	0.2348	0.65				intron	PIGA	0.0201	0.15		M5
chrX	23410816	+	C	T	PTCHD1	MS	T394M	0.0356	0.55						0.0515	0.25		M5
chrX	53610679	+	G	A	HUWE1	MS	R1787W	0.0114	0.55		rs782600084				0.0153	0.15		M5
chr1	27271965	+	G	A	NUDC	SS	_276_						intron	NUDC	0.0160	0.1		M6
chr5	150498926	+	A	G	ANXA6	MS	L458P	0.0064	0.2						0.0012	0.05		M6
chr11	78574079	+	C	T	TENM4	MS	D395N	0.1626	0.9		rs753490323				0.0668	0.3		M6
chr2	201853134	+	G	A	FAM126B	MS	A281V	0.1554							0.0028			M7
chr5	74842906	+	G	A	POLK	MS	G20E	0.3612							0.0853			M7
chr6	138619807	+	C	G	ARFGEF3	MS	T1238R	0.1118							0.0191	0.1		M9
chr8	77767958	+	C	T	ZFHX4	MS	T2934M	0.0682							0.0080	0.1		M9
chr11	6412780	+	T	C	SMPD1	MS	L162P	0.1554							0.0345	0.15		M9
chr1	11985396	+	A	G	KIAA2013	MS	L300P	0.088	0.9						0.0536	1		MEC001
chr3	42264721	+	C	T	TRAK1	MS	S785L	0.026	0.9		rs201913227				0.0304	1		MEC001
chr5	1293767	+	G	A	TERT	MS	H412Y	0.3482	0.9		rs34094720				0.0927	1		MEC001

chr5	56205489	+	T	C	SETD9	MS	L6P	0.168	0.9									0.0161	1		MEC001
chr5	178413163	+	G	C	GRM6	MS	L698V	0.2114	0.9		rs62638623							0.0245	1		MEC001
chr7	73184318	+	A	G	CLDN3	MS	I21T	0.196	0.9		rs113016282							0.0585	1		MEC001
chr9	140008443	+	G	A	DPP7	MS	S120F	0.2004	0.9		rs150267769							0.0431	1		MEC001
chr10	87379753	+	G	A	GRID1	MS	A744V	0.4906	0.95									0.0750	1		MEC001
chr11	88911186	+	G	A	TYR	MS	R22K	0.4438	0.9									0.0219	1		MEC001
chr12	51086762	+	A	G	DIP2B	MS	K562R	0.2786	0.9									0.0560	1		MEC001
chr13	114088074	+	G	A	ADPRHL1	MS	T163I	0.0682	0.9		rs139075628							0.0005	0.15		MEC001
chr15	65369947	+	G	A	KBTBD13	MS	G265D	0.0232	0.9		rs146917406				2k downstr.	RASL12		0.0025	0.5		MEC001
chr16	1396337	+	C	T	BAIAP3	MS	T858M	0.1802	0.9		rs116326490							0.0927	1		MEC001
chr16	67234674	+	G	A	ELMO3	MS	E263K	0.3122	0.9		rs764620212				2k downstr.	E2F4		0.0271	1		MEC001
chr17	18141081	+	T	C	LLGL1	MS	L633P	0.11	0.9		rs140611231				intron	LLGL1		0.0062	0.95		MEC001
chr17	64876751	+	G	A	CACNG5	MS	G121R	0.3482	0.9		rs201783473							0.0005	0.15		MEC001
chr19	1047537	+	A	C	ABCA7	MS	N718T	0.2996	0.9		rs3752239							0.0805	1		MEC001
chr19	7680165	+	T	C	CAMSAP3	MS	W950R	0.0682	0.9									0.0955	1		MEC001
chr20	6033004	+	G	A	LRRN4	MS	L148F	0.5992	0.95		rs6117050							0.0823	1		MEC001
chr20	60884427	+	C	T	LAMA5	MS	G3685R	0.5332	0.95		rs138468519				2k downstr.	ADRM1		0.0168	1		MEC001
chr21	46950811	+	C	A	SLC19A1	MS	V342F	0.731	1		rs142899279							0.0635	1		MEC001
chr1	47078924	+	C	T	MOB3C	MS	G76S	0.088	1		rs754376722							0.0410	0.25		MEC008
chr2	65237847	+	C	A	SLC1A4	MS	F250L	0.582	1									0.0920	0.3		MEC008
chr2	213403252	+	C	T	ERBB4	MS	M1I	0.53	1				Lapatinib					0.0169	0.15		MEC008
chr3	30648460	+	C	T	TGFBR2	SG	Q29*											0.0826	0.3		MEC008
chr4	1803568	+	C	G	FGFR3	MS	S249C	0.038	1	+	rs121913483	FGFR						0.0040	0.1	Oncogene	MEC008
chr7	66264309	+	G	C	RABGEF1	SS	_257_								intron	RABGEF1		0.0086	0.15		MEC008
chr9	35698498	+	C	T	TLN1	MS	R2398Q	0.1312	1		rs556571189							0.0494	0.25		MEC008
chr10	105658711	+	C	G	OBFC1	MS	E169Q	0.454	1									0.0712	0.3		MEC008
chr13	21306082	+	C	G	N6AMT2	MS	D136H	0.717	1									0.0025	0.1		MEC008
chr17	74060176	+	C	T	SRP68	MS	E148K	0.2004	1									0.0471	0.25		MEC008
chr18	8783957	+	G	C	MTCL1	MS	E643Q	0.3754	1									0.0668	0.3		MEC008
chr20	13260384	+	G	C	ISM1	MS	W161S	0.0898	1									0.0124	0.15		MEC008
chr6	138619807	+	C	G	ARFGEF3	MS	T1238R	0.1118										0.0191			MEC009
chr8	77767958	+	C	T	ZFHX4	MS	T2934M	0.0682										0.0080			MEC009
chr11	6412780	+	T	C	SMPD1	MS	L162P	0.1554										0.0345			MEC009
chr9	136215814	+	C	T	RPL7A	MS	P14L	0.0884	0.3		rs61733976				5' UTR	RPL7A		0.0581	0.25		MEC033
chr10	126682443	+	G	T	CTBP2	MS	H838N	0.1444	0.35		rs76582415							0.0323	0.2		MEC033
chr10	126682454	+	G	A	CTBP2	MS	A834V	0.053	0.25									0.0045	0.05		MEC033
chr10	126682455	+	C	T	CTBP2	MS	A834T	0.0486	0.25									0.0028	0.05		MEC033
chrX	55039955	+	G	A	ALAS2	MS	P522S	0.2348	0.4									0.0943	0.35		MEC033

Table 3-14. CRAVAT functional annotation of all mutation calls from the exome cohort.

CHASM and VEST scores for each variant were derived to assist in the functional prediction of each alteration. Table is filtered to include variants where VEST p value<0.1.

Gene	MEC001	MEC004	MEC008	MEC009	MEC033	M5	M6	M7	M9	M10	M17	M23	M43	M114
A1CF	3.0	2.0	1.0	1.0	1.0	2.0	2.0	2.0	4.0	2.0	2.0	2.0	2.0	2.0
ABI1	3.0	2.0	1.0	1.0	1.0	2.0	2.0	2.0	4.0	2.0	2.0	2.0	2.0	2.0
ABL1	2.0	2.0	2.0	2.0	2.0	2.0	3.0	2.0	2.0	2.0	2.2	3.0	2.2	2.0
ABL2	2.0	2.0	1.6	1.0	2.0	3.0	2.0	2.0	4.0	2.0	2.0	2.0	2.0	2.0
ACSL3	2.0	2.0	2.0	1.3	2.0	2.3	2.0	2.0	4.0	2.0	2.0	2.0	2.0	2.0
ACSL6	2.0	2.0	2.0	2.0	2.0	2.0	2.0	3.0	3.0	2.0	2.0	2.9	2.0	2.0
ACVR1	3.0	2.0	2.0	1.0	1.0	2.0	2.0	2.0	4.0	2.0	2.0	2.0	2.0	2.0
ACVR1B	2.0	2.0	2.0	2.0	2.0	2.0	2.2	2.0	2.0	2.0	2.0	3.0	2.0	2.0
ACVR2A	2.0	2.0	2.0	1.0	1.2	1.0	1.1	2.0	4.0	2.0	2.0	2.0	1.1	2.0
AFF1	2.3	2.0	1.9	1.0	2.0	1.6	2.0	2.0	4.0	2.0	2.0	2.0	2.0	2.0
AFF4	2.0	2.0	1.0	2.0	1.4	2.0	2.0	3.0	3.0	2.0	2.0	2.0	2.0	2.0
AKAP9	3.0	2.0	1.0	1.0	1.7	1.3	1.0	2.3	4.0	2.0	2.0	2.0	1.0	2.0
AKT1	1.0	1.0	2.0	2.0	4.0	3.0	3.0	2.0	2.0	2.0	3.0	4.0	2.6	2.0
AKT2	1.0	1.0	2.0	2.0	3.0	3.0	3.0	2.0	2.0	2.0	2.0	3.0	3.0	2.0
AKT3	3.0	2.0	1.3	1.0	2.0	2.7	1.0	2.0	4.0	2.0	2.0	2.0	2.0	2.0
ALDH2	2.0	2.0	2.0	2.0	2.0	3.0	2.0	2.0	2.4	2.0	2.0	3.0	2.0	2.0
ALK	2.0	2.0	2.0	2.0	2.0	2.0	2.8	2.0	2.2	2.0	2.0	2.9	2.0	2.0
ANK1	2.0	2.0	2.0	2.8	3.0	3.1	2.9	2.0	3.0	2.0	1.0	3.0	2.8	2.0
APC	2.0	2.0	1.0	1.0	2.0	2.0	1.0	3.0	4.0	2.0	2.0	2.0	1.6	2.0
APOBEC3B	2.0	2.0	2.0	2.0	2.0	3.0	3.0	2.0	2.0	2.0	2.0	3.0	3.0	2.0
AR	1.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	3.8	2.0	2.0	2.0	1.6	2.0
ARAF	1.0	2.0	2.0	2.0	2.0	2.0	3.0	2.0	2.0	2.0	2.0	3.0	3.0	2.0
ARHGAP26	2.0	2.0	1.3	2.0	1.6	1.8	2.0	3.0	4.0	2.0	2.0	2.0	2.0	2.0
ARHGAP35	1.0	2.0	2.0	2.0	2.0	2.4	3.0	2.0	2.0	2.0	2.0	3.0	3.0	2.0
ARHGAP5	3.0	2.0	1.0	2.0	2.0	2.0	1.1	2.0	4.0	2.0	2.0	2.0	1.0	2.0
ARHGEF10	2.0	2.0	1.0	2.0	3.0	2.0	2.0	2.0	3.0	2.0	1.0	3.0	2.0	2.0
ARHGEF10L	1.0	2.0	2.0	2.0	2.0	2.9	3.0	2.0	2.0	2.0	2.8	4.0	2.0	2.0
ARHGEF12	2.0	2.0	3.0	1.8	2.0	2.0	2.0	2.0	3.9	2.0	1.0	2.0	2.0	2.0
ARID1A	2.0	2.0	2.0	2.0	2.0	2.0	3.0	2.0	2.0	2.0	2.0	3.0	2.2	2.0
ARID1B	3.0	2.0	2.0	2.0	2.0	1.4	2.0	2.0	3.0	2.0	2.0	2.0	2.0	2.0
ARID2	3.0	2.0	1.0	1.0	2.0	2.0	1.0	2.0	4.0	2.0	2.0	2.0	2.0	2.0
ARNT	2.0	2.0	2.0	2.0	2.0	3.0	2.0	2.0	3.0	2.0	2.0	2.0	2.1	2.0
ASPSCR1	1.0	1.0	2.0	2.0	3.0	3.0	3.0	1.0	2.0	2.0	2.0	4.0	3.0	2.0
ASXL1	2.0	2.0	2.0	1.6	2.0	2.0	2.0	2.0	2.0	2.0	2.0	3.0	2.0	2.0
ASXL2	2.0	2.0	2.0	2.0	1.3	2.0	3.0	2.0	3.0	2.0	2.0	3.0	2.0	2.0
ATIC	3.0	2.0	1.0	1.3	1.1	2.0	2.0	2.0	4.0	2.0	2.0	2.0	2.0	2.0
ATM	3.0	2.0	>5.0	1.0	1.5	1.0	1.0	2.0	4.0	2.0	1.0	2.0	2.0	2.0
ATP1A1	3.0	2.0	2.0	2.0	1.2	1.0	2.0	2.0	4.0	2.0	2.0	2.0	2.0	2.0
ATP2B3	1.0	1.0	2.0	2.0	3.0	2.7	3.0	2.0	2.0	2.0	2.0	4.0	3.0	2.0
ATR	3.0	2.0	1.0	1.0	1.3	1.2	1.0	2.5	4.0	2.0	2.0	2.0	1.2	2.0
ATRX	1.9	2.0	1.0	1.0	2.0	1.1	1.1	2.0	4.0	2.0	2.0	2.0	2.0	2.0
AXIN1	1.0	1.0	2.0	2.0	3.0	3.0	3.0	2.0	2.0	2.0	2.0	4.0	3.0	2.0
AXIN2	2.0	2.0	1.0	2.0	2.0	2.0	2.8	2.0	3.0	2.0	2.0	3.0	2.0	2.0
B2M	2.0	2.0	2.0	1.0	2.0	2.0	2.0	2.0	3.0	2.0	2.0	2.0	2.0	2.0
BAP1	1.0	1.0	2.0	2.0	2.0	1.0	3.0	1.0	2.0	2.0	2.0	3.3	2.6	2.0
BARD1	3.0	2.0	1.0	1.0	1.0	2.0	2.0	3.0	4.0	2.0	2.0	2.0	1.4	2.0
BAX	1.2	2.0	2.0	2.0	3.0	3.0	3.0	2.0	2.0	2.0	2.0	3.0	3.0	2.0
BAZ1A	3.0	2.0	1.6	1.0	3.0	2.0	1.0	2.0	4.0	2.0	2.0	2.0	2.0	2.0
BCL10	3.0	2.0	2.0	1.0	2.0	1.0	2.0	2.0	4.0	2.0	2.0	2.0	2.0	2.0
BCL11A	3.0	2.0	2.0	1.0	2.0	2.0	2.0	2.0	4.0	2.0	2.0	2.0	1.2	2.0
BCL11B	2.0	2.0	2.0	2.0	3.0	2.0	2.0	2.0	3.0	2.0	2.0	2.0	2.0	2.0
BCL2	3.0	2.0	1.0	1.0	2.0	2.0	2.0	2.0	4.0	2.0	2.0	2.0	2.0	2.0
BCL2L12	1.0	1.0	1.9	2.0	2.3	2.3	3.0	1.0	2.0	2.0	2.0	3.0	3.0	2.0
BCL3	1.0	1.0	2.0	2.0	3.0	3.0	3.0	2.0	2.0	2.0	2.0	3.0	3.0	2.0
BCL6	2.0	2.7	3.0	2.0	2.0	2.0	2.0	2.0	3.0	2.0	2.0	2.0	2.0	2.0
BCL7A	2.0	2.0	2.0	2.0	2.0	3.0	3.0	2.0	2.0	2.0	2.0	3.0	2.0	2.0
BCL9	2.0	2.0	2.0	2.0	2.0	3.0	2.0	2.0	3.0	2.0	2.0	2.0	2.0	2.0
BCL9L	2.0	2.0	3.0	2.0	2.0	2.0	2.0	2.0	3.0	2.0	1.0	3.0	3.0	2.0
BCLAF1	3.0	2.0	1.0	1.0	2.0	1.0	1.4	2.0	4.0	2.0	2.0	2.0	2.0	2.0
BCOR	1.0	2.0	1.0	2.0	2.0	2.0	2.0	2.0	4.0	2.0	2.0	2.0	2.0	2.0
BCORL1	1.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	3.0	2.0	2.0	3.0	2.0	2.0
BCR	1.0	2.0	2.0	2.0	2.0	2.2	3.0	2.0	2.0	2.0	2.0	4.0	2.1	2.0
BIRC3	2.3	2.0	3.0	1.0	2.0	1.0	1.0	2.0	4.0	2.0	1.0	2.0	1.8	2.0
BIRC6	2.0	2.0	1.0	1.0	2.0	2.0	2.0	2.0	4.0	2.0	2.0	2.0	2.0	2.0
BLM	2.9	2.0	2.0	1.1	1.0	1.0	1.3	2.9	3.0	2.0	2.0	2.2	2.0	2.0
BMP5	3.0	2.0	2.0	1.0	1.8	1.3	2.0	2.8	4.0	2.0	2.0	2.0	1.0	2.0
BMPR1A	2.0	2.0	2.0	1.3	1.0	2.0	2.0	2.0	3.0	2.0	2.0	3.0	2.0	2.0

BRAF	2.0	2.0	1.0	1.0	1.3	2.0	1.2	2.3	4.0	2.0	2.0	2.0	>5.0	2.0
BRCA1	2.0	2.0	2.0	1.2	1.2	2.0	2.2	2.0	3.7	2.0	1.1	2.4	>5.0	2.0
BRCA2	3.0	3.0	3.0	1.0	2.0	2.0	1.0	2.0	4.0	2.0	2.0	2.0	1.9	2.0
BRD3	2.0	1.0	1.0	2.0	3.0	2.0	3.0	2.0	2.0	2.0	2.0	4.0	2.0	2.0
BRIP1	3.0	2.0	1.0	1.0	1.0	2.6	1.0	3.0	4.0	2.0	2.0	2.0	1.0	2.0
BTG1	3.0	2.0	1.0	1.0	2.0	2.0	1.0	2.0	4.0	2.0	2.0	2.0	2.0	2.0
BTG2	2.0	2.0	2.0	2.0	2.0	3.0	3.0	2.0	2.0	2.0	2.0	3.0	2.0	2.0
BTK	1.0	2.0	2.0	1.7	2.0	2.0	2.0	2.0	4.0	2.0	1.3	2.0	2.0	2.0
BUB1B	2.0	2.0	2.0	1.0	2.0	2.0	2.0	3.0	4.0	2.0	2.0	2.0	2.0	2.0
CACNA1D	2.0	2.0	2.0	2.0	2.0	2.0	2.6	2.0	3.0	2.0	2.0	3.0	2.0	2.0
CALR	1.0	1.0	1.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	4.0	3.0	2.0
CAMTA1	2.0	2.0	2.0	2.0	2.0	2.0	2.8	2.0	3.0	2.0	2.0	3.0	2.0	2.0
CANT1	1.0	1.0	2.0	2.0	2.0	3.0	3.0	2.0	2.0	2.0	2.0	3.0	3.0	2.0
CARD11	1.0	2.0	2.0	2.0	3.0	3.0	3.0	2.0	2.0	2.0	2.0	3.9	3.0	2.0
CARS	1.1	1.2	2.0	3.0	2.0	2.0	3.0	2.0	3.0	2.0	2.0	3.0	2.0	2.0
CASP3	2.0	1.0	1.0	1.0	2.0	2.0	2.0	2.0	4.0	2.0	2.0	2.0	2.0	2.0
CASP8	2.0	2.0	1.0	1.0	2.0	2.0	2.0	1.7	4.0	2.0	1.4	2.0	2.1	2.0
CASP9	2.0	2.0	2.0	2.0	2.0	2.0	3.0	2.0	2.0	2.0	2.0	3.0	2.0	2.0
CBFA2T3	1.0	1.0	2.0	2.0	3.0	3.0	3.0	2.0	2.0	2.0	2.0	4.0	3.0	2.0
CBFB	2.0	2.0	2.0	1.3	2.0	2.0	2.0	2.0	2.0	2.0	2.0	3.0	2.3	2.0
CBL	2.0	2.0	3.0	1.9	2.0	2.0	>5.0	2.0	3.0	2.0	1.0	>5.0	3.0	2.0
CBLB	3.0	2.0	1.0	1.0	2.0	1.0	1.3	2.0	4.0	2.0	2.0	2.0	1.9	2.0
CBLC	1.0	1.0	2.0	2.0	3.0	3.0	3.0	2.0	2.0	2.0	2.0	3.0	3.0	2.0
CCDC6	3.0	2.0	2.0	1.0	1.0	2.0	2.0	2.0	4.0	2.0	2.0	2.0	1.0	2.0
CCNB1IP1	2.0	2.0	1.0	2.0	2.0	2.0	2.0	2.0	3.0	2.0	2.0	2.0	2.0	2.0
CCNC	3.0	2.0	1.0	1.0	2.0	1.5	1.1	2.0	4.0	2.0	2.0	2.0	2.0	2.0
CCND1	1.0	1.0	3.0	3.0	3.0	2.0	3.0	2.0	3.0	2.0	2.0	3.0	2.0	2.0
CCND3	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	3.0	3.0	2.0
CCNE1	2.0	2.0	2.0	2.0	2.0	3.0	3.0	2.0	3.0	2.0	2.0	3.0	2.4	2.0
CCR4	2.0	2.0	1.0	2.0	2.0	2.0	2.0	2.0	3.0	2.0	2.0	2.0	2.0	2.0
CCR7	2.0	2.0	2.0	2.0	2.0	2.0	3.0	2.0	3.0	2.0	2.0	3.0	2.0	2.0
CD209	1.0	2.0	2.0	2.0	3.0	3.0	3.0	2.0	2.0	2.0	2.0	4.0	3.0	2.0
CD274	3.0	1.0	1.0	1.0	2.0	2.0	2.0	2.0	3.0	2.0	2.0	2.0	2.0	2.0
CD28	2.0	2.0	2.0	1.0	2.0	2.0	2.0	2.0	4.0	2.0	2.0	2.0	2.0	2.0
CD74	1.0	2.0	2.0	2.0	2.0	2.0	3.0	2.0	2.0	2.0	2.0	3.0	3.0	2.0
CD79A	1.0	2.0	2.0	2.0	3.0	3.0	3.0	1.0	2.0	2.0	2.0	3.0	3.0	2.0
CD79B	2.0	2.0	2.0	2.0	2.0	2.0	3.0	2.0	2.0	2.0	2.0	3.0	3.0	2.0
CDC73	3.0	2.0	1.0	1.0	2.0	2.0	1.0	2.1	4.0	2.0	2.0	2.0	2.0	2.0
CDH1	2.0	2.6	2.0	2.0	1.1	2.0	3.0	2.0	3.0	2.0	2.0	3.0	2.0	2.0
CDH10	3.0	2.0	2.0	1.8	1.7	2.0	>5.0	3.0	4.0	2.0	2.0	2.0	2.0	2.0
CDH11	2.0	2.0	2.0	2.0	1.3	2.0	2.0	2.0	3.0	2.0	2.0	2.0	2.0	2.0
CDH17	2.0	2.0	2.0	2.0	3.0	2.0	2.0	2.0	4.0	2.0	2.0	2.0	2.0	2.0
CDK12	2.0	2.0	2.0	1.2	2.0	2.0	3.0	2.0	3.0	2.0	2.0	3.0	3.0	2.0
CDK6	3.0	2.0	1.0	1.0	2.0	2.0	1.0	2.0	4.0	2.0	2.0	2.0	2.0	2.0
CDKN1B	3.0	2.0	1.0	1.0	2.0	2.0	2.0	2.0	4.0	2.0	2.0	2.0	2.0	2.0
CDKN2A	3.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	3.0	2.0	2.0	2.0	2.0	2.0
CDKN2C	2.0	2.0	1.0	1.0	2.0	2.0	2.0	2.0	4.0	2.0	2.0	3.0	2.0	2.0
CDX2	2.0	3.0	3.0	2.0	2.0	2.0	2.0	2.0	4.0	2.0	2.0	3.0	2.0	2.0
CEP89	2.0	2.0	2.0	1.4	2.0	3.0	2.6	2.0	3.4	2.0	2.0	3.0	2.0	2.0
CHCHD7	2.0	2.0	1.0	2.0	3.0	2.0	2.0	2.0	4.0	2.0	2.0	2.0	2.0	2.0
CHD2	2.0	2.0	2.0	1.4	2.0	1.3	2.0	3.0	3.0	2.0	2.0	2.0	2.0	2.0
CHD4	2.0	2.0	2.0	2.0	2.0	3.0	2.0	1.3	2.0	2.0	1.9	3.0	2.8	2.0
CHEK2	2.0	2.0	2.0	2.0	1.2	2.0	2.0	2.0	3.8	2.0	1.5	3.0	2.2	2.0
CHIC2	3.0	2.0	2.0	1.3	2.0	1.0	2.0	2.0	4.0	2.0	2.0	2.0	2.0	2.0
CHST11	2.0	2.0	1.0	1.0	2.0	2.0	2.0	2.0	4.0	2.0	2.0	2.0	2.0	2.0
CIC	1.0	1.0	2.0	2.0	3.0	3.0	3.0	1.0	2.0	2.0	2.0	3.0	3.0	2.0
CIITA	1.3	2.0	2.0	2.0	2.0	2.0	2.1	2.0	3.0	2.0	2.0	3.0	3.0	2.0
CLIP1	2.0	2.0	1.4	2.0	2.0	3.0	2.0	2.0	3.0	2.0	2.0	2.0	2.0	2.0
CLP1	2.0	2.0	2.0	3.0	2.0	2.0	2.0	2.0	3.0	2.0	2.0	2.0	2.0	2.0
CLTC	2.0	2.0	1.0	1.0	1.3	2.0	1.0	2.0	4.0	2.0	1.9	2.0	2.0	2.0
CLTCL1	2.0	2.0	2.0	2.0	2.3	2.7	3.0	2.0	2.3	2.0	2.0	3.0	2.5	2.0
CNBD1	2.0	2.0	1.0	1.0	3.0	2.0	2.0	2.0	4.0	2.0	2.0	2.0	2.0	2.0
CNOT3	1.3	1.0	2.0	2.0	2.7	2.6	3.0	2.0	2.0	2.0	2.0	4.0	3.0	2.0
CNTNAP2	2.0	2.0	1.0	1.5	2.0	2.0	2.0	2.0	4.0	2.0	2.0	2.0	2.0	2.0
CNTRL	2.0	2.0	1.0	1.0	2.0	2.0	1.8	2.0	4.0	2.0	2.0	2.0	2.0	2.0
COL1A1	2.0	2.0	1.0	2.0	2.0	2.8	3.0	2.0	2.0	2.0	2.0	3.7	2.6	2.0
COL2A1	1.6	2.0	2.0	2.0	1.8	2.7	2.8	2.0	2.0	2.0	2.0	3.7	2.0	1.9
COL3A1	3.0	2.0	1.0	1.0	2.0	2.0	1.7	2.7	4.0	2.1	2.0	2.0	1.0	2.0

CPEB3	3.0	2.0	2.0	1.0	1.0	2.0	2.0	2.0	4.0	2.0	2.0	2.0	2.0	2.0
CREB1	2.0	2.0	2.0	1.0	2.0	2.0	2.0	2.0	4.0	2.0	2.0	2.0	2.0	2.0
CREB3L1	1.7	2.0	2.0	3.0	2.0	2.0	2.8	2.0	2.7	2.0	2.0	3.0	3.0	2.0
CREB3L2	2.0	2.0	1.0	2.0	1.5	2.0	2.0	2.0	4.0	2.0	2.0	2.0	2.0	2.0
CREBBP	2.0	2.0	2.0	2.0	2.0	>5.0	2.1	2.0	2.8	2.0	2.0	3.0	>5.0	2.0
CRLF2	1.5	2.0	2.0	2.0	2.0	3.0	3.0	1.6	2.0	1.9	2.0	4.0	3.0	2.0
CRNKL1	2.0	2.0	2.0	1.0	1.2	1.1	2.0	2.0	4.0	2.0	2.0	2.0	2.0	2.0
CRTC1	1.0	1.0	2.0	2.0	3.0	3.0	3.0	2.0	2.0	2.0	2.0	4.0	3.0	2.0
CRTC3	2.2	2.0	2.0	2.0	2.0	1.9	2.0	2.0	3.0	2.0	2.0	2.0	2.0	2.0
CSF1R	2.0	2.0	2.0	2.0	2.0	2.0	2.0	3.0	2.0	2.0	2.0	3.0	2.0	2.0
CSF3R	2.0	2.0	2.0	2.0	2.0	2.0	3.0	2.0	3.0	2.0	2.0	3.0	2.0	2.0
CSMD3	2.0	2.0	1.0	1.0	2.0	2.0	2.0	3.0	4.0	2.0	2.0	2.0	1.0	2.0
CTCF	2.0	1.0	2.0	2.0	3.0	2.0	3.0	2.0	2.0	2.0	1.2	3.0	3.0	2.0
CTNNA1	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	3.0	2.0	2.0	3.0	2.0	2.0
CTNNA2	2.0	2.0	2.0	2.0	1.9	2.0	2.0	2.0	3.6	2.0	2.0	2.0	2.0	2.0
CTNNB1	2.0	2.0	2.0	2.0	2.0	1.2	2.0	2.0	3.9	2.0	2.0	2.0	2.0	2.0
CTNND1	2.0	2.0	2.0	3.0	2.0	2.0	2.0	2.0	4.0	2.0	2.0	2.0	2.0	2.0
CTNND2	2.0	2.0	2.0	2.0	2.0	2.0	2.0	3.0	3.0	2.0	2.0	2.0	2.0	2.0
CUL3	2.0	2.0	2.0	1.1	2.0	2.3	1.2	2.0	4.0	2.0	2.0	2.0	2.0	2.0
CUX1	1.0	2.0	1.1	2.0	2.9	3.0	3.0	2.0	2.0	2.0	2.0	4.0	3.0	2.0
CXCR4	2.0	2.0	2.0	1.0	2.0	2.0	2.0	2.0	4.0	2.0	2.0	2.0	2.0	2.0
CYLD	2.0	2.0	1.2	1.2	1.7	1.2	1.2	2.0	4.0	2.0	2.0	2.0	2.0	2.0
CYP2C8	3.0	2.0	2.0	1.0	1.0	2.0	2.0	2.0	4.0	2.0	2.0	2.0	1.3	2.0
CYSLTR2	3.0	3.0	3.0	1.0	2.0	1.0	2.0	2.0	4.0	2.0	2.0	2.0	2.0	2.0
DAXX	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	3.0	3.0	2.0
DCAF12L2	1.0	2.0	2.0	1.0	2.0	2.0	2.0	2.0	4.0	2.0	2.0	2.0	2.0	2.0
DCC	3.0	2.0	1.1	1.3	1.0	1.4	2.0	2.4	3.0	2.0	2.0	2.0	1.3	2.0
DCTN1	2.0	2.0	2.0	2.0	1.8	>5.0	2.0	2.0	2.0	2.0	2.0	3.0	2.0	2.0
DDB2	2.0	2.0	2.0	3.0	2.0	2.0	2.0	2.0	3.0	2.0	2.0	3.0	2.0	2.0
DDIT3	2.0	2.0	2.0	2.0	2.0	2.0	2.0	1.0	2.0	2.0	2.0	3.0	2.0	2.0
DDR2	2.0	2.0	2.0	1.9	1.0	2.0	2.0	2.0	4.0	2.0	2.0	2.0	1.1	2.0
DDX10	2.0	2.0	3.0	1.3	2.0	2.0	1.4	2.0	4.0	2.0	1.0	2.0	2.0	2.0
DDX3X	1.0	2.0	1.0	1.0	2.0	2.0	2.0	2.0	4.0	2.0	2.0	2.0	2.0	2.0
DDX5	2.0	2.0	1.0	1.2	2.0	1.1	1.1	2.0	4.0	2.0	2.0	2.0	2.0	2.0
DDX6	2.0	2.0	3.0	2.0	2.0	2.0	>5.0	2.0	3.0	2.0	1.0	3.0	3.0	2.0
DEK	3.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	4.0	2.0	2.0	2.0	2.0	2.0
DGCR8	1.0	1.6	2.0	2.0	3.0	2.2	3.0	2.0	2.0	2.0	2.0	4.0	2.6	2.0
DICER1	3.0	2.0	1.3	1.4	2.0	2.0	1.0	2.0	3.8	2.0	2.0	2.0	2.0	2.0
DNAJB1	1.0	1.0	1.3	2.0	2.0	2.0	3.0	2.0	2.0	2.0	2.0	4.0	3.0	2.0
DNM2	1.8	2.0	2.0	2.0	2.1	2.3	3.0	2.0	2.0	2.0	2.0	4.0	3.0	2.0
DNMT3A	1.0	2.0	2.0	2.0	2.0	3.0	3.0	2.0	2.3	2.0	2.0	3.0	2.8	2.0
DROSHA	3.0	2.0	1.0	1.8	2.0	2.0	2.0	3.0	4.0	2.0	2.0	2.0	2.0	2.0
EBF1	2.0	2.0	1.9	2.0	2.0	2.0	2.0	3.0	3.0	2.0	2.0	2.0	2.0	2.0
ECT2L	3.0	2.0	2.0	1.9	1.9	2.0	2.0	2.0	4.0	2.0	2.0	2.0	2.0	2.0
EED	3.0	2.0	3.0	1.2	2.0	2.0	1.0	3.0	4.0	2.0	2.0	2.0	2.0	2.0
EGFR	2.0	2.0	2.0	2.0	1.6	1.1	2.0	2.0	3.0	2.3	2.0	2.0	1.1	1.8
EIF1AX	1.0	2.0	1.0	1.0	2.0	2.0	2.0	2.0	4.0	2.0	2.0	2.0	2.0	2.0
EIF3E	2.0	2.0	1.0	1.1	2.0	2.0	1.4	2.0	4.0	2.0	2.0	2.0	2.0	2.0
EIF4A2	2.0	2.0	1.0	1.0	2.0	1.1	2.0	2.0	4.0	2.0	2.0	2.0	2.0	2.0
ELF3	2.0	2.0	2.0	2.0	2.0	3.0	3.0	2.0	2.0	2.0	2.0	3.0	2.0	2.0
ELF4	1.0	2.0	2.0	2.0	2.0	2.0	2.0	1.0	3.0	2.0	2.0	3.0	2.0	2.0
ELK4	2.0	2.0	2.0	2.0	2.0	3.0	2.0	2.0	3.0	2.0	2.0	3.0	2.0	2.0
ELL	1.0	1.0	2.0	2.0	3.0	3.0	3.0	2.0	2.0	2.0	2.0	4.0	3.0	2.0
ELN	1.0	2.0	1.6	2.0	2.0	3.0	3.0	2.0	2.0	2.0	2.0	3.0	3.0	2.0
EML4	2.0	2.0	2.0	1.1	2.0	2.0	2.0	2.0	4.0	2.0	2.0	2.0	2.0	2.0
EP300	2.0	2.0	2.0	1.3	1.1	2.0	2.0	2.0	3.7	2.0	1.5	>5.0	>5.0	2.0
EPHA3	2.0	2.0	1.6	1.0	2.0	2.0	2.0	3.0	4.0	2.0	2.0	2.0	2.0	2.0
EPHA7	3.0	2.0	1.0	1.0	1.0	1.9	2.0	2.0	4.0	2.0	2.0	2.0	1.0	2.0
EPS15	2.0	2.0	2.0	1.0	2.0	2.0	2.0	2.0	4.0	2.0	2.0	2.0	2.0	2.0
ERBB2	2.0	2.0	2.0	2.0	2.0	2.0	3.0	2.0	2.0	2.0	2.0	3.0	3.0	2.0
ERBB3	2.0	2.0	2.0	2.0	1.0	2.0	2.0	1.5	2.0	2.0	1.5	3.0	2.0	2.0
ERBB4	3.0	2.0	1.0	1.0	1.1	2.0	2.0	2.8	4.0	2.0	2.0	2.0	1.1	2.0
ERC1	2.0	2.0	2.0	2.0	2.0	2.3	2.0	2.0	3.0	2.0	2.0	3.0	2.0	2.0
ERCC2	1.0	1.0	2.0	2.0	2.0	3.0	3.0	2.0	2.0	2.0	2.0	3.0	3.0	2.0
ERCC3	2.0	2.0	2.0	2.0	2.0	2.0	2.1	2.0	3.0	2.0	2.0	2.9	2.0	2.0
ERCC4	2.0	2.0	2.0	1.3	1.5	2.0	2.0	2.0	3.9	2.0	2.0	2.0	2.0	2.0
ERCC5	3.0	2.0	1.0	1.0	1.2	1.2	2.0	2.0	4.0	2.0	2.0	2.0	2.0	2.0
ERG	2.0	2.0	3.0	2.0	2.0	2.0	2.0	2.0	4.0	2.0	2.0	2.0	2.0	2.0

ESR1	3.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	4.0	2.0	2.0	2.0	2.0	2.0
ETNK1	3.0	2.0	1.0	1.0	2.0	2.0	1.0	2.0	4.0	2.0	2.0	2.0	1.0	2.0
ETV1	3.0	2.0	1.6	1.0	1.0	1.0	2.0	3.0	4.0	2.0	1.6	2.0	1.0	2.0
ETV4	2.0	2.0	2.0	2.0	2.0	2.0	3.0	2.0	2.0	2.0	2.0	3.0	2.0	2.0
ETV5	2.0	2.3	2.0	2.0	2.0	2.0	2.0	2.0	3.0	2.0	2.0	2.0	2.0	2.0
ETV6	2.0	2.0	2.0	2.0	1.7	2.0	2.0	2.0	4.0	2.0	2.0	2.0	2.0	2.0
EWSR1	1.9	2.0	1.7	2.0	2.0	2.0	2.3	2.0	2.0	2.0	2.0	2.7	3.0	2.0
EXT1	2.0	2.0	1.0	2.0	2.0	2.0	2.0	2.0	4.0	2.0	2.0	2.0	2.0	2.0
EXT2	2.0	2.0	2.0	3.0	2.0	2.0	2.0	2.0	3.8	2.0	2.0	2.3	2.0	2.0
EZH2	2.0	2.0	1.0	1.5	2.0	2.0	2.0	2.0	3.7	2.0	2.0	2.0	2.0	2.0
EZR	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	3.0	2.0	2.0	2.4	2.0	2.0
FAM131B	2.0	2.0	1.0	3.0	2.0	2.0	2.0	2.0	3.0	2.0	2.0	3.0	2.0	2.0
FAM135B	2.0	2.0	2.0	2.0	2.9	2.0	3.0	2.0	3.0	2.0	2.0	3.0	2.0	2.0
FAM47C	1.0	2.0	1.0	1.0	2.0	2.0	2.0	3.0	4.0	2.0	2.0	2.0	2.0	2.0
FANCA	1.9	1.8	2.0	2.0	>5.0	3.0	3.0	2.0	2.0	2.0	2.0	4.0	3.0	2.0
FANCC	2.0	2.0	1.0	1.7	2.0	1.0	1.7	2.0	3.0	>5.0	1.6	3.0	1.0	2.0
FANCD2	1.0	2.0	2.0	1.0	1.3	2.0	2.0	2.0	4.0	2.0	1.5	2.0	2.0	2.0
FANCE	1.0	2.0	2.0	2.0	2.0	2.0	3.0	2.0	2.0	2.0	2.0	3.0	3.0	2.0
FANCF	2.0	2.0	2.0	1.0	2.0	2.0	2.0	3.0	4.0	2.0	2.0	2.0	2.0	2.0
FANCG	2.0	2.0	2.0	2.0	2.0	2.0	2.0	1.3	2.0	2.0	2.0	3.0	2.0	2.0
FAS	3.0	2.0	1.0	1.0	1.0	1.9	1.3	2.0	4.0	2.0	2.0	2.0	2.0	2.0
FAT1	3.0	1.0	1.0	2.0	2.0	1.0	2.0	2.0	3.0	2.0	2.0	3.0	1.0	2.0
FAT3	2.0	2.0	3.0	3.0	2.0	2.0	2.0	3.0	4.0	2.0	2.0	2.0	1.0	2.0
FAT4	3.0	2.0	1.0	1.0	2.0	2.0	2.0	2.0	4.0	2.0	2.0	2.0	2.0	2.0
FBLN2	1.0	2.0	2.0	2.0	2.0	2.7	3.0	2.0	2.0	2.0	2.0	3.0	2.0	2.0
FBXO11	2.0	2.0	1.0	1.0	2.0	2.0	2.0	2.0	4.0	2.0	2.0	2.0	2.0	2.0
FBXW7	2.0	2.0	1.0	1.0	2.0	2.0	2.0	2.0	3.8	2.0	2.0	2.0	2.0	2.0
FCGR2B	2.0	2.0	2.0	2.0	1.5	2.3	2.0	2.0	3.0	2.0	2.0	3.0	3.0	2.0
FCRL4	2.0	2.0	2.0	2.0	2.0	3.0	2.0	2.0	3.0	2.0	1.8	3.0	2.0	2.0
FEN1	2.0	2.0	2.0	3.0	2.0	2.0	3.0	2.0	3.0	2.0	2.0	3.0	3.0	2.0
FES	1.1	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	3.0	2.0	2.0
FEV	1.0	2.0	2.0	2.0	2.0	3.0	3.0	2.0	2.0	2.0	2.0	3.0	3.0	2.0
FGFR1	2.0	2.0	2.0	2.0	3.0	3.0	2.0	2.0	3.0	2.0	1.0	3.5	2.0	2.0
FGFR1OP	2.0	2.0	1.0	2.0	2.0	2.0	2.0	2.0	4.0	2.0	2.0	3.0	2.0	2.0
FGFR2	3.0	2.0	2.0	2.0	1.0	2.0	2.0	2.0	4.0	2.0	2.0	2.0	1.0	2.0
FGFR3	1.0	1.0	2.0	2.0	2.0	3.0	3.0	2.0	2.0	2.0	2.0	4.0	3.0	2.0
FGFR4	1.0	2.0	2.0	2.0	2.0	3.0	3.0	2.0	2.5	2.0	2.0	3.0	3.0	2.0
FH	3.0	2.0	1.1	1.0	1.1	3.0	2.0	2.0	4.0	2.0	2.0	2.0	1.2	2.0
FIP1L1	3.0	2.0	1.7	1.4	1.6	1.0	1.7	2.0	4.0	2.0	2.0	2.0	1.3	2.0
FLCN	1.0	2.0	2.0	2.0	2.0	2.0	3.0	2.0	2.0	2.0	2.0	3.0	2.0	2.0
FLI1	2.0	2.0	3.0	2.0	2.0	2.0	2.0	2.0	3.0	2.0	1.0	2.0	2.0	2.0
FLNA	0.8	1.0	2.0	2.0	3.0	2.0	3.0	1.0	2.0	2.0	3.0	4.0	3.0	2.0
FLT3	3.0	3.0	3.0	1.0	2.0	2.0	2.0	2.0	4.0	2.0	2.0	2.0	2.0	2.0
FLT4	1.0	1.2	2.0	2.0	2.0	3.0	3.0	2.0	>5.0	2.0	2.0	4.0	2.9	2.0
FNBP1	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.8	2.0	1.6	3.0	2.8	2.0
FOXA1	3.0	2.0	1.0	2.0	3.0	2.0	1.0	3.0	4.0	2.0	2.0	2.0	1.0	2.0
FOXL2	2.0	2.0	2.0	1.0	2.0	2.0	2.0	2.0	4.0	2.0	2.0	2.0	2.0	2.0
FOXO1	3.0	3.0	3.0	1.0	2.0	1.0	2.0	2.0	4.0	2.0	2.0	2.0	2.0	2.0
FOXO3	3.0	2.0	1.0	1.0	2.0	2.0	2.0	2.0	4.0	2.0	2.0	2.0	2.0	2.0
FOXO4	1.0	2.0	2.0	2.0	2.0	2.0	2.0	1.0	3.0	2.0	2.0	3.0	3.0	2.0
FOXP1	2.0	2.0	1.0	2.0	1.4	1.0	2.0	2.0	4.0	2.0	2.0	2.0	1.6	2.0
FOXR1	2.0	2.0	3.0	2.0	2.0	3.0	2.0	2.0	3.0	2.0	1.0	3.0	3.0	2.0
FSTL3	1.0	1.0	2.0	2.0	3.0	3.0	4.0	2.0	2.0	2.0	2.0	4.0	3.0	2.0
FUBP1	3.0	2.0	1.0	1.0	2.0	1.0	1.0	2.0	4.0	2.0	2.0	2.0	2.0	2.0
FUS	2.0	2.0	1.3	2.0	2.0	>5.0	2.0	2.0	2.0	2.0	1.4	2.5	3.0	2.0
GAS7	2.0	2.0	2.0	2.0	2.0	2.0	3.0	2.0	3.0	2.0	2.0	3.0	2.0	2.0
GATA1	1.0	2.0	2.0	2.0	2.0	2.8	3.0	2.0	2.0	2.0	2.0	3.0	3.0	2.0
GATA3	2.0	2.0	1.0	2.0	1.0	2.0	2.0	2.0	3.0	2.0	2.0	2.0	3.0	2.0
GLI1	2.0	2.0	2.0	2.0	2.0	3.0	2.0	2.0	2.0	2.0	2.0	3.0	3.0	2.0
GMPS	3.0	2.0	2.0	1.0	2.0	2.0	2.0	2.0	4.0	2.0	2.0	2.0	2.0	2.0
GNA11	1.0	1.0	2.0	2.0	3.0	3.0	3.0	2.0	2.0	2.0	2.0	4.0	3.0	2.0
GNAS	2.0	2.0	2.0	2.0	1.3	1.3	2.0	2.0	3.0	2.0	2.0	3.0	2.0	2.0
GOLGA5	2.0	2.0	2.0	1.6	3.0	2.0	2.0	2.0	3.0	2.0	2.0	3.0	2.0	2.0
GOPC	3.0	2.0	1.0	1.0	2.0	2.0	2.0	2.3	4.0	2.0	2.0	2.0	1.2	2.0
GPC3	1.0	2.0	1.0	1.1	1.0	2.0	2.0	2.0	4.0	2.0	2.0	2.0	2.0	2.0
GPC5	2.9	2.0	1.3	1.0	2.0	2.0	2.0	2.0	4.0	2.0	2.0	2.0	2.0	2.0
GPHN	2.0	2.0	1.0	1.1	2.0	2.0	2.0	2.0	3.0	2.0	2.0	2.0	1.2	2.0
GRIN2A	2.0	2.8	2.0	2.0	2.0	2.0	2.0	2.0	3.8	2.0	2.0	2.0	2.0	2.0

GRM3	3.0	2.0	1.0	1.8	2.0	2.0	2.0	2.0	4.0	2.0	2.0	2.0	1.0	2.0
H3F3A	2.0	2.0	2.0	1.0	2.0	3.0	2.0	2.0	3.0	2.0	2.0	3.0	2.0	2.0
H3F3B	1.0	2.0	1.0	2.0	3.0	3.0	3.0	2.0	2.0	2.0	2.0	4.0	3.0	2.0
HERPUD1	2.0	2.0	1.4	2.0	1.0	1.0	2.0	2.0	3.0	2.0	2.0	3.0	3.0	2.0
HEY1	2.0	2.0	1.0	2.0	3.0	2.0	2.0	2.0	4.0	2.0	2.0	2.0	2.0	2.0
HIF1A	2.0	2.0	1.0	2.0	3.0	2.0	2.0	2.0	4.0	2.0	2.0	2.0	2.0	2.0
HIP1	1.0	2.0	1.0	2.0	2.0	3.0	3.0	2.0	2.0	2.0	2.0	3.0	3.0	2.0
HIST1H3B	2.0	2.0	2.0	2.0	2.0	1.0	2.0	2.0	3.0	2.0	3.0	2.0	3.0	2.0
HLF	2.0	2.0	1.0	2.0	1.3	2.0	2.0	2.0	3.0	2.0	2.0	2.0	2.0	2.0
HMGA1	2.0	2.0	2.0	2.0	2.0	2.0	3.0	2.0	2.0	2.0	2.0	3.0	3.0	2.0
HMGA2	2.6	2.0	1.0	1.0	2.0	2.0	2.0	2.0	4.0	2.0	1.5	2.0	1.5	2.0
HMGN2P46	3.0	2.0	2.0	2.0	2.0	2.0	2.0	3.0	4.0	2.0	2.0	2.0	2.0	2.0
HNF1A	1.0	2.0	2.0	2.0	2.0	3.0	3.0	2.0	2.0	2.0	2.0	4.0	2.0	2.0
HNRNPA2B1	2.0	2.0	1.0	1.0	2.0	2.0	1.4	2.0	4.0	2.0	2.0	2.0	2.0	2.0
HOOK3	2.0	2.0	1.2	2.0	3.0	2.0	2.0	2.0	4.0	2.0	2.0	2.0	2.0	2.0
HOXA11	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	3.0	2.0	2.0	3.0	2.0	2.0
HOXA13	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	3.0	2.0	2.0	3.0	2.0	2.0
HOXA9	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	3.0	2.0	2.0	3.0	2.0	2.0
HOXD11	1.0	2.0	2.0	2.0	2.0	2.0	2.0	1.0	3.0	2.0	2.0	3.0	2.0	2.0
HOXD13	1.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	3.0	2.0	2.0	3.0	2.0	2.0
HRAS	1.0	1.0	2.0	3.0	3.0	3.0	4.0	2.0	2.0	2.0	2.0	4.0	3.0	2.0
HSP90AA1	2.0	2.0	2.0	2.0	2.1	2.0	1.2	1.1	3.0	2.0	2.0	2.2	2.0	2.0
HSP90AB1	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	1.5	2.2	2.3	2.0
IDH1	2.0	2.0	1.0	1.0	2.0	2.0	1.5	2.0	4.0	2.0	2.0	2.0	2.0	2.0
IDH2	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.3	3.0	2.0	2.0
IGF2BP2	2.0	2.0	2.5	2.0	1.5	2.0	2.0	2.0	4.0	2.0	2.0	2.0	2.0	2.0
IKBKB	2.0	2.0	1.0	2.0	2.2	2.0	2.0	2.0	3.0	2.0	1.0	2.0	2.0	2.0
IKZF1	2.0	2.0	2.0	1.9	2.0	2.0	2.0	2.0	3.0	2.0	2.0	2.1	2.0	2.0
IKZF3	2.0	2.0	2.0	1.1	2.0	2.0	2.5	2.0	2.9	2.0	2.0	3.0	3.0	2.0
IL2	3.0	2.0	1.0	1.0	2.0	2.0	2.0	2.0	4.0	2.0	2.0	2.0	2.0	2.0
IL21R	2.0	2.0	2.0	2.0	2.0	3.0	3.0	2.0	2.0	2.0	2.0	3.0	3.0	2.0
IL6ST	2.0	2.0	2.0	1.0	1.6	2.0	1.3	3.0	4.0	2.0	2.0	2.0	2.0	2.0
IL7R	3.0	2.0	2.0	1.3	2.0	2.0	2.0	3.0	4.0	2.0	1.0	2.0	2.0	2.0
IRF4	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	3.0	2.0	2.0	3.0	3.0	2.0
IRS4	1.0	2.0	1.0	1.0	2.0	2.0	2.0	2.0	4.0	2.0	2.0	2.0	1.0	2.0
ISX	2.0	2.0	2.0	2.0	2.0	2.0	3.0	2.0	3.0	2.0	2.0	2.0	2.0	2.0
ITGAV	3.0	2.0	1.0	1.0	2.0	2.0	1.0	3.0	4.0	2.0	2.0	2.0	2.0	2.0
ITK	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	3.0	2.0	1.0	2.0	2.9	2.0
JAK1	3.0	2.0	2.0	1.0	2.0	2.0	2.0	2.0	4.0	2.0	2.0	2.0	2.0	2.0
JAK2	3.0	1.0	1.0	1.0	2.0	2.0	1.4	2.0	3.0	2.0	2.0	2.0	2.0	2.0
JAK3	1.0	1.7	2.0	2.0	2.0	3.0	3.0	2.0	2.0	2.0	2.0	4.0	3.0	2.0
JUN	3.0	2.0	2.0	1.0	2.0	1.0	2.0	2.0	4.0	2.0	2.0	2.0	2.0	2.0
KAT6A	2.0	2.0	1.6	2.0	3.0	>5.0	2.0	2.0	3.0	2.0	1.0	>5.0	2.0	2.0
KAT6B	2.8	2.0	1.1	1.3	1.0	2.0	1.2	2.0	4.0	2.0	2.0	2.1	2.0	2.0
KAT7	2.0	2.0	1.0	1.1	1.0	1.4	2.0	2.0	3.0	2.0	2.0	3.0	2.0	2.0
KCNJ5	2.0	2.0	3.0	2.0	2.0	2.0	2.0	2.0	3.0	2.0	1.0	2.0	2.0	2.0
KDM5A	2.0	2.0	2.0	1.1	1.2	2.5	1.3	2.0	3.0	2.0	1.8	2.0	2.0	2.0
KDM5C	1.0	2.0	2.0	2.0	2.0	2.0	3.0	2.0	3.0	2.0	2.0	3.0	2.0	2.0
KDM6A	1.9	2.0	3.7	1.1	1.3	1.0	1.1	2.0	3.9	2.0	2.0	2.0	2.0	2.0
KDR	2.9	2.0	1.0	1.0	1.5	1.6	2.0	2.0	4.0	2.0	2.0	2.0	1.0	2.0
KDSR	3.0	2.0	1.0	1.0	2.0	2.0	2.0	2.0	4.0	2.0	2.0	2.0	2.0	2.0
KEAP1	1.0	1.0	2.0	2.0	3.0	3.0	3.0	2.0	2.0	2.0	2.0	4.0	3.0	2.0
KIAA1549	2.0	2.0	1.0	2.0	2.0	2.0	2.4	2.0	3.4	2.0	2.0	2.0	2.0	2.0
KIF5B	3.0	2.0	1.0	1.0	1.0	2.0	1.0	2.0	4.0	2.0	2.0	2.0	2.0	2.0
KIT	3.0	2.0	1.1	1.2	1.0	1.0	2.0	2.0	4.0	2.0	2.0	2.0	1.0	2.0
KLF4	3.0	2.0	1.0	2.0	2.0	2.0	2.0	2.0	4.0	2.0	2.0	2.0	2.0	2.0
KLF6	2.0	2.0	1.0	2.0	1.0	2.0	2.0	2.0	3.0	2.0	2.0	2.0	2.0	2.0
KLK2	1.0	2.0	3.0	2.0	3.0	3.0	3.0	2.0	2.0	2.0	2.0	3.0	3.0	2.0
KRAS	3.0	2.0	2.0	1.0	2.0	2.0	1.0	2.0	4.0	2.0	2.0	2.0	1.2	2.0
KTN1	3.0	2.0	1.0	1.0	3.0	2.0	1.0	2.0	4.0	2.0	2.0	2.0	2.0	2.0
LARP4B	3.0	2.0	1.0	2.0	1.0	2.0	2.0	2.0	3.0	2.0	2.0	3.0	2.0	2.0
LASP1	2.0	2.0	2.0	2.0	2.0	2.0	3.0	2.0	3.0	2.0	2.0	3.0	3.0	2.0
LATS1	3.0	2.0	2.0	1.0	2.0	2.0	2.0	2.0	4.0	2.0	2.0	2.0	2.0	2.0
LATS2	2.6	2.0	2.0	1.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0
LCK	2.0	2.0	2.0	2.0	2.0	2.0	3.0	2.0	2.0	2.0	2.0	3.0	3.0	2.0
LCP1	3.0	3.0	3.0	1.0	1.0	1.0	2.0	2.0	4.0	2.0	2.0	2.0	2.0	2.0
LEF1	2.9	2.0	1.0	1.0	2.0	2.0	2.0	2.0	4.0	2.0	2.0	2.0	2.0	2.0
LEPROTL1	2.0	2.0	2.0	2.0	3.0	2.0	2.0	2.0	4.0	2.0	1.0	2.0	2.9	2.0

LIFR	3.0	2.0	1.0	1.0	1.1	1.1	1.0	3.0	4.0	2.0	1.0	2.0	1.0	2.0
LMNA	1.0	1.5	2.0	2.0	2.0	3.0	3.0	2.0	2.0	2.0	2.0	3.0	3.0	2.0
LMO1	2.0	2.0	2.0	3.0	2.0	2.0	2.0	3.0	3.4	2.0	2.0	2.0	2.0	2.0
LPP	2.0	3.0	2.5	2.0	2.0	2.0	2.0	2.0	3.0	2.0	2.0	2.0	2.0	2.0
LRIG3	3.0	2.0	1.0	1.0	1.1	1.6	1.0	2.0	4.0	2.0	2.0	2.0	1.0	2.0
LRP1B	2.0	2.0	2.0	1.0	1.4	2.0	2.0	2.8	4.0	2.0	2.0	2.0	1.0	2.0
LSM14A	1.9	2.0	2.0	1.4	2.0	2.0	2.0	2.0	3.0	2.0	2.0	2.9	3.0	2.0
LYL1	1.0	1.0	2.0	2.0	3.0	2.3	3.0	2.0	2.0	2.0	2.0	4.0	3.0	2.0
LYN	2.0	2.0	1.1	1.8	2.0	2.0	2.0	2.0	4.0	2.0	2.0	2.0	2.0	2.0
LZTR1	1.0	1.0	2.0	2.0	3.0	3.0	3.0	2.0	2.0	2.0	2.0	4.0	3.0	2.0
MACC1	3.0	2.0	2.0	1.0	2.0	1.0	2.0	3.0	4.0	2.0	2.0	2.0	2.0	2.0
MAF	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	3.0	2.0	2.0	2.5	2.0	2.0
MAFB	2.0	2.0	1.0	2.0	2.0	2.0	2.0	2.0	3.0	2.0	2.0	2.0	2.0	2.0
MALAT1	1.0	1.0	2.0	3.0	3.0	2.0	3.0	2.0	2.0	2.0	2.0	3.0	3.0	2.0
MALT1	3.0	2.0	2.0	1.0	2.0	2.0	1.7	2.0	3.0	2.0	1.7	2.0	2.0	2.0
MAML2	3.0	2.0	3.0	2.0	2.0	1.0	1.0	3.0	4.0	2.0	1.6	2.0	1.0	2.0
MAP2K1	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	3.0	2.0	1.6	2.0	2.0	2.0
MAP2K2	1.0	1.0	2.0	2.0	3.0	3.0	3.0	2.0	2.0	2.0	2.0	4.0	3.0	2.0
MAP2K4	2.0	2.0	2.0	1.4	2.0	2.0	>5.0	2.0	3.0	2.0	2.0	2.0	2.0	2.0
MAP3K1	2.0	2.0	1.1	1.1	1.1	1.1	1.1	3.0	4.0	2.0	2.0	2.0	2.0	2.0
MAP3K13	2.0	2.4	2.0	2.0	1.5	2.0	2.0	2.0	4.0	2.0	2.0	2.0	2.0	2.0
MAPK1	1.0	2.0	2.0	1.1	3.0	3.0	3.0	2.0	2.0	2.0	2.0	4.0	3.0	2.0
MAX	2.0	2.0	1.4	2.0	2.1	2.0	2.0	2.0	3.0	2.0	2.0	2.0	2.0	2.0
MB21D2	2.0	2.0	2.0	1.0	2.0	2.0	2.0	2.0	4.0	2.0	2.0	2.0	2.0	2.0
MDM2	3.0	2.0	1.0	1.0	2.0	2.0	1.0	2.0	4.0	2.0	2.0	2.0	2.0	2.0
MDM4	2.0	2.0	2.0	1.0	2.0	3.0	2.0	2.0	2.0	2.0	2.0	3.0	2.0	2.0
MECOM	2.9	>5.0	1.0	1.0	2.0	1.0	2.0	2.3	4.0	2.0	2.0	2.0	1.3	2.0
MED12	1.0	2.0	2.0	2.0	2.0	>5.0	2.3	1.0	2.3	2.0	1.8	3.0	3.0	2.0
MEN1	1.0	1.0	2.0	3.0	3.0	2.0	3.0	2.0	2.0	2.0	2.0	4.0	3.0	2.0
MET	3.0	2.0	1.0	1.1	2.0	2.0	1.1	2.4	4.0	2.0	2.0	2.0	1.0	2.0
MGMT	2.0	2.0	2.0	2.0	1.0	2.0	2.0	2.0	3.0	2.0	2.0	3.0	2.0	2.0
MITF	2.0	2.0	1.0	1.1	1.0	1.0	2.0	2.0	4.0	2.0	2.0	2.0	1.0	2.0
MLF1	2.3	2.0	1.0	1.0	2.0	2.0	1.3	2.3	4.0	2.0	2.0	2.0	2.0	2.0
MLH1	2.0	2.0	2.0	1.1	2.0	2.0	2.0	2.0	4.0	2.0	1.6	2.0	2.0	2.0
MLLT1	1.0	2.0	2.0	2.0	3.0	3.0	3.0	2.0	2.1	2.0	2.0	4.0	3.0	2.0
MLLT10	2.8	2.0	1.0	1.1	1.0	2.0	2.0	2.0	4.0	2.0	2.0	2.0	2.0	2.0
MLLT11	2.0	2.0	2.0	2.0	2.0	3.0	3.0	2.0	2.0	2.0	2.0	3.0	3.0	2.0
MLLT3	3.0	2.0	2.0	1.0	1.5	1.5	2.0	2.0	3.0	2.0	1.2	2.0	2.0	2.0
MLLT6	1.9	1.0	2.0	2.0	2.0	2.0	3.0	2.0	2.0	2.0	2.0	3.0	3.0	2.0
MN1	1.0	2.0	2.0	2.0	2.0	3.0	3.0	2.0	3.0	2.0	2.0	3.0	2.0	2.0
MNX1	2.0	2.0	1.0	2.0	2.0	2.0	2.0	2.0	4.0	2.0	2.0	2.0	2.0	2.0
MPL	2.0	2.0	2.0	2.0	2.0	2.0	3.0	2.0	2.0	2.0	2.0	3.0	2.0	2.0
MSH2	2.0	2.0	1.0	1.0	2.0	2.0	2.0	2.0	4.0	2.0	1.4	2.0	2.0	2.0
MSH6	2.0	2.0	1.0	1.0	2.0	2.0	2.0	2.0	4.0	2.0	2.0	2.0	2.0	2.0
MSI2	2.0	2.0	1.0	2.0	2.0	2.0	2.1	2.0	3.0	2.0	1.6	2.2	2.0	2.0
MSN	1.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	3.0	2.0	2.0	2.0	2.0	2.0
MTCP1	1.0	2.0	1.0	1.0	2.0	2.0	2.0	2.0	4.0	2.0	2.0	2.0	2.0	2.0
MTOR	2.0	2.0	2.0	2.0	2.0	>5.0	3.0	2.0	2.8	2.0	2.0	3.0	2.2	2.0
MUC1	2.0	2.0	2.0	2.0	2.0	3.0	3.0	1.0	2.0	2.0	2.0	3.0	3.0	2.0
MUC16	1.0	>5.0	2.6	2.6	>5.0	3.0	3.0	2.4	3.0	2.0	2.0	3.0	2.0	2.0
MUC4	2.0	2.0	3.0	2.0	2.0	3.0	3.0	2.0	2.0	2.0	2.0	4.0	3.0	2.0
MUTYH	2.0	2.0	2.0	2.0	2.0	2.0	3.0	2.0	2.0	2.0	2.0	3.0	2.0	2.0
MYB	3.0	2.0	1.0	1.0	2.0	1.7	2.0	2.0	4.0	2.0	2.0	2.0	2.0	2.0
MYC	2.0	2.0	1.0	2.0	2.0	2.0	2.0	2.0	4.0	2.0	2.0	2.0	2.0	2.0
MYCN	2.0	2.0	2.0	1.0	2.0	2.0	2.0	2.0	4.0	2.0	2.0	2.0	2.0	2.0
MYD88	1.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	3.0	2.0	2.0
MYH11	2.0	2.3	2.9	2.0	2.0	2.5	2.8	2.0	2.3	2.0	2.0	3.0	3.0	2.0
MYH9	1.1	1.3	2.0	2.0	2.0	2.0	2.5	2.0	2.0	2.0	2.0	3.0	2.7	2.0
MYO5A	2.0	2.0	2.0	1.0	1.5	1.4	1.7	2.8	4.0	2.0	2.0	>5.0	2.0	2.0
MYOD1	1.0	2.0	2.0	3.0	3.0	3.0	3.0	3.0	3.0	2.0	2.0	3.0	2.0	2.0
N4BP2	2.0	2.0	1.0	1.0	2.0	2.0	1.3	2.0	4.0	2.0	2.0	2.0	2.0	2.0
NAB2	2.0	2.0	2.0	2.0	2.0	3.0	2.6	2.0	2.0	2.0	2.0	3.0	2.0	2.0
NACA	2.0	2.0	2.0	1.1	2.0	3.0	2.0	2.0	3.0	2.0	2.0	2.0	2.0	2.0
NBEA	3.0	3.0	3.0	1.0	2.0	1.0	1.4	2.8	4.0	2.0	2.0	2.0	1.0	2.0
NBN	2.0	2.0	1.0	1.0	3.0	2.0	1.1	2.0	4.0	2.0	2.0	2.0	2.0	2.0
NCKIPSD	1.0	1.0	2.0	2.0	2.0	2.0	3.0	2.0	2.0	2.0	2.0	4.0	2.0	2.0
NCOA1	2.0	2.0	2.0	1.7	2.0	2.0	2.0	2.0	3.0	2.0	2.0	2.2	2.0	2.0
NCOA2	2.0	2.0	1.5	1.7	2.0	2.0	2.0	2.0	4.0	2.0	2.0	2.0	2.0	2.0

NCOA4	3.0	2.0	1.0	1.0	1.0	2.0	2.0	2.0	4.0	2.0	1.8	2.0	2.0	2.0
NCOR1	2.0	2.0	2.0	1.4	1.7	2.0	1.7	2.0	4.0	2.0	2.0	2.0	2.0	2.0
NCOR2	1.0	1.1	2.0	2.0	2.0	3.0	3.0	2.0	>5.0	2.0	2.0	3.0	2.0	2.0
NDRG1	2.0	2.0	2.0	2.0	3.0	2.0	2.0	2.0	3.0	2.0	2.0	3.0	2.1	2.0
NF1	3.0	2.0	1.0	1.0	2.0	1.6	1.0	2.3	4.0	2.0	1.9	2.0	1.7	2.0
NF2	2.0	2.0	2.0	2.0	2.0	2.0	3.0	2.0	2.6	2.0	2.0	2.5	2.0	2.0
NFATC2	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	3.0	2.0	2.0	3.0	2.0	2.0
NFE2L2	2.0	2.0	1.0	1.0	2.0	2.0	1.0	2.0	4.0	2.0	1.8	2.0	2.0	2.0
NFIB	3.0	2.0	1.0	1.0	1.0	1.1	2.0	2.0	3.0	2.0	2.0	2.0	1.0	2.0
NFKB2	2.0	2.0	2.0	2.0	1.0	2.0	3.0	2.0	2.0	2.0	2.0	3.0	2.5	2.0
NFKBIE	2.0	2.0	2.0	2.0	2.0	2.2	2.0	2.0	2.0	2.0	2.0	3.0	3.0	2.0
NIN	2.0	2.0	1.0	1.5	2.0	2.0	1.5	2.0	4.0	2.0	2.0	2.0	2.0	2.0
NONO	1.0	2.0	2.0	2.0	2.0	2.0	3.0	2.0	3.0	2.0	2.0	2.3	3.0	2.0
NOTCH1	1.0	1.0	1.0	2.0	2.0	3.0	3.0	1.4	2.0	2.0	3.0	4.0	3.0	2.6
NOTCH2	2.0	2.0	2.0	2.0	1.2	1.1	2.0	2.0	3.6	2.0	2.0	2.0	1.5	2.0
NPM1	2.0	2.0	2.0	2.0	2.0	2.0	2.0	3.0	3.0	1.6	1.2	>5.0	2.0	2.0
NR4A3	2.0	2.0	2.0	1.0	2.0	2.0	2.0	2.0	4.0	2.0	2.0	2.0	2.0	2.0
NRAS	3.0	2.0	1.0	1.0	1.0	2.0	2.0	2.0	4.0	2.0	2.0	2.0	2.0	2.0
NRG1	2.0	2.0	2.0	2.0	2.3	2.0	2.0	2.3	4.0	2.0	1.0	2.0	2.0	2.0
NSD1	1.0	2.0	2.0	2.0	2.0	2.1	2.0	2.0	3.0	2.0	2.0	3.0	2.2	2.0
NT5C2	2.0	2.0	2.0	1.3	1.0	2.0	2.0	2.0	4.0	2.0	2.0	2.0	2.0	2.0
NTHL1	1.0	1.0	2.0	2.0	3.0	2.0	3.0	2.0	2.0	2.0	2.0	4.0	2.0	2.0
NTRK1	1.1	2.0	2.0	2.0	2.0	4.0	3.0	2.0	2.0	2.0	2.0	3.7	3.0	2.0
NTRK3	2.0	2.0	2.0	2.0	2.0	2.0	2.0	3.0	3.0	2.0	2.0	2.4	2.0	2.0
NUMA1	2.0	2.0	3.0	3.0	2.0	2.0	3.0	2.0	3.0	2.0	2.0	3.0	2.0	2.0
NUP214	2.0	2.0	2.0	2.0	1.7	2.0	>5.0	2.0	2.0	2.0	2.0	3.0	2.0	2.0
NUP98	2.0	2.0	2.0	2.2	2.0	2.0	2.0	2.0	3.7	2.0	2.0	2.2	2.0	2.0
P2RY8	1.5	2.0	2.0	2.0	2.5	3.0	3.0	2.0	2.0	2.0	2.0	4.0	3.0	2.0
PABPC1	2.0	2.0	1.0	1.5	2.0	2.0	2.0	2.0	4.0	2.0	2.0	2.0	2.0	2.0
PAFAH1B2	2.0	2.0	4.0	2.0	2.0	2.0	2.0	2.0	3.0	2.0	1.0	3.0	2.0	2.0
PALB2	2.0	2.0	1.5	1.1	1.1	2.0	2.0	2.0	4.0	2.0	2.0	3.0	2.0	2.0
PATZ1	2.0	2.0	2.0	2.0	2.0	2.0	3.0	2.0	2.7	2.0	2.0	3.0	3.0	2.0
PAX5	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	3.0	2.0	2.0	3.0	3.0	2.0
PAX7	1.0	2.0	2.0	2.0	2.0	2.8	3.0	2.0	2.0	2.0	2.0	4.0	2.0	2.0
PAX8	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	3.0	2.0	2.0	3.0	2.0	2.0
PBRM1	2.0	2.0	2.0	1.0	2.0	2.0	2.0	2.0	3.0	2.0	2.0	2.0	2.0	2.0
PBX1	2.0	2.0	1.6	1.8	2.0	2.0	2.0	2.0	3.2	2.0	2.0	2.0	2.0	2.0
PCM1	2.0	2.0	2.0	1.0	2.0	2.0	1.0	2.0	4.0	2.0	1.0	2.0	1.0	2.0
PDCD1LG2	3.0	1.0	1.0	1.0	2.0	2.0	2.0	2.0	3.0	2.0	2.0	2.0	2.0	2.0
PDE4DIP	2.0	2.0	2.0	1.6	2.0	>5.0	2.0	2.0	3.0	2.0	2.0	2.0	2.0	1.9
PDGFB	1.0	2.0	2.0	2.0	2.0	2.1	3.0	2.0	2.0	2.0	2.0	3.0	3.0	2.0
PDGFRB	2.0	2.0	2.0	2.0	2.0	2.0	2.5	3.0	2.6	2.0	2.0	3.0	2.0	2.0
PER1	1.8	2.0	2.0	2.0	2.0	2.0	3.0	2.0	2.0	2.0	2.0	3.0	3.0	2.0
PHF6	1.0	2.0	1.0	1.0	1.3	1.0	1.0	2.0	4.0	2.0	1.3	2.0	2.0	2.0
PHOX2B	2.0	2.0	1.0	1.0	2.0	1.0	2.0	2.0	4.0	2.0	2.0	2.0	1.0	2.0
PICALM	3.0	2.0	3.0	1.2	2.0	2.0	1.2	3.0	4.0	2.0	2.0	2.0	2.0	2.0
PIK3CA	3.0	2.0	2.0	1.0	2.0	2.0	1.0	2.0	4.0	2.0	2.0	2.0	2.0	2.0
PIK3CB	2.0	2.0	2.0	1.0	1.0	2.0	>5.0	2.0	4.0	2.0	2.0	2.0	2.0	2.0
PIK3R1	2.0	2.0	1.0	2.0	1.4	1.0	1.4	3.0	4.0	2.0	2.0	2.0	2.0	2.0
PIM1	2.0	2.0	2.0	2.0	2.0	2.0	2.0	1.0	2.2	2.0	2.0	3.0	2.8	2.0
PLAG1	2.0	2.0	1.0	2.0	3.0	2.0	2.0	2.0	4.0	2.0	2.0	2.0	2.0	2.0
PML	1.0	2.0	2.0	2.0	2.0	2.0	3.0	2.0	2.0	2.0	2.0	3.1	2.0	2.0
PMS1	3.0	2.0	1.0	1.0	1.0	2.0	1.0	2.0	4.0	2.0	2.0	2.0	2.0	2.0
PMS2	2.0	2.0	2.0	2.0	2.0	2.3	2.0	2.0	3.0	1.3	2.0	3.0	2.0	2.0
POLD1	1.0	1.0	2.0	2.0	3.0	3.0	3.0	2.0	2.0	2.0	2.0	3.0	3.0	2.0
POLE	1.9	1.8	2.0	2.0	2.0	3.0	3.0	2.0	2.0	2.0	2.0	3.7	2.0	2.0
POLG	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.4	3.0	2.0	2.0
POLQ	>5.0	2.0	2.0	1.0	2.0	2.0	2.0	2.0	4.0	2.0	2.0	2.0	2.0	2.0
POT1	2.0	2.0	1.0	1.1	2.0	2.0	2.0	2.8	4.0	2.0	2.0	2.0	2.0	2.0
POU2AF1	2.0	2.0	3.0	2.0	2.0	2.0	2.0	2.0	3.1	2.0	1.0	2.0	2.0	2.0
POU5F1	1.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	3.0	3.0	2.0
PPFIBP1	3.0	2.0	1.0	1.0	2.0	2.0	2.0	2.0	4.0	2.0	2.0	2.0	1.0	2.0
PPM1D	2.0	2.0	1.0	1.0	2.0	2.0	2.0	2.0	4.0	2.0	2.0	2.0	2.0	2.0
PPP2R1A	1.2	2.0	2.0	2.0	2.0	2.0	3.0	2.0	3.0	2.0	2.0	2.8	>5.0	2.0
PPP6C	2.0	2.0	2.0	1.0	2.0	2.0	2.0	2.0	3.0	2.0	2.0	2.0	2.0	2.0
PRCC	2.0	2.0	2.0	2.0	2.0	3.0	3.0	2.0	2.0	2.0	2.0	3.0	3.0	2.0
PRDM1	3.0	2.0	1.0	1.0	2.0	1.1	2.0	2.0	4.0	2.0	2.0	2.0	2.0	2.0
PRDM16	1.0	1.0	2.0	2.0	3.0	3.0	3.0	2.0	2.0	2.0	3.0	4.0	3.0	2.0

PRDM2	2.0	2.0	2.0	2.0	3.0	2.0	3.0	2.0	3.0	2.0	2.0	3.0	2.0	2.0
PREX2	2.0	2.0	1.0	1.1	3.0	2.0	2.0	2.0	4.0	2.0	2.0	2.0	2.0	2.0
PRF1	2.0	2.0	2.0	2.0	1.0	3.0	3.0	2.0	2.0	2.0	2.0	3.0	2.0	2.0
PRKACA	1.0	2.0	2.0	2.0	2.0	2.0	3.0	2.0	2.0	2.0	2.0	4.0	3.0	2.0
PRKAR1A	2.0	2.0	1.0	1.3	1.0	1.0	1.3	2.0	3.0	2.0	2.0	2.0	2.0	2.0
PRKCB	2.0	2.0	2.0	2.0	2.0	>5.0	3.0	2.0	3.0	2.0	1.4	2.1	2.0	2.0
PRKD1	2.0	2.0	1.0	1.1	2.0	2.0	1.8	2.0	4.0	2.0	2.0	2.0	1.0	2.0
PRPF40B	2.0	2.0	2.0	2.0	2.0	3.0	3.0	2.0	2.0	2.0	2.0	3.0	3.0	2.0
PRRX1	3.0	2.0	2.0	1.0	2.0	2.0	2.0	2.0	4.0	2.0	2.0	2.0	2.0	2.0
PSIP1	3.0	2.0	2.0	1.0	2.0	2.0	1.4	2.0	3.0	2.0	2.0	2.0	2.0	2.0
PTCH1	2.0	2.0	2.0	2.0	1.3	1.6	2.0	2.0	3.0	2.0	2.0	3.0	2.0	2.0
PTEN	3.0	2.0	1.0	1.0	1.0	1.2	1.0	2.0	3.0	2.0	1.8	2.0	2.0	2.0
PTK6	1.0	1.0	2.0	2.0	3.0	3.0	4.0	2.0	2.0	2.0	3.0	4.0	3.0	2.0
PTPN11	2.0	2.0	2.0	2.0	1.1	3.0	2.0	2.0	3.0	2.0	1.2	3.0	2.0	2.0
PTPN13	2.7	2.0	1.0	1.0	2.0	2.0	2.0	2.0	4.0	2.0	2.0	2.0	1.2	2.0
PTPN6	2.0	1.0	2.0	2.0	2.0	2.0	2.0	1.0	2.0	2.0	2.0	3.0	3.0	2.0
PTPRB	>5.0	2.0	1.0	1.0	1.7	2.0	2.0	2.0	4.0	2.0	2.0	2.0	1.0	2.0
PTPRC	3.0	2.0	1.0	1.0	2.0	>5.0	1.0	2.0	4.0	2.0	1.7	2.0	2.0	2.0
PTPRD	3.0	1.5	1.0	1.0	1.7	2.0	2.0	2.0	3.0	2.0	2.0	2.0	2.0	2.0
PTPRK	3.0	2.0	1.0	1.0	2.0	1.1	1.6	2.0	4.0	2.0	1.9	2.0	1.0	2.0
PTPRT	2.0	2.0	2.0	2.0	1.9	2.0	2.0	2.0	3.0	2.0	2.0	2.9	2.0	2.0
PWWP2A	2.0	2.0	1.2	2.0	2.0	2.0	2.0	3.0	4.0	2.0	2.0	2.0	2.0	2.0
QKI	3.0	2.0	1.0	2.0	2.0	2.0	2.0	2.0	3.0	2.0	2.0	2.0	2.0	2.0
RABEP1	2.0	2.0	2.0	1.0	2.0	2.0	5.0	2.0	4.0	2.0	2.0	2.0	5.0	2.0
RAC1	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	3.0	2.0	2.0	3.0	2.3	2.0
RAD17	>5.0	2.0	2.0	2.0	2.0	2.0	2.0	3.0	4.0	2.0	2.0	2.0	2.0	2.0
RAD21	2.0	2.0	1.0	2.0	2.0	2.0	2.0	2.0	4.0	2.0	2.0	2.0	2.0	2.0
RAD51B	2.0	2.0	1.0	2.0	2.0	2.0	2.0	2.0	3.0	2.0	1.8	2.0	2.0	2.0
RAF1	1.0	2.0	2.0	2.0	2.0	2.0	>5.0	2.0	3.0	2.0	2.0	2.0	2.0	2.0
RALGDS	1.0	1.0	1.0	2.0	2.0	2.0	3.0	2.0	2.0	2.0	2.8	3.6	2.3	2.0
RANBP2	3.0	2.0	2.0	1.0	2.0	2.0	1.0	2.0	4.0	2.0	2.0	2.0	2.0	2.0
RAP1GDS1	2.0	2.0	1.0	1.0	2.0	1.0	2.0	2.0	4.0	2.0	2.0	2.0	2.0	2.0
RARA	2.0	2.0	1.7	2.0	2.0	2.0	3.0	2.0	3.0	2.0	2.0	3.0	2.0	2.0
RB1	3.0	3.0	3.0	1.0	2.0	1.1	1.0	2.0	4.0	2.0	2.0	2.0	2.0	2.0
RBM10	1.0	2.0	2.0	2.0	2.0	2.0	3.0	1.8	2.0	2.0	2.0	3.0	3.0	2.0
RBM15	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	3.0	2.0	2.0	3.0	2.0	2.0
RECQL4	1.0	1.0	2.0	2.0	3.0	3.0	3.0	1.0	2.0	2.0	2.0	4.0	3.0	2.0
REL	3.0	2.0	1.0	1.0	2.0	2.0	2.0	2.0	4.0	2.0	2.0	2.0	2.0	2.0
RET	2.0	2.0	2.0	2.0	1.0	2.7	2.9	2.0	2.1	2.0	2.0	3.8	2.0	2.0
RFWD3	2.0	2.0	2.0	1.3	1.8	2.0	2.8	2.0	3.0	2.0	2.0	2.8	2.0	2.0
RGPD3	3.0	2.0	2.0	1.0	2.0	2.0	2.0	2.6	4.0	2.0	2.0	2.0	2.0	2.0
RGS7	3.0	2.0	1.7	1.0	2.0	3.0	2.0	2.3	4.0	2.0	2.0	2.0	1.0	2.0
RHOA	1.0	2.0	2.0	2.0	2.0	2.0	3.0	1.7	2.0	2.0	2.0	3.0	2.3	2.0
RHOH	2.0	2.0	1.0	1.0	2.0	2.0	2.0	2.0	3.0	2.0	2.0	2.0	2.0	2.0
RMI2	2.0	2.0	2.0	2.0	2.0	3.0	3.0	2.0	3.0	2.0	2.0	3.0	3.0	2.0
RNF213	1.9	2.0	2.0	2.0	1.8	>5.0	2.4	2.0	2.0	2.0	2.0	3.2	2.1	2.0
RNF43	2.0	2.0	1.0	2.0	2.0	3.0	3.0	2.0	2.0	2.0	2.0	3.0	2.9	2.0
ROBO2	2.0	2.2	1.1	1.0	1.0	1.8	2.0	2.0	4.0	2.0	2.0	2.0	1.5	2.0
ROS1	3.0	2.0	1.0	1.0	2.0	2.0	2.0	2.0	4.0	2.0	2.0	2.0	1.0	2.0
RPL10	1.0	1.0	2.0	2.0	2.0	2.0	3.0	1.0	2.0	2.0	2.3	4.0	3.0	2.0
RPL22	1.0	1.0	2.0	2.0	3.0	3.0	3.0	2.0	2.0	2.0	2.3	4.0	3.0	2.0
RPL5	3.0	2.0	2.0	1.0	2.0	2.0	1.0	2.0	4.0	2.0	2.0	2.0	2.0	2.0
RPN1	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.7	2.0	2.0	3.0	2.0	2.0
RSPO2	2.0	2.0	1.0	2.0	2.0	2.0	2.0	2.0	4.0	2.0	2.0	2.0	2.0	2.0
RSPO3	3.0	2.0	1.0	1.0	2.0	1.0	1.0	2.0	4.0	2.0	2.0	2.0	1.0	2.0
RUNX1	2.0	2.0	3.0	2.0	2.0	2.0	2.0	2.0	4.0	2.0	2.0	3.0	2.0	2.0
RUNX1T1	2.0	2.0	1.0	2.0	3.0	2.0	2.0	2.0	4.0	2.0	2.0	2.0	2.0	2.0
S100A7	2.0	2.0	2.0	2.0	2.0	3.0	3.0	2.0	2.0	2.0	2.0	3.0	2.0	2.0
SALL4	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	3.0	2.0	2.0	3.0	2.0	2.0
SBDS	1.8	2.0	2.2	2.2	1.3	2.0	2.2	2.0	2.8	2.0	1.3	2.2	2.2	2.0
SDHA	2.0	2.0	2.0	3.0	2.0	2.0	3.0	3.0	3.0	2.0	2.0	3.0	2.0	2.0
SDHB	2.0	2.0	2.0	2.0	2.1	2.0	3.0	2.0	2.0	2.0	2.1	3.0	2.0	2.0
SDHC	2.0	2.0	2.0	2.0	1.5	3.0	2.0	2.0	3.0	2.0	1.2	3.0	3.0	2.0
SET	2.0	2.0	2.0	1.5	2.0	2.0	2.0	2.0	2.0	2.0	2.0	3.0	2.0	2.0
SETBP1	3.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	4.0	2.0	2.0	2.0	2.0	2.0
SETD1B	2.0	1.9	2.0	2.0	2.0	3.0	3.0	2.0	2.0	2.0	2.0	3.8	3.0	2.0
SETD2	1.0	2.0	2.0	1.4	2.0	2.0	3.0	2.0	2.0	2.0	1.1	3.0	>5.0	2.0
SETDB1	2.0	2.0	2.0	2.0	2.0	3.0	2.9	2.0	3.0	2.0	2.0	2.0	3.0	2.0

SF3B1	2.9	2.0	1.0	1.0	2.0	2.0	1.5	2.0	4.0	2.0	2.0	2.0	2.0	2.0
SFPQ	2.0	2.0	2.0	1.0	2.0	2.0	2.0	2.0	4.0	2.0	2.0	2.0	2.0	2.0
SGK1	3.0	2.0	1.0	1.0	2.0	1.2	2.0	2.0	4.0	2.0	2.0	2.0	2.0	2.0
SH2B3	1.0	2.0	2.0	2.0	2.0	3.0	2.2	2.0	2.3	2.0	2.0	3.0	3.0	2.0
SH3GL1	1.0	1.0	2.0	2.0	3.0	3.0	3.0	2.0	2.0	2.0	2.0	4.0	3.0	2.0
SIX1	2.0	2.0	1.0	2.0	2.0	2.0	2.0	2.0	4.0	2.0	2.0	2.0	2.0	2.0
SIX2	2.0	2.0	1.0	1.0	2.0	2.0	2.0	2.0	4.0	2.0	2.0	2.0	2.0	2.0
SKI	1.0	1.0	2.0	2.0	3.0	3.0	3.0	2.0	2.0	2.0	3.0	4.0	3.0	2.0
SLC34A2	2.0	2.0	2.0	2.0	1.2	1.2	2.0	2.0	4.0	2.0	2.0	3.0	1.0	2.0
SLC45A3	2.0	2.0	2.0	2.0	2.0	3.0	2.0	2.0	3.0	2.0	2.0	3.0	2.0	2.0
SMAD2	3.0	2.0	1.0	1.1	2.0	2.0	1.1	2.0	3.0	2.0	2.0	2.0	1.1	2.0
SMAD4	3.0	2.0	2.0	1.2	2.0	2.0	2.0	2.0	3.0	2.0	2.0	2.0	2.0	2.0
SMARCA4	1.0	1.3	2.0	2.0	2.1	2.3	3.0	2.0	2.0	2.0	2.6	4.0	3.0	2.0
SMARCB1	1.0	2.0	2.0	2.0	2.0	2.0	3.0	2.0	2.0	2.0	2.0	3.0	3.0	2.0
SMARCD1	2.0	2.0	2.0	2.0	2.0	>5.0	2.0	2.0	2.0	2.0	2.0	3.0	2.0	2.0
SMARCE1	2.0	2.0	2.0	1.1	2.0	2.0	2.2	2.0	3.0	2.0	2.0	3.0	2.0	2.0
SMC1A	1.0	2.0	2.0	2.0	2.0	2.0	>5.0	2.0	3.0	2.0	2.0	3.0	2.0	2.0
SMO	1.0	2.0	1.0	2.0	2.0	2.0	2.0	2.0	3.0	2.0	2.0	3.0	2.0	2.0
SND1	2.0	2.0	1.0	2.0	1.8	2.0	2.0	2.0	3.0	2.0	2.0	3.0	2.0	2.0
SOCS1	2.0	2.0	2.0	2.0	2.0	3.0	3.0	2.0	3.0	2.0	2.0	3.0	3.0	2.0
SOX21	2.0	2.0	2.0	1.0	2.0	2.0	2.0	2.0	4.0	2.0	2.0	2.0	2.0	2.0
SPECC1	2.0	2.0	2.0	2.0	2.0	2.0	3.0	2.0	3.0	2.0	2.0	2.0	2.0	2.0
SPEN	2.0	2.0	2.0	2.0	2.0	2.0	3.0	2.0	2.9	2.0	2.0	3.1	2.0	2.0
SPOP	2.0	2.0	1.0	1.1	1.0	2.0	2.0	2.0	3.1	2.0	1.8	2.0	2.0	2.0
SRGAP3	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	3.0	2.0	2.0	2.8	2.0	2.0
SRSF2	1.0	2.0	2.0	1.0	2.0	2.0	2.0	2.0	3.0	2.0	2.0	3.0	2.0	2.0
SS18	3.0	2.0	1.0	1.0	2.0	2.0	2.0	2.0	4.0	2.0	2.0	2.0	1.2	2.0
SS18L1	1.0	1.1	2.0	2.0	2.0	2.7	4.0	2.0	2.0	2.0	2.0	4.0	3.0	2.0
SSX1	1.0	2.0	3.0	2.0	2.0	2.0	3.0	2.0	3.0	2.0	1.0	3.0	2.0	2.0
SSX2	1.0	2.0	2.5	2.0	2.0	2.0	3.0	2.0	3.0	2.0	2.0	3.0	2.0	2.0
SSX4	1.0	2.0	3.0	2.0	2.0	3.0	3.0	2.0	3.0	2.0	1.6	3.0	2.0	2.0
STAG1	3.0	2.0	2.0	1.0	2.0	2.0	2.0	2.0	4.0	2.0	2.0	2.0	2.0	2.0
STAG2	1.0	2.0	2.0	1.0	2.0	2.0	1.3	2.0	4.0	2.0	2.0	2.0	2.0	2.0
STAT3	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	3.0	>5.0	2.0	3.0	2.0	2.0
STAT5B	1.0	2.0	2.0	2.0	2.0	2.0	3.0	2.0	2.0	2.0	2.0	3.0	3.0	2.0
STAT6	2.0	2.0	2.0	2.0	2.0	3.0	3.0	2.0	2.0	2.0	2.0	3.0	2.7	2.0
STIL	2.0	2.0	2.0	1.0	1.1	2.0	2.0	2.0	3.0	2.0	2.0	>5.0	2.0	2.0
STK11	1.0	1.0	2.0	2.0	3.0	3.0	4.0	2.0	2.0	2.0	2.0	4.0	3.0	2.0
STRN	2.0	2.0	2.0	1.0	2.0	2.0	2.0	2.0	4.0	2.0	2.0	2.0	2.0	2.0
SUFU	2.0	2.0	2.0	2.0	1.0	2.0	2.0	2.0	2.0	2.0	2.0	3.0	2.0	2.0
SUZ12	2.0	2.0	2.0	1.0	2.0	2.0	2.0	2.0	4.0	2.0	2.0	3.0	2.0	2.0
SYK	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	3.0	2.0	2.0	3.0	2.0	2.0
TAF15	2.0	2.0	2.0	1.4	2.0	2.0	1.4	2.0	3.0	2.0	1.3	2.1	2.0	2.0
TAL1	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	3.0	2.0	2.0	3.0	2.0	2.0
TAL2	3.0	2.0	1.0	1.0	2.0	2.0	2.0	2.0	4.0	2.0	2.0	2.0	2.0	2.0
TBL1XR1	3.0	2.0	2.0	1.0	1.9	1.0	1.0	2.0	4.0	2.0	2.0	2.0	1.9	2.0
TBX3	2.0	2.0	1.6	2.0	1.4	2.3	2.0	2.0	3.0	2.0	2.0	2.3	2.0	2.0
TCEA1	2.0	2.0	1.0	2.0	3.0	2.0	2.0	2.0	4.0	2.0	2.0	2.0	2.0	2.0
TCF12	2.0	2.0	2.0	1.0	1.7	1.8	2.0	3.0	4.0	2.0	2.0	>5.0	1.0	2.0
TCF3	1.0	1.0	2.0	2.0	3.0	3.0	4.0	2.0	2.0	2.0	2.0	4.0	3.0	2.0
TCF7L2	3.0	2.0	2.0	2.0	1.0	2.0	2.0	2.0	3.0	2.0	2.0	2.0	2.0	2.0
TEC	3.0	2.0	1.5	1.0	2.0	1.0	2.0	2.0	4.0	2.0	2.0	2.0	1.0	2.0
TERT	1.0	1.0	2.0	3.0	2.0	2.9	3.0	2.0	2.0	2.0	2.0	4.0	2.0	2.0
TET1	3.0	2.0	2.0	1.0	1.0	2.0	2.0	2.0	4.0	2.0	2.0	2.0	2.0	2.0
TET2	3.0	2.0	1.3	1.0	2.0	1.0	2.0	2.0	4.0	2.0	2.0	2.0	1.0	2.0
TFE3	1.0	2.0	2.0	2.0	2.0	2.0	3.0	2.0	2.0	2.0	2.0	3.0	3.0	2.0
TFEB	2.0	2.0	2.0	2.0	2.0	2.1	3.0	2.0	2.0	2.0	2.0	3.0	3.0	2.0
TFG	3.0	2.0	1.0	1.0	2.0	2.0	2.0	2.0	4.0	2.0	2.0	2.0	2.0	2.0
TFPT	1.0	2.0	2.0	2.0	2.0	2.0	3.0	2.0	2.0	2.0	2.0	4.0	3.0	2.0
TFRC	2.0	2.0	2.0	1.2	2.0	2.8	2.4	2.0	2.9	2.0	2.0	3.0	2.0	2.0
TGFBR2	2.0	2.0	1.0	1.0	2.0	2.0	2.0	2.0	4.0	2.0	2.0	2.0	2.0	2.0
THRAP3	2.0	2.0	2.0	2.0	2.0	2.0	3.0	2.0	3.0	2.0	2.0	3.0	2.0	2.0
TLX3	2.0	2.0	2.0	2.0	2.0	2.0	2.0	3.0	3.0	2.0	2.0	2.0	2.0	2.0
TMEM127	2.0	2.0	2.0	2.0	2.0	2.0	3.0	2.0	2.0	2.0	2.0	3.0	3.0	2.0
TMSB4X	1.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	4.0	2.0	2.0	2.0	2.0	2.0
TNC	2.0	2.0	1.4	2.0	2.0	2.0	2.5	2.0	3.0	2.0	2.0	2.0	2.0	2.0
TNFAIP3	3.0	2.0	2.0	2.0	2.0	1.0	2.0	2.0	4.0	2.0	2.0	2.0	2.0	2.0
TNFRSF14	1.0	1.0	2.0	2.0	3.0	3.0	3.0	2.0	2.0	2.0	3.0	4.0	3.0	2.0

TOP1	2.0	2.0	1.0	1.5	2.0	2.0	2.0	2.0	3.0	2.0	2.0	2.0	2.0	2.0
TP53	2.0	2.0	2.0	2.0	2.0	2.0	3.0	2.0	2.0	2.0	2.0	3.0	3.0	2.0
TP63	2.0	3.0	1.0	2.0	2.0	2.0	2.0	2.0	4.0	2.0	2.0	2.0	2.0	2.0
TPM3	2.0	2.0	2.0	2.0	2.0	3.0	2.0	2.0	3.0	2.0	1.4	2.0	2.9	2.0
TPR	3.0	2.0	1.3	1.0	2.0	2.0	1.0	2.8	4.0	2.0	2.0	2.0	1.0	2.0
TRAF7	1.0	1.0	2.0	2.0	3.0	3.0	3.0	2.0	2.0	2.0	2.8	4.0	3.0	2.3
TRIM24	2.0	2.0	1.0	1.2	2.0	2.0	2.0	2.0	4.0	2.0	2.0	2.0	2.0	2.0
TRIM33	3.0	2.0	1.6	1.0	2.0	2.0	2.0	2.0	4.0	2.0	2.0	2.0	2.0	2.0
TRIP11	3.0	2.0	1.0	1.0	2.0	1.5	1.1	2.0	4.0	2.0	2.0	>5.0	2.0	2.0
TRRAP	2.0	2.0	2.0	2.0	2.0	2.5	3.0	2.0	3.0	2.0	2.0	3.0	>5.0	2.0
TSC1	2.0	2.0	1.0	2.0	1.0	2.0	2.0	2.0	3.0	2.0	2.0	3.0	2.0	2.0
TSC2	1.0	1.0	2.0	2.0	3.0	2.0	3.0	2.0	2.0	2.0	2.6	4.0	2.0	2.7
UBR5	2.0	2.0	1.0	1.0	2.1	2.0	1.7	2.0	4.0	2.0	2.0	2.0	2.0	2.0
USP44	3.0	2.0	2.0	1.0	1.6	2.0	2.0	2.0	4.0	2.0	2.0	2.0	2.0	2.0
USP6	2.0	2.0	2.0	1.5	1.7	2.0	2.6	2.0	2.6	2.0	2.0	3.0	2.0	2.0
USP8	3.0	2.0	2.0	1.0	2.0	2.0	2.0	3.0	4.0	2.0	1.5	2.0	2.0	2.0
USP9X	1.9	2.0	1.0	1.0	1.7	2.0	1.7	2.0	4.0	2.0	2.0	2.0	2.0	2.0
VAV1	1.0	2.0	2.0	2.0	2.6	3.0	3.0	2.0	2.0	2.0	2.0	4.0	3.0	2.0
VHL	1.0	2.0	2.0	2.0	2.0	2.0	3.0	2.0	2.0	2.0	2.0	3.0	3.0	2.0
VTI1A	3.0	2.0	2.0	1.1	1.0	2.0	2.0	2.0	3.8	2.0	2.0	2.0	2.0	2.0
WAS	1.0	2.0	2.0	2.0	2.0	2.0	3.0	1.4	2.0	2.0	2.0	3.0	3.0	2.0
WIF1	3.0	2.0	1.0	1.0	1.2	2.0	2.0	2.0	4.0	2.0	2.0	2.0	1.0	2.0
WNK2	1.0	2.0	2.0	2.0	2.0	2.0	3.0	2.0	2.0	2.0	2.0	4.0	2.9	2.0
WRN	2.0	2.0	1.0	1.0	2.0	1.4	1.0	3.0	4.0	2.0	1.0	2.0	1.4	2.0
WT1	2.0	2.0	2.0	2.0	2.0	2.0	2.0	3.0	4.0	2.0	2.0	2.0	2.0	2.0
WWTR1	2.0	2.0	1.0	1.9	2.0	1.1	2.0	2.0	4.0	2.0	2.0	2.0	2.0	2.0
XPA	2.0	2.0	1.0	1.0	2.0	2.0	1.1	2.0	3.0	2.0	2.0	2.0	2.0	2.0
XPC	1.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	3.0	2.0	2.0
XPO1	3.0	2.0	1.5	1.0	2.0	2.0	1.0	2.0	4.0	2.0	2.0	2.0	2.0	2.0
YWHAE	1.8	1.5	2.0	1.2	2.0	2.0	3.0	2.0	2.0	2.0	1.2	3.0	3.0	2.0
ZBTB16	2.0	2.0	4.0	2.0	2.0	2.0	2.0	2.0	3.0	2.0	1.0	3.0	2.0	2.0
ZCCHC8	2.0	2.0	1.0	1.1	2.0	2.1	1.4	2.0	3.9	2.0	2.0	2.0	2.0	2.0
ZEB1	3.0	2.0	1.0	1.0	1.0	2.0	1.5	2.0	3.8	2.0	2.0	2.0	2.0	2.0
ZFHX3	2.0	2.0	2.0	2.0	2.0	2.0	3.0	2.0	3.0	2.0	2.0	3.0	2.0	2.0
ZMYM2	3.0	2.0	2.0	1.0	2.0	1.6	1.1	2.0	4.0	2.0	2.0	2.0	2.0	2.0
ZMYM3	1.0	2.0	2.0	2.0	2.0	2.0	3.0	2.0	2.5	2.0	2.0	3.0	2.8	2.0
ZNF331	2.0	2.0	3.0	2.0	2.0	3.0	3.0	2.0	3.0	2.0	2.0	4.0	2.0	2.0
ZNF384	2.0	2.0	2.0	2.0	2.0	3.0	2.0	1.0	2.0	2.0	2.0	3.0	3.0	2.0
ZNF429	2.0	2.0	2.0	1.0	2.0	3.0	2.0	2.0	3.0	2.0	2.0	2.0	2.0	2.0
ZNF479	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	3.0	2.0	2.0	3.0	2.0	2.0
ZNF521	3.0	2.0	1.5	1.0	2.0	2.0	2.0	2.0	4.0	2.0	2.0	2.0	1.0	2.0
ZNRF3	2.0	2.0	2.0	2.0	1.9	2.0	3.0	2.0	3.0	2.0	2.0	3.0	3.0	2.0
ZRSR2	1.0	2.0	2.0	1.0	2.0	2.0	2.0	2.0	4.0	2.0	2.0	2.0	2.0	2.0

Table 3-15. Copy number annotation details for the exome cohort.

List of derived copy number values per sample, filtered to include genes with known somatic driver mutations in human cancers (curated by the COSMIC database)

Cell Line	Patient Age at model creation
NCI-H292	32
UM-SCC-104	56
MDA8788-6	56
UM-SCC-47	54

Table 3-16. Age of Patient When Model Was Derived.

Chapter 4 Transcriptomic Analysis defines *CRTC1-MAML2* and Grade-Associated Molecular and Immune Profiles in Mucoepidermoid Carcinoma

4.1 Abstract

Mucoepidermoid carcinoma is one of the most common malignancies of the salivary gland, of which about 50% are known to harbor *CRTC1/3-MAML2* gene fusions. The majority of MECs respond well to conventional therapy, but 15-20% of patients will recur and die of disease. Unfortunately, there are no known molecular markers to predict aggressive disease and the somatic genetic alterations remain largely uncharacterized. Here, we perform RNA sequencing on 48 MEC tumors, discovering gene expression patterns that are associated with *CRTC1/3-MAML2* fusion status and tumor grade. These gene expression patterns affect the molecular etiology of disease and changes to the tumor microenvironment, including T and B cell immune infiltration. We therefore proceeded to measure spatially resolved changes in gene expression using spatial RNA sequencing, identifying a pattern of spatial overlap between *CRTC1-MAML2* associated gene expression and several other transcripts, including *VEGFA* and *CTNNB1*, which encodes the beta-catenin protein. Taken together, these data support *CRTC1-MAML2*-mediated changes to gene expression affecting MEC tumor phenotypes throughout the tumor microenvironment.

4.2 Introduction

Mucoepidermoid carcinoma (MEC) is one of the most common malignancies of the salivary glands representing between 30-40% of all cases (117-119). The disease most commonly arises in the parotid gland, but occasionally will form in the submandibular gland and minor salivary glands (120). MEC is primarily treated surgically with adjuvant radiation used in advanced disease in frontline settings (184). In general, survival is favorable for most cases and can be stratified by histologic grade, tumor location, tumor stage, nodal status, patient age, margin status, and perineural invasion (118, 119, 121); however, it remains extremely challenging to differentiate between aggressive and non-aggressive MEC.

At the molecular level, a series of highly recurrent genetic alterations in MEC that lead to the translocation of *CRTC1* to *MAML2*, ((11;19)(q14-21; p12-13)), or *CRTC3* to *MAML2* (70, 124, 126, 129), (130) has been suggested to have a potential prognostic role in this disease (129, 185). This suggests that molecular features may play an important role in improving our ability to discriminate between lethal and non-lethal disease, but a more nuanced understanding of the molecular networks associated with fusion status may be required to better understand the disease process. The resultant *CRTC1-MAML2* and *CRTC3-MAML2* fusion genes are known to play a critical role in regulating in vitro oncogenic phenotypes, including transformation. Early studies evaluated oncogenic pathways regulated by CREB and NOTCH, transcription factors activated by wild type CRTC1 and MAML2 (70, 127). Later work has determined that CRTC1-MAML2 lacks the MAML2 NOTCH binding domain, and therefore does not affect NOTCH signaling (73, 74, 132). Therefore, MEC oncogenic phenotypes have been found to be primarily regulated by a CREB-mediated pathway, and possibly also through novel MYC and/or IGF1-dependent mechanisms (72-74, 128, 131, 132). However, further studies are needed to identify

other transcription factor effectors of CRTC1-MAML2 activity, and to characterize the effects these broadly reaching transcriptional changes have on MEC etiology.

In addition to remaining questions about how to improve MEC prognostication, there also continue to be outstanding questions regarding how to advance novel immunotherapies in this space. To date, most knowledge about MEC immune landscape centers PD-L1 expression. 63-71% of MEC tumors express PD-L1, and in pan-salivary gland cancer combined analysis, this PD-L1 expression negatively affects prognosis (186, 187). However, the two pan-salivary gland phase II clinical trials to date have included too few MEC tumors to determine the efficacy of PD-L1 inhibition (MEC N= 3, 3, respectively) (188, 189). It is unknown which other checkpoint inhibitors may be the best to advance for clinical studies because the immune landscape of this disease is mostly uncharacterized. Given these open questions, in this study, we sought to use whole transcriptome sequencing analysis of a cohort of MEC tumors to both define gene sets associated with *CRTC1/3-MAML2* and to help de-convolute the immune composition of this disease.

Beyond these molecular questions, MEC is also known to have an extremely high level of cellular heterogeneity, as it is comprised of mucosal, epidermoid and intermediate cell types. Here, we hypothesize that understanding the molecular composition of the cell types that comprise MEC, including the identification of the cell type that harbors recurrent *CRTC1-MAML2* alterations (as this remains unknown), would provide pivotal information to facilitate improved prognostic strategies for this disease. We therefore sought to spatially resolve gene expression, CRTC1-MAML2 transcription factor activity, and cell type localization in MEC tumors.

4.3 Materials and Methods

4.3.1 Clinical Specimens and clinical data

A retrospective cohort of patients with MEC was identified from the University of Michigan pathology archive using a protocol IRB-approved for next generation sequencing of DNA and RNA (HUM00080561), though patients were not consented for deposit of data in public databases. The cohort was previously described and typed for gene fusion status by RT-qPCR (98). As previously noted, clinical, histologic, and outcome data was collected from medical records and death was documented for electronic medical record notes and the Social Security Death Index. In total, we identified 90 MEC patients in this cohort that were collected from 1998-2015. DNA and RNA were simultaneously isolated from these patients' FFPE-preserved primary tumors using the Qiagen AllPrep kit as described (98) and advanced for NGS if it met our previously defined quality standards defined by Qubit and Bioanalyzer analysis (4, 99).

4.3.2 Transcriptome Sequencing

Total RNA was submitted to the University of Michigan Advanced Genomics core for library preparation and sequencing. Briefly, if available, up to 500ng of RNA was used for library preparations with the Illumina TruSeq Stranded Total RNA library prep kit (Cat#: RS-122-2201/2) according to the manufacturer's recommendations, with a modification that 14 cycles of PCR were performed to amplify the library prior to the final bead purification. The samples were then loaded on to a total of 12 lanes of an Illumina HiSEQ4000 and paired-end

sequenced to 75nt length in each direction. A summary of sequencing quality statistics including total unique mapped reads for each sample is provided in **Table 4-1**.

4.3.3 Transcriptome Quantification

Read quality was assessed using FastQC (v0.11.5). No quality issues were detected. The two-step STAR workflow was used to map the reads. In step 1, STAR (v2.5.3a) was used to generate the genome index database with the help of the reference human genome hg19 and annotated transcriptome files. In step 2, read mapping was guided by this generated genome index database generated in step 1. Only reads that map uniquely were retained by using SAMtools (v1.2). To compute FPKM, Cufflinks (v2.2.1) was used with default parameters except for "--max-bundle-frags" which was changed to 100000000 to avoid raising of the HIDATA flag at loci that have more fragments than the pre-set threshold for every locus.

4.3.4 Correlation of gene sets with clinical variables

Before the analysis, genes with average FPKM less than 1 were removed. Patient sex and tumor fusion status were treated as binary variables. For these two binary variables, logistic regression and two-sample t-test were used. Both pooled variance and Welch variance were tried for two-sample t-tests. Grade was treated as an ordinal factor. We denote "Low" as 0, "Intermediate" as 1, "High" as 2. Ordinal logistic regression was fitted for Grade. In addition, Grade is also treated as a continuous covariate and fitted with linear regression. For clinical stage, we code "I" as 1, "II" as 2, "III" as 3 and "IV" as 4. Both ordinal logistic regression and linear regression were fitted with clinical stage. The above procedure was also repeated with the

log FPKM. Because multiple genes were tested simultaneously, false discovery rate (FDR) method was used to adjust p-values.

4.3.5 Gene set enrichment

Gene Set Enrichment Analysis (GSEA) was used to identify enrichment of CRTC1-MAML2 fusion-dependent gene expression in eight data sets (190). The gene sets used are as follows: two sets of genes associated with the CRTC1-MAML2 fusion in human MEC cells (71, 73); genes associated with LINC00473, a CRTC1-MAML2 fusion induced lncRNA (175); genes upregulated and genes downregulated in MOLT-4 cells after NOTCH inhibition (191); genes upregulated and genes downregulated in cultured breast cancer cells after Myc overexpression (192); and genes in the neighborhood of Myc (http://software.broadinstitute.org/gsea/msigdb/cards/MORF_MYC.html). Read counts (FPKM) of fusion-positive and fusion-negative samples were used as input. An FDR q-value of 0.1 was used as a cutoff for statistical significance.

Using the weighted correlation network analysis (WGCNA) R package, we assigned a gene co-expression network to each clinical characteristic (193). WGCNA clusters co-expressed transcripts, using a weighted correlation network analysis, which is unbiased to sample phenotype. All transcripts with a mean read count greater than 1 FPKM across all samples were included in clustering. Samples were clustered according to transcription patterns to remove any outlier samples, and two samples were removed. Using a soft thresholding power of eight, we generated a co-expression network. This network clusters co-expressed genes into modules, which were assigned arbitrary color names. The expression of these modules was then associated with the sex, fusion status, clinical stage (ranked 1-4), and tumor grade (ranked 0-2). Pearson

pairwise correlation coefficient and p value, calculated by Student's t test, are reported for each module-trait combination.

4.3.6 Annotation of Viral Genomes, Gene Fusion Analysis and Immune Annotation

Because gene fusions were already identified in these samples, FusionCatcher was used as an informatics method to identify and validate these fusion transcripts (194). We required a minimum of 3 unique spanning reads and longest anchor of at least 20 nucleotides. FusionCatcher also calls the presence of viruses in samples. Immune cell infiltrate content from bulk RNAseq data was assessed using the CIBERSORT algorithm (195).

4.3.7 Immunohistochemistry

Immunohistochemical staining was performed on the DAKO Autostainer (Agilent, Carpinteria, CA) using Envision+ and diaminobenzidine (DAB) as the chromogen. De-paraffinized sections were labeled with the antibodies listed in **Table 2** for 30 minutes at ambient temperature. Microwave epitope retrieval as specified was used prior to staining for all antibodies. Appropriate negative (no primary antibody) and positive controls (as listed) were stained in parallel with each set of slides studied. SHH immunostained slides were analyzed using a modified Allred scoring system with intensity reported as negative (0), weak (1), moderate (2) or strong (3) and the percentage of tumor staining. Tumor infiltrating CD8+ or CD20+ cells were scored as previously described by our team (16, 17).

4.3.8 Spatial RNA sequencing

Starting with FFPE preserved tissue, tumor regions of seven MEC tumors and one normal salivary gland control were annotated by an experienced pathologist (J.B.M.). Samples were then prepared for sequencing, including tissue regions with the highest tumor content. Two replicates were prepared for each sample sequenced. Visium spatial RNA sequencing was performed at the University of Michigan's Advanced Genomics Core.

4.3.9 Spatial RNA sequencing analysis

Spatial RNA sequencing results were analyzed using Seurat (v 4.0.6) (196). Briefly, sequencing results were normalized by SCT transformation, followed by UMAP dimensional reduction and clustering. Spatial feature plots were generated using SCT transformed expression levels for each sample.

4.3.10 Signature correlation analysis

Gene wise linear regression was performed using the SCENIC R package (v 1.1.2) (197). First, using a table of gene expression per sequencing spot, genes were filtered to exclude genes expressed in less than 1% of all spots, and genes with a read count less than 0.03 times the number of spots in the sample (197). Then, a pairwise Spearman correlation was performed, resulting in a matrix of Spearman correlation coefficients (R) for each pair of genes included in the analysis.

For following steps, we used a gene signature associated with CRTC1-MAML2 fusion status in primary MEC tumors, generated by bulk RNA sequencing. Generating this signature is

described above. We then calculated a summary pseudo-correlation coefficient to quantify the gene expression correlation between the fusion associated gene signature with each gene in the analysis. We started with a matrix of Spearman R values, showing the gene expression correlation between each pair of genes within the spots of each sample. Next, we filtered this matrix to report the correlation coefficients of only the fusion-associated signature for each gene. We then calculated the geometric mean of all fusion signature Spearman R values per gene. to determine a gene signature pseudo-correlation coefficient. This method was repeated for all samples sequenced. Genes that moderately correlate with the fusion-associated gene signature (mean R>0.3) in four or more tumor samples were considered for downstream analysis. No genes met these criteria in the healthy control.

4.3.11 Spatial Deconvolution Analysis

We used spatial deconvolution to estimate the relative abundance of cell types in our spatial RNA sequenced samples. We performed spatial deconvolution with the R package SpatialDecon (v3.16) (198). We defined cell types using a cell profile matrix, derived from single cell RNA sequencing data. For this reference, we used data from 12 healthy parotid salivary glands, downloaded from the Tabula Sapiens database (199). We then performed deconvolution using default parameters.

4.3.12 RNAscope

We visualize RNA transcripts of interest via RNAscope, using the multiplex fluorescent detection kit (Advanced Cell Diagnostics; Newark, CA, USA). To hybridize, we prepared a probe mixture 48 parts RNAscope probe Hs-LINC00473, 1 part probe Hs-EPCAM-C2, and one

part either probe Hs-VEGFA-C3 or probe Hs-CTNNB1-C4 (Catalog 464821, 310281-C2, 423161-C3, 311731-C4), and used the secondary stains TSA-Cy3, TSA-Cy5, and TSA-Fluorescein, respectively (Akoya Biosciences; Marlborough, MA, USA). Slides were stored at 4°C prior to being imaged with the Leica Stellaris 5 confocal microscope at the University of Michigan Microscopy Core. Probe-positive cells are quantified using QuPath (v0.4.2) (200)

4.4 Results

We performed comprehensive transcriptome sequencing on 48 formalin fixed paraffin embedded (FFPE) MEC tissues with a majority of tumors arising in the parotid. Clinical features associated with the cohort were broken down by CRTC1/3-MAML2 fusion status (**Table 3**). Overall, had 32 fusion positive MECs and 16 fusion negative MECs that successfully passed or quality metrics for analysis. We generated an average of 69,358,792 reads per sample of which an average of 82.99% uniquely mapped to the genome in this cohort (**Table 1**).

Using this data, we performed a series of statistical tests to identify genes that were individually associated with clinical variables including patient sex and fusion status. In summary, for patient sex, 21 genes were detected as significant among all tests and we noted that of the 21 genes the drove the sex-associated signature 14/21 genes were located on the Y-linked genes (**Table 5**). This also approach also nominated genes that correlated with fusion status (133 genes, **Figure 4-1A, Table 6**). To then determine if the genes that correlated with fusion status in the tumors also related to previously derived fusion-regulated gene sets from cell lines, we utilized gene signatures generated by other groups including two independent studies that knocked down the *CRTC1-MAML2* fusion and assessed differential gene expression (71, 73). Our signature had significant overlap with one of the two *CRTC1-MAML2* fusion-dependent signatures and trended to significant overlap with the second gene signature, P = 0.017, 0.147;

FDR Q=0.059, 0.281, respectively (**Figure 4-1B**). Further gene set enrichment analysis of the fusion gene signature demonstrated a strong enrichment of concepts to oxidative phosphorylation and the electron transport chain (**Figure 4-1C**). For example, HALLMARK_OXIDATIVE_PHOSPHORYLATION (FDR q-value 0.0016), GO_ATP_SYNTHESIS_COUPLED_ELECTRON_TRANSPORT (FDR q-value 0.0004), and GO_RESPIRATORY_ELECTRON_TRANSPORT_CHAIN (FDR q-value 0.0003) were among the most statistically significant enriched concepts for fusion related genes (**Table 4-7**), suggesting that *CRTC1/3-MAML2* fusion positive tumors depend on an increase in efficiency of this molecular process.

Previous studies have indicated that CRTC1/3-MAML2 fusions drive expression of both CREB-related genes and cMYC-regulated genes, but not Notch-regulated genes (73). Thus, we queried CRTC 1-MAML2 fusion-associated gene expression signatures for overlap with five publicly available gene lists of either Notch or MYC associated genes (191, 192). Our signature was not found to be significantly enriched for MYC or Notch-associated gene sets; however, we did see strong association of fusion status with the Notch effector *HEY1* (P = 0.0064, Two sample T-test with Welch Variance) (**Figure 4-1D**). Furthermore, recent studies have demonstrated that LINC00473 is regulated by the CRTC1-MAML2 gene fusion and we observed a strong correlation of LINC00473 expression with fusion status in the tumors (P = 3.35 x 10^-4, Two sample T-test with Welch Variance). Interestingly, we also observed strong individual associations of *FLT1* (VEGFR1) and *VEGFA* with fusion status (P = 0.0463 and P = 0.0218, respectively, by Two sample T-test with Welch Variance), suggesting mechanism driving increased angiogenesis in fusion positive tumors, as well as with expression of *PTCHD4* (P = 0.0418, Two sample T-test with Welch Variance), an important regulator of hedgehog signaling.

Next, we sought to characterize the immune cell composition of the MECs using CIBERSORTx (195), an algorithm that leverages reference gene sets to estimate relative immune cell abundance from mixed populations. This demonstrated a range of immune "hot" and immune "cold" tumors from the cohort, and we then assessed potential correlations of the relative immune cell compositions with clinical features and CRTC1/3-MAML2 fusion status. Although we did not identify significant enrichments with sex, or stage, surprisingly we observed that fusion negative tumors had an enrichment in proportion of CD8+ T-cells (P = 0.008, equal variance T-test) as compared to fusion positive tumors (**Figure 4-1E**). Accordingly, IHC analysis of fusion positive and negative MEC tissue sections confirmed that fusion positive tumors had lower CD8+ T-cell infiltration that fusion negative tumors (**Figure 4-1F**).

In addition to the fusion associated gene and immune signatures, we then sought to identify signatures associated with grade, which uncovered a 68 gene signature that was strongly correlated with grade, **Figure 4-2A, Table 4-8**). Using this gene signature (**Figure 4-2A**), we then performed gene set enrichment using the "MEC grade" rank list signature using GSEA, which identified a strong negative correlation of gene sets involved in B-cell regulation (**Figure 4-2B**). For example, we observed negative correlations with GO_B_CELL_RECEPTOR_SIGNALING_PATHWAY (FDR q-value < 0.0001), GO_B_CELL_MEDIATED_IMMUNITY (FDR q-value < 0.0001), and GO_REGULATION_OF_B_CELL_ACTIVATION (FDR q-value = 0.0005). In contrast, we observed few molecular enrichments positively correlated in grade, which were limited to HALLMARK_E2F_TARGETS (FDR q-value < 0.0001), HALLMARK_MYC_TARGETS_V1 (FDR q-value = 0.0007), and ATAGGAA_MIR202 (FDR q-value = 0.0123) (**Table 4-9**), suggesting a potential molecular role for these cancer drivers in high grade MEC.

In our analysis of individual genes associated with grade, we identified several expected including the mucin genes, *MUC5B* (P = 0.04, Linear regression with log FPKM) and *MUC7* (P = 0.04, Linear regression with log FPKM) as well as several novel genes that were not previously linked to MEC (**Figure 4-2C**). For example, we identified strongly associated genes also included several transcriptional regulators including *FOS, FOSB, JUN* and *EGR1* as well as non-coding RNAs including *LINC-PINT* and *MIR29A* (P < 0.05 for all genes individually by Linear regression with log FPKM). *LINC-PINT* has previously been mechanistically linked to *EGR1*, in that overexpression of *LINC-PINT* caused a decrease in *EGR1* and *FOS* expression as well as a PRC2-dependent downregulation of invasion associated genes in colorectal carcinoma models (201). We did not see correlation between expression of *LINC-PINT* and *EGR1*, but did see modest associations with both *FOS* and *JUN*. Collectively, our individual gene and gene set enrichment analysis showed several strong associations of potentially mechanistically critical genes with both fusion status and grade.

In addition to identifying correlated genes, we also noticed a significant association between tumor grade and immune cell content. In contrast to fusion status, which was highly associated with CD8 content, when comparing grade, we observed a significant depletion of the proportion of CD20+ B-cells in high and intermediate grade as compared to low grade tumors (P = 0.020, equal variance T-test) (**Figure 4-2D**), which we confirmed by IHC on independent sections (**Figure 4-2E**). Consistent with the calls from CIBERSORT, we also observed a correlation of expression of *MZB1*, a B-cell specific gene, with grade (P = 0.049) (**Figure 4-2C**).

Given the strong enrichment of infiltrating immune cells in several of the tumors, we next characterized the RNA expression of 13 immune checkpoint proteins in the cohort. Importantly, while the expression of several of these genes including *CD274* (PDL1) was relatively low (0-2

FPKM), we noticed that expression of *C10ORF54* (*VSIR*, *VISTA*) and *VTCN1* (*B7H4*) was relatively high suggesting a role for these genes in this disease (**Figure 4-3A**). Interestingly, *VSIR* expression trended towards inverse correlation with grade (adjusted $P = 0.07$, linear regression); however, no other correlations with clinical variables were noted (**Figure 4-3B**). By then leveraging the CIBERSORTx-derived immune content variables, we identified significant correlations of *VISTA* RNA expression with CD8 T-cell content and *B7H4* RNA expression with CD20+ B cell content.

To further characterize the RNAseq cohort, we then used FusionCatcher to identify novel gene fusions as well as to confirm the presence of known gene fusions in each specimen. Of our *CRTC1/3-MAML2* fusion positive samples, the FusionCatcher algorithm identified the appropriate fusion in 12/32 (37.5%) qPCR-verified fusion positive samples with high read confidence for either *CRTC1-MAML2* or *CRTC3-MAML2* in MEC samples previously identified to harbor the respective fusion genes (**Table 4-10**). However, the algorithm identified a list of 98 additional candidate fusion genes. Of these fusions, one was predicted to result in an out-of-frame gene (*PKD1-NPIPA1* in 2 samples, MEC10 and MEC11), none were predicted to drive in-frame coding sequences, and the remainder were largely predicted to fuse non-coding RNA transcripts with pseudogenes and were not explored further.

4.4.1 Spatial RNA sequencing of primary MEC tumors

We hypothesized that within individual fusion-positive MEC tumors, gene expression would vary based on cell type, with differences in mucosal, epidermoid, and intermediate cells. To test this hypothesis, we performed Visium spatial RNA sequencing on 7 FFPE MEC tumors and 1 healthy parotid gland control. Samples were selected to include 6 fusion-positive and 1

fusion-negative tumor, from a variety of tumor sites (**Table 4-11**). Tumors were evaluated by a pathologist (J.B.M.), and regions with a high proportion of tumor cells were sequenced. Technical duplicates were sequenced for each sample. The resulting sequencing data was then analyzed as outlined in **Figure 4-4A**. We identified high reproducibility between technical duplicates of each sample, using principal component analysis (**Figure 4-5A**). After samples were normalized using SCTransform, a regularized negative binomial regression, we performed dimensional reduction via Uniform Manifold Approximation and Projection (UMAP). By UMAP, we observed no differences in clustering of cell types between samples (**Figure 4-5B**). We then performed nearest neighbor clustering on the combined dataset. These clusters spatially localize in each sample, and the localization of each cluster is highly reproducible between technical replicates (**Figure 4-5C**). Next, we performed differential expression analysis on each cluster, only including the 3000 most variable genes. 15-63% of all genes queried are differentially expressed between fusion-positive tumors and the CRTC1-MAML2 negative control tumor (**Figure 4-5D**). This suggests that CRTC1-MAML2 activity leads to changes in gene expression throughout the tumor microenvironment, potentially affecting many different pathways and cancer phenotypes. However, due to our small sample size and the dramatic effects suggested by this analysis, these changes were not explored further.

We next characterized spatial gene expression patterns of several genes of interest in fusion-positive and negative tumors (**Figure 4-4B**). We see *EPCAM* expression distributed throughout all MEC tumors, but no *EPCAM* expression in the normal salivary control tissue. EpCAM is a tumor cell marker in many epithelial cancers, including head and neck squamous cell carcinoma (202). *GRAMD1B* and *RAMP1* are selected from the CRTC1-MAML2-associated gene signature described above, by bulk RNAseq; the expression of both genes is moderately

correlated with *EPCAM* expression in fusion-positive, but not fusion-negative, tumors (fusion positive mean R=0.356 +/- 0.137, 0.378+/- 0.154; fusion negative R=0.145+/-0.0646, 0.0539+/- 0.0378, respectively, determined by linear regression).

Next, we sought to define the genes with expression across EpCAM positive spots in the all samples correlated with CRTC1-MAML2 expression. However, CRTC1-MAML2 transcripts could not be reliably differentiated from wild type CRTC1 or MAML2 reads. Therefore, for each MEC sample, we leveraged genes positively correlated with the "fusion signature" gene set, which we used as a proxy for CRTC1-MAML2 transcription factor activity (methods outlined in **Figure 4-4A**). The fusion signature gene set consists of 133 genes differentially expressed between fusion-positive and negative MEC tumors, determined by bulk RNAseq (**Table 4-6**). We next filtered each sample to only include spots that express EpCAM, a known epithelial tumor marker that is not expressed in the normal control sample. We then performed a linear regression, then took the geometric mean of all R values. Genes with a geometric mean R value of 0.3 were considered significantly correlated with fusion signature expression (**Figure 4-4C**). Using this method, we identified genes that correlate with the fusion signature in fusion-positive samples. 25 genes correlate with this signature in all fusion-positive tumors, including *KRT7*; 67 genes correlate with this signature in 5/6 fusion-positive tumors, including *CTNNB1*; 117 genes correlate with this signature in 4/6 fusion-positive tumors, including *VEGFA* (**Figure 4-4B, C**). An additional 172 genes correlate with this signature in 3/6 fusion-positive tumors (**Table 4-12**). We then performed gene set enrichment analysis, ranking genes by geometric mean R value. In all samples tested, genes correlated with the CRTC1-MAML2 signature are also positively correlated with gene sets related to cellular respiration and oxidative phosphorylation, as well as HALLMARK_MYC_TARGETS_V1 (**Figure 4-4D).**

We next sought to validate the findings of the bulk RNAseq data, of decreased CD8 T cells infiltration in CRTC1-MAML2-positive tumors. We counted the number of Visium spots expressing *CD8A* (SCT transformed read count>0) for each sample tested (**Figure 4-6A)**. Though multiple cells can be sequenced by a single Visium spot, this is mostly likely to lead to underestimated T cell counts in tumors with high degrees of infiltration; therefore, we use this method to semi-quantitatively measure the amount of CD8 T cell infiltration per sample. We find decreased frequency of CD8+ spots in fusion-positive tumors, compared to both the fusion-negative and normal control ($p=0.026,5.7x10^{-4}$, respectively, Student's T test of per-sample mean expression), consistent with results from bulk RNAseq data showing that CRTC1-MAML2+ tumors were associated with lower CD8 TIL content.

4.4.2 Cell type categorization of spatial RNAseq

We hypothesize that gene expression patterns differ within the MEC tumor, based on cell type. Using the Visium method, we identify gene expression pattern per capture spot (55μM diameter), rather than per individual tumor cell. Therefore, we performed spatial deconvolution to estimate relative cell type abundance in spatial RNAseq datasets, which can then be used to impute per-cell type gene expression patterns. The R package SpatialDecon estimates relative cell type abundance per spot, using single cell RNA sequencing (scRNAseq) data as a reference. During the period of this study, we were not able to acquire MEC tumors for scRNAseq, so we instead use scRNAseq data from 12 healthy parotid glands, downloaded from the publicly available Tabula Sapiens database (199). Because this method cannot identify malignant cells, it has limited efficacy deconvoluting tumor samples, though it may be more successful for the normal control, a cancer-free parotid gland (**Figure 4-7**). Therefore, we next sought to identify

regions enriched for mucosal, epidermoid, or intermediate cells within the sequenced tumor section, by examination by an experienced head and neck pathologist (J.B.M.). It was determined that each sequenced sample is too admixed for effective differentiation between regions. Therefore, further studies, which include single cell RNA sequencing of MEC tumors, will be necessary to determine the cellular composition of spatially resolved tumors.

4.4.3 Detecting transcript coexpression with RNAscope

We hypothesized that genes with expression highly correlated with the fusion signature would be more frequently expressed in CRTC1-MAML2 expressing tumor cells. Because spatial RNA sequencing lacks single-cell resolution, we questioned if genes highly correlated with the CRTC1-MAML2 fusion signature due to coexpression in the same cell, or due to colocalization of CRTC1-MAML2 expressing cells and cells expressing these genes of interest, such as *CTNNB1* and *VEGFA*. Therefore, we performed RNAscope to determine the expression patterns of *EpCAM*, *LINC00473*, *VEGFA* and *CTNNB1* transcripts at single-cell resolution. *LINC00473* expression is directly regulated by CRTC1-MAML2, and is therefore a marker of transcription factor activity (175). Both *VEGFA* and *CTNNB1* expression are associated with the CTRC1-MAML2 signature, and *EpCAM* is a marker of epithelial tumor cells. CRTC1-MAML2-positive and -negative tumors were probed for EpCAM, LINC00473, and either CTNNB1 or VEGFA transcripts. Cells from each tumor image were identified, and cells expressing each transcript were counted using QuPath. Frequency of cells containing each transcript were highly variable between each sample (**Figure 4-8A**). Next, we queried if distance between CTNNB1+ or VEGFA+ cells and EpCAM+ or LINC00473+ cells varied based on fusion status. Therefore, we quantified the distance between each cell and nearby VEGFA+ or CTNNB1+ cells (**Figure 4-8B**). We counted the number of VEGFA+ or CTNNB1+ cells near each cell, using a 50μm

nearness cutoff (as in (203)). In addition, we report the distance to the nearest VEGFA+ or CTNNB1+ cell. While there are no significant differences in either measurement between CRTC1-MAML2-positive and CRTC1-MAML2-negative cells, this is likely due to high inter-tumoral variability (**Figure 4-8C**). Therefore, future research includes a MEC tissue microarray that is currently being developed to determine a fusion-specific effect, using an increased sample size, which is beyond the scope of this thesis.

4.5 Discussion

Previous studies have successfully leveraged the available MEC cell line models to manipulate *CRTC1-MAML2* fusion gene expression and define cell line associated fusion gene signatures (71, 73, 175). Our data significantly extends these signatures by being the largest MEC tissue cohort to date. Our data confirm that *LINC00473* is associated with fusion status, and discover that *HEY1*, *VEGFA* and *PTCHD4* are also associated with the gene fusion. Similarly, our data identified associations of Mucin proteins with disease grade confirming the previously identified associations with mucin and MEC in recent immunohistochemistry cohorts (169, 204, 205), and our data also discover that high grade tumors are also associated with *FOS*, *JUN* and *LINC-PINT* expression, suggesting that these genes are pivotal for advanced grade disease.

. The discovery of the *HEY1* association with CRTC1-MAML2 fusion status was unexpected, as previous functional studies have suggested that *CRTC1-MAML2* does not drive expression of Notch signaling effectors (71, 73, 175), and suggests that potential secondary or alternative mechanisms associated with the gene fusion may be responsible for driving *HEY1* expression. One such mechanism is transcription activation by MYC. The *HEY1* promoter contains a MYC binding site, measured by ChIP-seq of A549, a lung cancer derived cell line, in

the publicly available ENCODE database (206). MYC has been previously shown to interact with CRTC1-MAML2 (73). If *HEY1* is regulated by MYC in fusion-positive MEC tumors, this mirrors the first studies of CRTC1-MAML2 mediated Notch signaling, which first demonstrate CRTC1-MAML2-mediated Notch activation via a *HES1* reporter, only to later determine that this activation is better explained by the discovery of a *HES1* CREB binding domain (70, 132). Taken together, these findings both describe the limitations of reducing individual genes to categories like Notch-regulated, though further study is needed to measure MYC binding at the *HEY1* promoter in MEC tumors.

In addition to gene set enrichments, our transcriptome data is the first to identify associations between tumor infiltrating immune cells and genetic or clinical characteristics of MEC tumors. Surprisingly, we show a significant depletion of infiltrating CD8+ T-cells in *CRTC1/3-MAML2* fusion positive tumors suggesting that perhaps CD8+ T-cell exclusion may be important cellular mechanism for tumors expressing fusion gene-derived neoantigens. Further, we also discovered a novel relationship between MEC grade and CD20+ B-cell content, showing that total B-cell infiltration was also decreased in high grade tumors, consistent with recently described analyses of B-cells in other malignancies(132, 207). Overall, however, our data show that the infiltrating immune content is closely related with both genetic status and grade of the tumor, and future studies will need to carefully analyze the role of these immune components to understand how they contribute to disease progression as well as how tumor infiltrating immune content may need to regulated in order to improve the efficacy of immunotherapy-based approaches in the future.

By spatial RNA sequencing, we identify a putative interaction between CRTC1-MAML2 transcription factor activity and both *CTNNB1* and *VEGFA* expression within heterogeneous

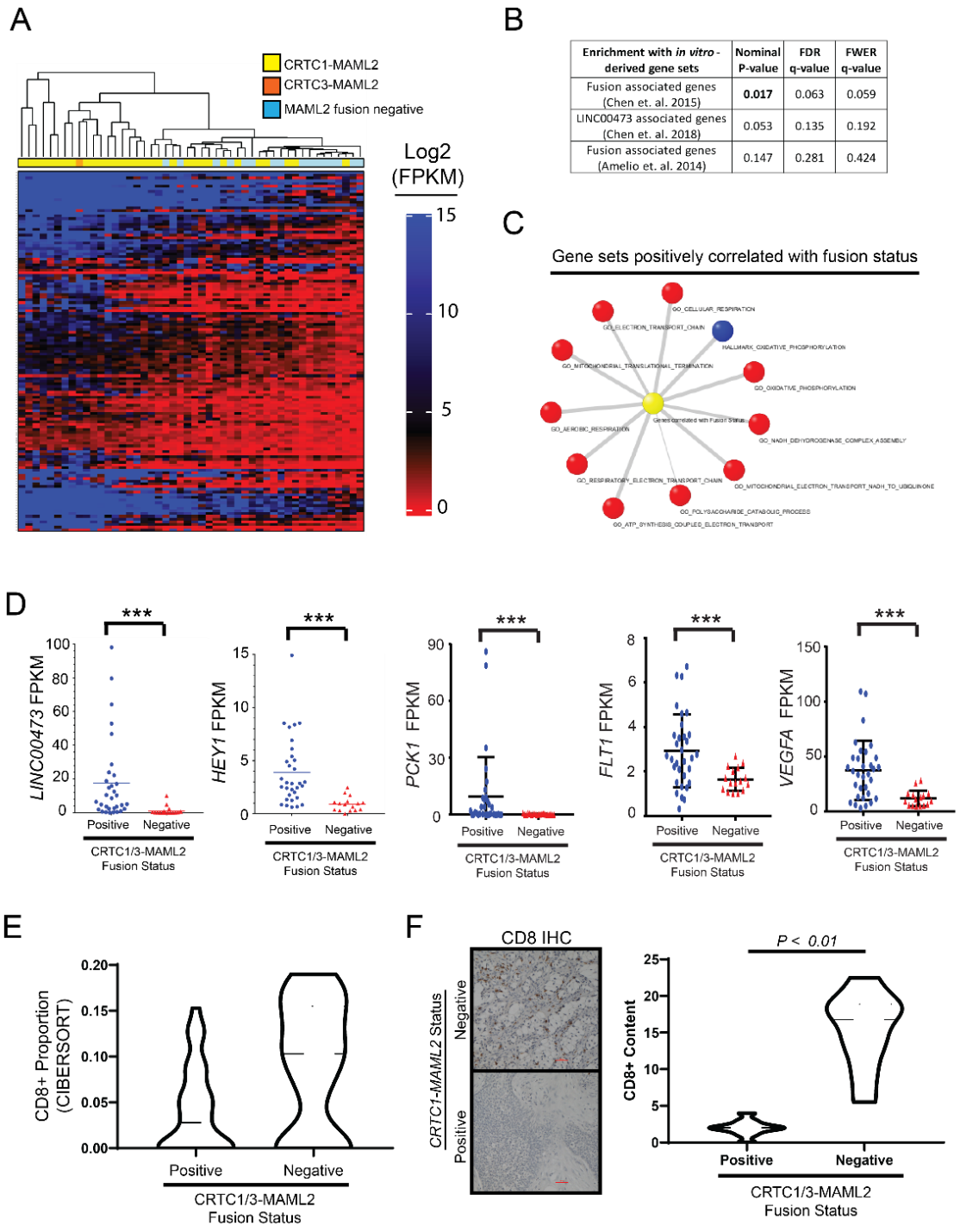
tumors. In MEC tumors, beta catenin is preferentially expressed in epidermoid cells; however, the role, prevalence of accumulation, and effect on prognosis of beta-catenin is controversial in MEC tumors with an unknown MAML2 fusion status (208-211). Therefore, the interaction between CRTC1-MAML2 and beta-catenin in epidermoid cells is a promising target for future study. By contrast, *VEGFA* overexpression is reproducibly associated with increased tumor size and worse disease-free survival for MEC patients (212, 213). VEGFA is associated with MAML2 fusion positivity by both bulk RNAseq and spatial RNAseq, and its increased expression could impact tumor growth by increased angiogenesis, suppressing immune cell activity, or suppressing immune infiltration (reviewed in (214, 215)). Therefore, the CRTC1-MAML2-associated increased *VEGFA* expression may impact many cancer phenotypes, and the effects of VEGF pathway inhibition in MEC is a promising topic for further study.

The data presented here represent a significant step forward in the ability to detect aggressive MECs that may benefit from alternative therapeutic strategies and/or enhanced monitoring following standard of care therapy. We hope that in the future, studies assessing therapeutic strategies to disrupt the CRTC1/3-MAML2 gene fusion or its pivotal targets, MEC-immune cell interactions will serve as the basis of improved outcomes for patients with clinically aggressive MEC.

4.6 Acknowledgements

Thank you to Padma Kadiyala and Marina Pasca di Magliano for help developing this RNAscope assay method. Thank you to Jonathon B. McHugh (J.B.M.) for expert tumor pathology evaluation. Thank you to Apurva Bhangale, Molly Heft Neal, Collin Brummel, Matthew Spector, and Chad Brenner for assistance with data curation and analysis.

4.7 Figures



Enrichment with *in vitro* - derived gene sets	Nominal P-value	FDR q-value	FWER q-value
Fusion associated genes (Chen et. al. 2015)	**0.017**	0.063	0.059
LINC00473 associated genes (Chen et. al. 2018)	0.053	0.135	0.192
Fusion associated genes (Amelio et. al. 2014)	0.147	0.281	0.424

Figure 4-1. We identify a CRTC1-MAML2 associated gene signature.

A) RNA levels (FPKM) of genes differentially expressed in MAML2 fusion-positive and fusion-negative tumors. **B)** This gene signature significantly overlaps with previously published studies,

which define genes associated with CRTC1-MAML2-expressing cell lines. **C)** Gene sets positively correlated with the MAML2 fusion-associated signature, determined by gene set enrichment analysis. **D)** Expression of genes within this signature in CRTC1/3-MAML2-positive and CRTC1/3-MAML2 negative tumors. **E)** CD8+ T cell proportion is increased in fusion-negative tumors.CD8+ T cell prevalence is determined using CIBERSORT cell type deconvolution. **F)** CD8-expressing cells are highly elevated in fusion negative tumors, as measured by IHC.

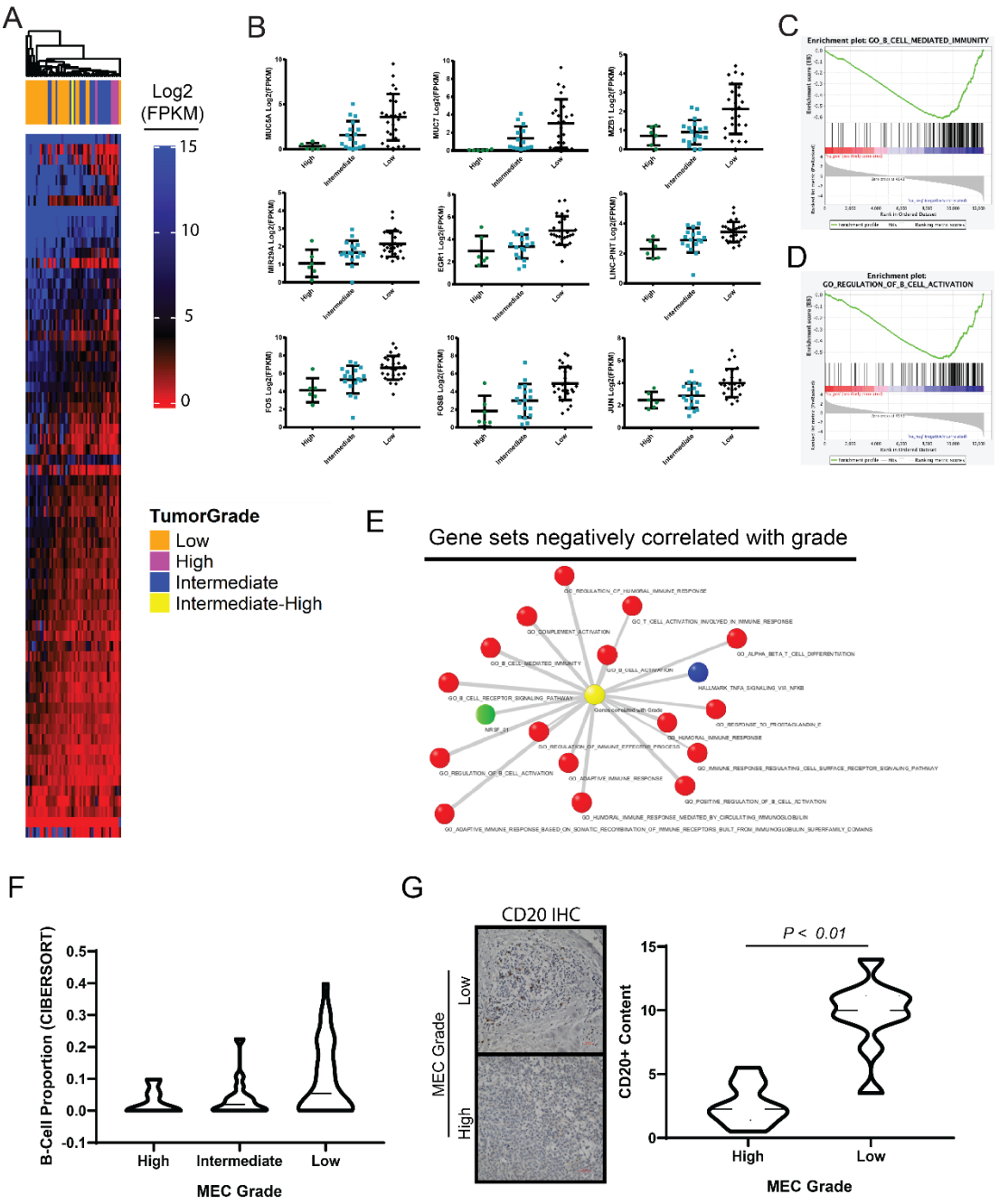


Figure 4-2. We define a grade-associated gene expression signature, which is enriched for genes with a role in B cell activity.

A) RNA levels (FPKM) of genes differentially expressed, based on tumor grade. **B)** Gene expression of genes within this signature in low, intermediate, and high-grade tumors. Genes in the grade-associated signature are negatively correlated with for **C)** B cell-mediated immunity and **D)** regulation of B cell activation, tested via gene set enrichment analysis. **E)** All gene sets negatively correlated with the tumor grade-associated gene signature. **F)** B cell proportion is negatively associated with tumor grade. B cell prevalence is determined via cell type deconvolution with CIBERSORT. **G)** CD20+ cells are more prevalent in low-grade tumors, as measured by IHC.

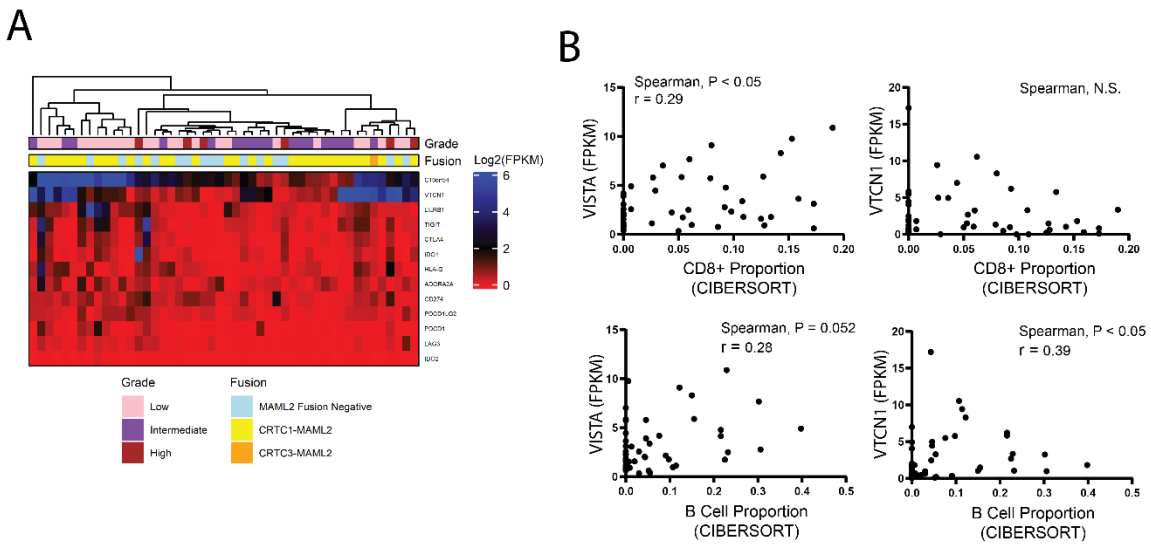


Figure 4-3. MEC tumors express immune regulators VISTA and VTCN1 transcripts.

A) Heat map depicts gene expression (FPKM) of known immune checkpoint genes *c10orf54* (VISTA) and *VTCN1* are the most commonly expressed immune checkpoint genes across the tumors in this cohort. **B)** Association of *VTCN1* and *c10orf54* with T and B cell proportion in the tumor microenvironment.

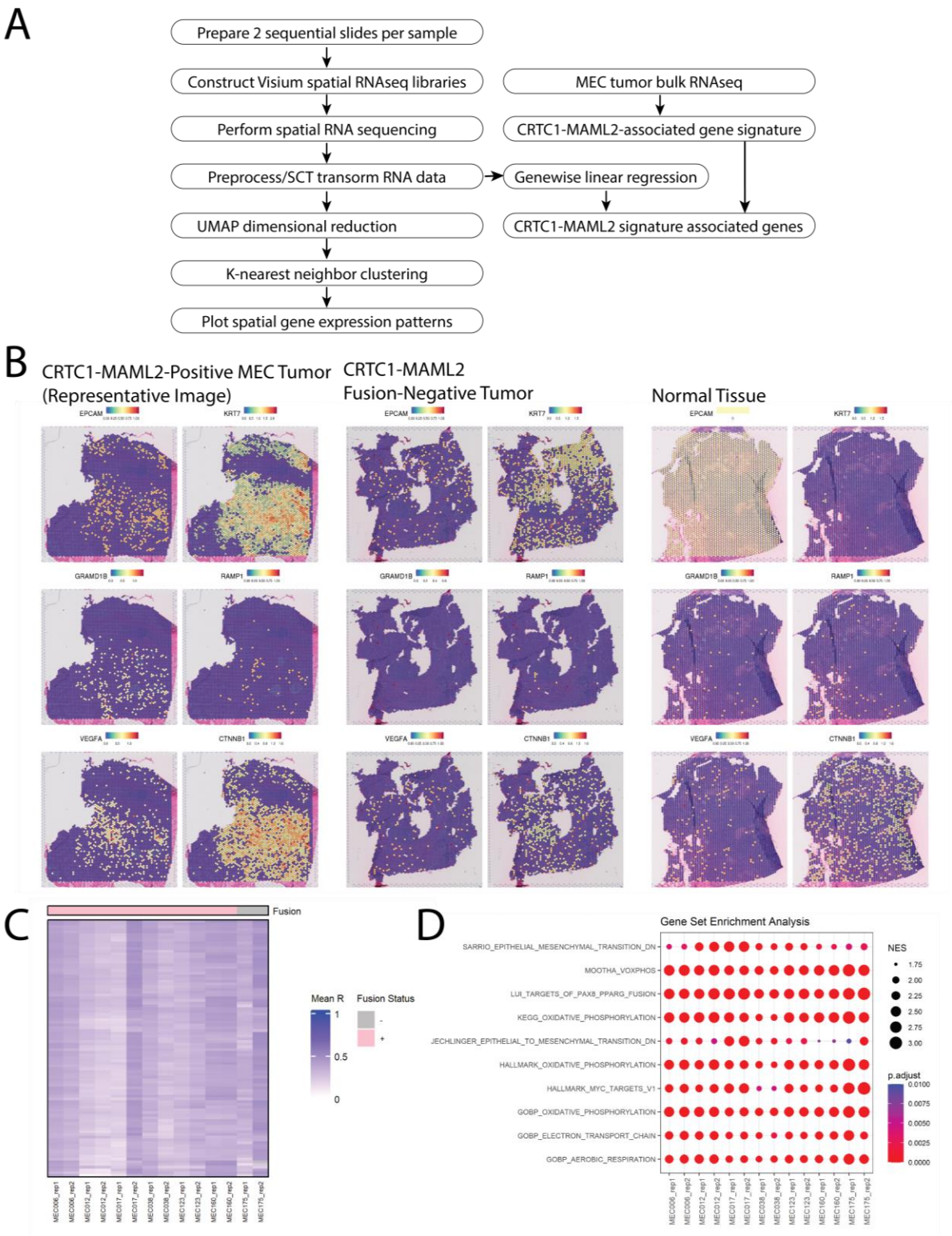


Figure 4-4. Spatial RNAseq reveals spatially resolved gene expression patterns.

A) Analysis method for spatial RNAseq is defined. Details, including bioinformatics tools used, are provided in the methods section. **B)** Spatial expression plots of a representative CRTC1-

MAML2-positive tumor, a CRTC1-MAML2-negative tumor, and the healthy parotid gland. *EpCAM* is an epithelial tumor marker, *GRAM1DB* and *RAMP1* expression is increased in CRTC1-MAML2 positive tumors, determined by bulk RNAseq. *KRT7, VEGFA*, and *CTNNB1* expression is associated with the MAML2 fusion gene signature. Spatial expression plots for all samples are in **figure 4-9**. **C)** Genes associated with MAML2 fusion signature. For each gene, we calculate the Spearman correlation coefficient, comparing gene expression to the expression of each gene in the MAML2 fusion gene signature. The geometric mean of all R coefficients in the signature is reported. Genes with a geometric mean R>0.3 in greater than 4 CRTC1-MAML2 positive tumors are included in this heat map. **D)** Gene sets significantly enriched in genes associated with the fusion signature.

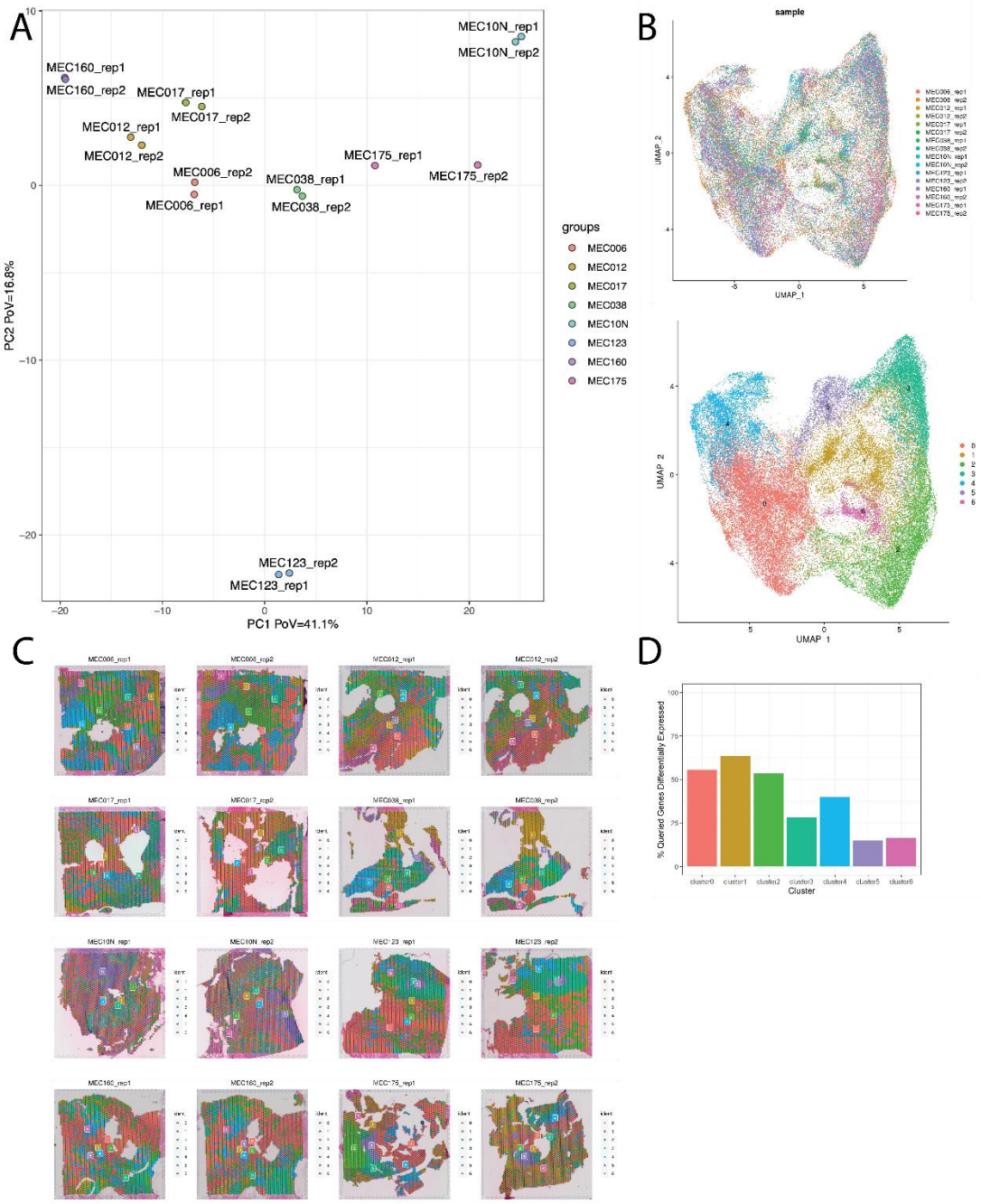


Figure 4-5. Spatially respolved spots cluster by gene expression.

A) Principal component analysis of each tumor undergoing spatial RNAseq. Technical replicates are plotted separately. **B)** Per-sample distribution and k-means clustering identities following UMAP dimensional reduction. **C)** Spatial distribution of each cluster identity, in each sample sequenced. Cluster definitions are consistent between each sample. **D)** Percentage of genes differentially expressed per cluster between CRTC1-MAML2 fusion positive and fusion negative tumors.

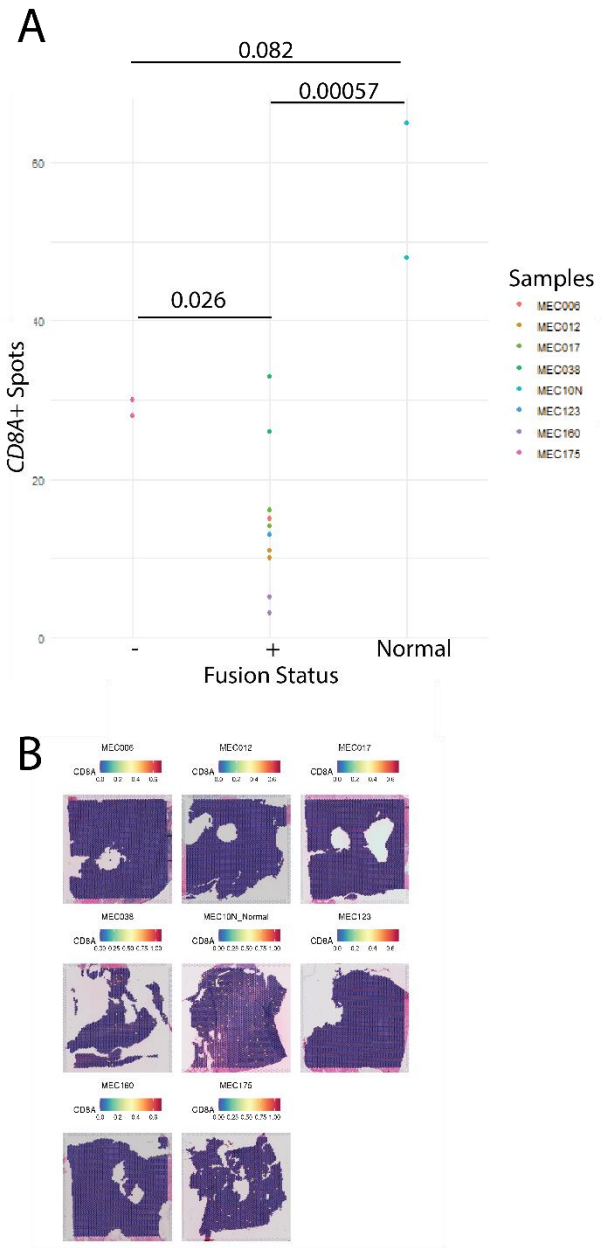


Figure 4-6. VISIUM spatial RNAseq also reports decreased CD8+ T cell infiltration in CRTC1-MAML2+ tumors.

A) The number of spots containing *CD8A* (a CD8+ T cell marker) is reported. The number of spots containing *CD8A* is significantly lower in CRTC1-MAML2-positive tumors, compared to both CRTC1-MAML2-negative tumors and the healthy parotid gland. Technical duplicates of the same tumor are plotted separately. **B)** Spatial plots of *CD8A* gene expression in each tumor sequenced.

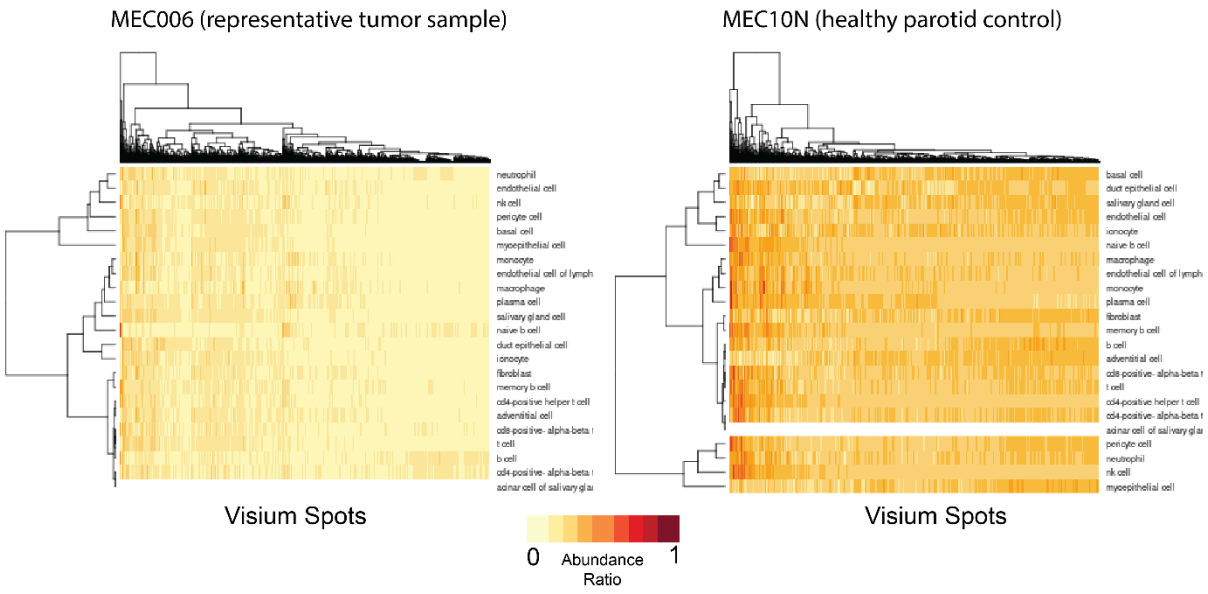
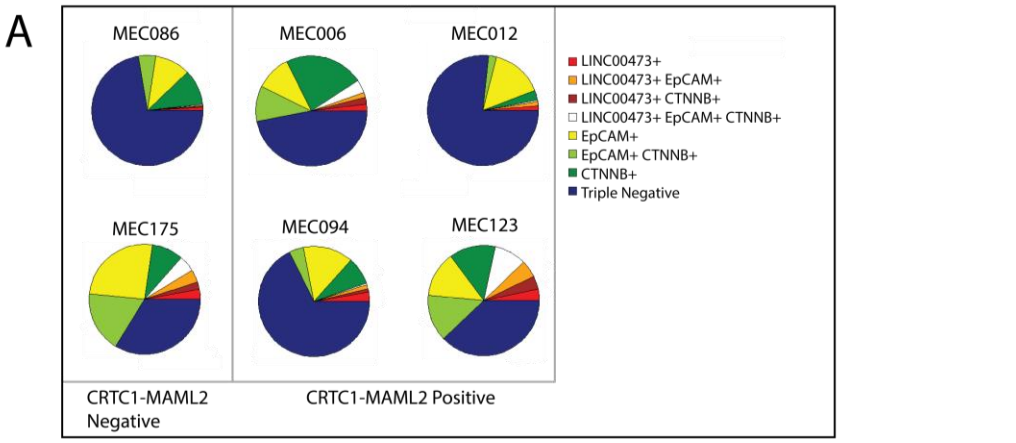
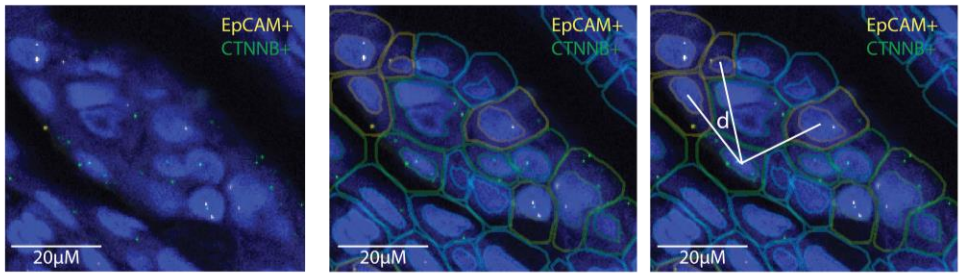
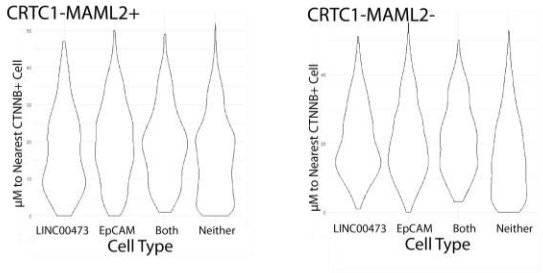


Figure 4-7. Spatial cell type deconvolution is unsuccessful in tumor cells.

Each heat map reports the proportion of gene expression of each Visium spot associated with each cell type listed. Healthy parotid gland scRNAseq data is used as a cell type reference.







Count of CTNNB+ Cells <50μm from cell

	Fusion +	Fusion -
LINC00473 +	12.8 ± 18.9	10.3
LINC00473 -	17.5 ± 21.7	13.9 ± 4.1

	Fusion +	Fusion -
EpCAM +	14.6 ± 18.5	10.2
EpCAM -	17.7 ±22.0	14.3 ± 3.7

Distance to Nearest CTNNB+ cell (μm)

	Fusion +	Fusion -
LINC00473 +	5 ± 7.0	20
LINC00473 -	5 ± 6.8	11.5 ± 7.8

	Fusion +	Fusion -
EpCAM +	6 ± 7.6	20
EpCAM -	5 ± 6.5	11 ± 7.1

Median ± standard deviation of all per-sample medians reported

Figure 4-8. RNAscope reveals tumor gene expression patterns at single-cell resolution.

A) Cell type distribution per sample. **B)** Method for distance calculations between cells. First, cells RNAscope is performed and imaged by confocal microscopy. Yellow spots are EpCAM RNA transcripts, and green spots are CTNNB1 transcripts. Samples are also probed for LINC00473 (red spots, not pictured) Second, cells are automatically detected by QuPath image analysis, and cells are identified as EpCAM+ (yellow circle), CTNNB1+ (green circle), EpCAM+ CTNNB1+, or double negative (blue circle). Third, location coordinates are recorded each cell, and used to calculate distance between each cell and nearest CTNNB1+ (green) cell. **C)** Distance to nearest CTNNB1+ cell for a CRTC1-MAML2 positive and negative tumor (representative images). These results are summarized in the included table (N=4 fusion-positive, 2 fusion-negative)

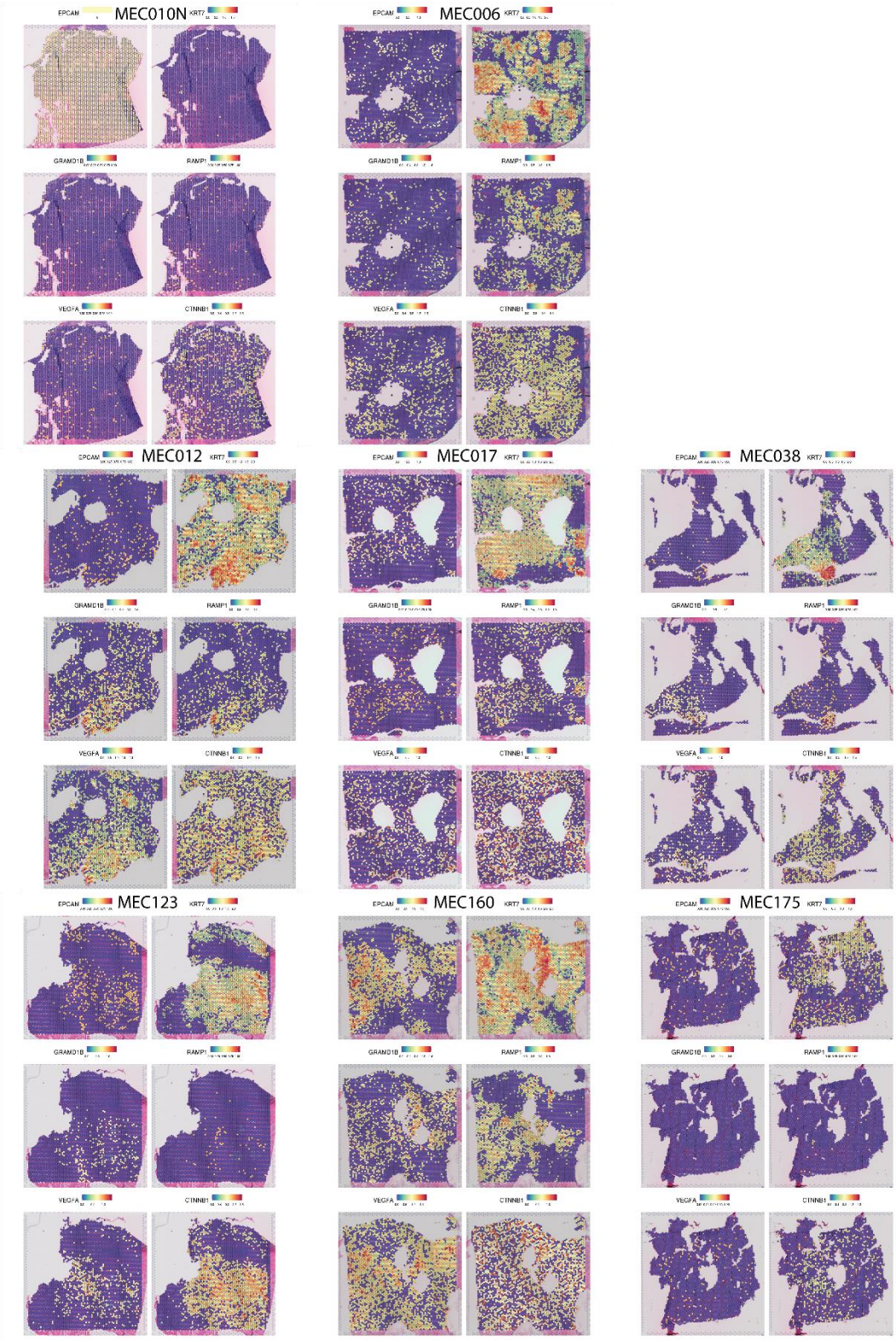


Figure 4-9. Spatial gene expression plots of every tumor sequenced.

Genes of interest selected as in figure 4-4B.

4.8 Tables

Feature		Fusion Status	
		Positive (N = 32)	Negative (N = 16)
Gender	Female	16	6
	Male	16	10
Ethnicity	Caucasian	15	14
	African-American	8	-
	Asian	-	1
	Unknown	1	1
Age	Years	47.1	53.3
Grade	Low	16	8
	Intermediate	14	4
	High	2	4
Clinical Stage	I	14	7
	II	10	5
	III	3	2
	IV	4	2
	Unknown	1	-
Site	Parotid	15	7
	Other	17	9
Tobacco Use	None	18	9
	Former	9	5
	Current	5	2
Pack Years (Smokers)	Years (Range)	14.5 (0.5 to 57)	12.5 (3 to 25)
Alcohol Use	None	18	8
	Occasional	13	8
	Heavy	1	0
Comorbidity Index (ACE)	0	8	7
	1	13	5
	2	6	2
	3	5	2
Died of Disease		6	2

Table 4-1. Clinical information of MEC samples sequenced

Number	Name	Total Reads	% Uniquely Mapped	Non-Zero FPKMs	FPKM > 1
81352	MEC5	79331941	82.62	27339	11796
81353	MEC6	65777804	89.4	26326	11042
81354	MEC7	60104171	82.88	26170	12399
81355	MEC8	49583315	87.64	25537	11754
81356	MEC9	53676260	87.35	25313	11892
81357	MEC10	61390865	87.58	25503	11408
81358	MEC11	48021793	84.16	25932	12144
81359	MEC12	56633399	89.31	26420	11984
81360	MEC13	66524943	73.98	19118	8620
81362	MEC15	58981555	90.45	25610	11133
81363	MEC16	54509380	79.31	22982	10481
81364	MEC17	51720568	91.4	25429	11395
81365	MEC18	65811995	88.59	23883	10990
81366	MEC19	68232756	85.98	23403	10856
81367	MEC20	79568084	91.89	23312	10831
81368	MEC21	72317942	91.57	25970	11611
81369	MEC22	72029831	85.89	24129	10644
81370	MEC23	73853131	63.24	18663	9342
81371	MEC24	57511761	60.24	17763	8496
81372	MEC25	63879153	75.02	18895	6460
81374	MEC27	71546360	89.48	20943	10079
81375	MEC3	63378164	83.92	23930	12318
81378	MEC1	55489888	76.84	23548	11132
81379	MEC28	49070141	89.26	25185	10635
81384	MEC33	56434070	86.64	24123	10360
81386	MEC35	68475798	89.29	23431	11272
81387	MEC36	66729661	89.22	27186	12553
81389	MEC38	64625364	87.58	24572	10540
81390	MEC39	69721872	89.91	26500	11570
81391	MEC40	73078039	88.02	25878	11600
81392	MEC41	52845850	65.9	19764	9129
81396	MEC45	76267155	80	24469	11812
81397	MEC46	83295429	84.78	25196	10658
81560	MEC84	60302481	88.02	26120	11651
85191	MEC167	69719295	87.76	26160	12916
85192	MEC175	50610276	75.21	25794	13481
85193	MEC98	65750358	74.6	15552	6212
85194	MEC117	52453294	88.06	25253	12286
85195	MEC123	59522270	89.51	27531	13426
94491	MEC079	132221255	74	15163	7795
94492	MEC121	101717421	84.03	25005	12023
94493	MEC087	74856828	80.48	23964	12046
94494	MEC089	72432629	75.94	23737	12721
94495	MEC096	93523338	84.43	20637	12372
94496	MEC114	91917972	82.55	26164	12616
94498	MEC160	131152819	78.46	24518	12477
94499	MEC170	67860165	69.55	18745	11463
94500	MEC173	94763182	81.68	24413	12992

Table 4-2. Read depths and total mapped reads for each transcriptome sample.

Target	Company	Cat #	Source	Clone	Dil'n	Epitope retrieval	Detection	Ctrl Tissue
CD8	BioCare Medical	ACI3160	Mouse MonoAb	C8/144B	1:200	pH9 HIER	Env+	Tonsil
CD20	AbCam	Ab78237	Rabbit MonoAb	EP459Y	1:250	pH6 HIER	Env+	Tonsil
VISTA	Cell Signaling	64953	Rabbit MonoAb	D1L2G				Tonsil
B7-H4	Cell Signaling	14572	Rabbit MonoAb	D1M8I				Tonsil

Table 4-3. Antibodies used in this study.

Gender	Tracking ID	Gene	Logistic reg. log FPKM		2-sample t-test pooled var.		2-sample t-test pooled var. log		2-sample Welch t-test		2-sample Welch t-test log FPKM	
			P-value	Adjusted P	P-value	Adjusted P	P-value	Adjusted P	P-value	Adjusted P	P-value	Adjusted P
221	ENSG00000012817	KDM5D	1.50E-05	0.02614	5.44E-06	0.006233	2.04E-09	2.14E-06	7.1E-06	0.008169	6.38E-09	6.69E-06
735	ENSG00000067048	DDX3Y	1.37E-05	0.02614	2.32E-10	0.000003	4.05E-12	1.71E-08	3.8E-10	0.000004	2.53E-12	1.52E-08
756	ENSG00000067646	ZFY	1.66E-05	0.02614	5.58E-07	0.000703	5.88E-11	1.23E-07	5.4E-07	0.000685	3.30E-11	8.31E-08
1545	ENSG00000099725	PRKY	1.88E-05	0.02627	6.99E-08	0.000098	1.01E-09	1.38E-06	4.2E-08	0.000065	1.06E-09	1.22E-06
3028	ENSG00000114374	USP9Y	1.40E-05	0.02614	8.93E-10	0.000006	4.49E-11	1.13E-07	5.9E-10	0.000004	2.80E-11	8.31E-08
4266	ENSG00000129824	RPS4Y1	1.55E-05	0.02614	5.84E-05	0.043875	4.06E-12	1.71E-08	5.3E-05	0.041792	2.27E-12	1.52E-08
4388	ENSG00000131002	TXLNG2P	1.08E-05	0.02614	1.28E-08	0.000027	1.39E-10	2.50E-07	8.0E-09	0.000028	2.12E-10	3.80E-07
7596	ENSG00000165246	NLGN4Y	2.87E-05	0.03325	3.71E-08	0.000067	1.09E-09	1.38E-06	2.5E-08	0.000055	6.88E-10	9.63E-07
8895	ENSG00000176728	TTTY14	4.48E-05	0.04342	5.46E-08	0.000086	1.68E-09	1.92E-06	3.9E-08	0.000065	9.67E-10	1.22E-06
9427	ENSG00000183878	UTY	1.05E-05	0.02614	1.24E-08	0.000027	1.97E-10	3.10E-07	9.0E-09	0.000028	2.41E-10	3.80E-07
10289	ENSG00000198692	EIF1AY	3.93E-05	0.04121	9.97E-09	0.000027	3.68E-12	1.71E-08	2.6E-08	0.000055	3.62E-12	1.52E-08
11464	ENSG00000229807	XIST	1.12E-05	0.02614	2.65E-09	0.000011	3.02E-11	9.51E-08	6.0E-08	0.000084	2.09E-10	3.80E-07
11564	ENSG00000233864	TTTY15	2.90E-05	0.03325	2.71E-05	0.026274	9.06E-09	8.78E-06	1.8E-05	0.017078	7.36E-09	7.13E-06
1195	ENSG00000084073	ZMPSTE24	0.001905	0.23164	6.37E-05	0.044571	4.71E-05	0.031197	1.1E-04	0.063149	0.000210	0.09108
2478	ENSG00000108100	CCNY	0.001716	0.23164	1.12E-05	0.011764	1.08E-05	0.009727	1.5E-05	0.015826	5.20E-05	0.04370
2603	ENSG00000109534	GAR1	0.001024	0.23164	5.92E-05	0.043875	2.24E-05	0.016573	6.9E-05	0.051317	6.22E-05	0.04612
9780	ENSG00000188010	MORN2	0.001658	0.23164	5.79E-05	0.043875	2.11E-05	0.016573	8.4E-05	0.056340	8.41E-05	0.05882
10274	ENSG00000198612	COPS8	0.002717	0.23164	3.65E-05	0.032861	7.45E-05	0.042628	5.1E-05	0.041792	0.000264	0.09616
4217	ENSG00000129116	PALLD	0.001071	0.23164	7.90E-05	0.052372	1.32E-05	0.011048	5.2E-05	0.041792	2.41E-05	0.02172
8018	ENSG00000168246	UBTD2	0.001800	0.23164	9.51E-05	0.059890	5.74E-05	0.034434	8.5E-05	0.056340	0.000120	0.07189
10181	ENSG00000197965	MPZL1	0.002125	0.23164	0.000266	0.111496	4.05E-05	0.028366	1.9E-04	0.083846	9.35E-05	0.06199
11308	ENSG00000223773	CD99P1	0.001091	0.23164	0.000152	0.079704	5.61E-05	0.034434	1.1E-04	0.063149	6.02E-05	0.04612

Table 4-4. Sex-associated gene signature.

Gene	2-sample t-test pooled var. log FPKM		2-sample Welch t-test		2-sample Welch t-test log FPKM	
	P-value	Adjusted P-value	P-value	Adjusted P-value	P-value	Adjusted P-value
GRAMD1B	0.000001	0.005553	0.000021	0.021860	0.000000	0.000408
PLEKHH1	0.000050	0.034252	0.000039	0.029182	0.000507	0.064461
MYLK	0.000077	0.046180	0.005782	0.177297	0.001371	0.107478
TMEM260	0.000089	0.049988	0.000002	0.005137	0.000072	0.023302
ANKRD10	0.000037	0.034252	0.000428	0.053911	0.001797	0.120398
MSLN	0.000039	0.034252	0.000467	0.056976	0.000029	0.012980
WSB1	0.000048	0.034252	0.000086	0.033291	0.000043	0.016274
PTP4A1	0.000002	0.006863	0.000000	0.001114	0.000009	0.006709
PDE10A	0.000012	0.018783	0.000250	0.044321	0.000000	0.000765
EDA2R	0.000052	0.034252	0.000007	0.013184	0.000014	0.007716
RAMP1	0.000032	0.034100	0.001306	0.086799	0.000000	0.000408
KSR1	0.000001	0.005553	0.000002	0.005137	0.000016	0.008242
XPR1	0.000047	0.034252	0.000064	0.032126	0.000187	0.039019
MSI2	0.000021	0.026407	0.000014	0.019528	0.000133	0.032765
KCNMA1	0.000045	0.034252	0.003907	0.148670	0.000293	0.049219
SYT12	0.000069	0.043259	0.000094	0.033670	0.000005	0.004465
CALML5	0.000007	0.013276	0.036835	0.425627	0.003332	0.171290
CFI	0.000004	0.007668	0.000002	0.005137	0.000000	0.000408
LINC00473	0.000003	0.007668	0.000528	0.058854	0.000000	0.000336
LINC01059	0.000020	0.026407	0.000196	0.041834	0.000001	0.002015
SDIM1	0.000091	0.049988	0.000541	0.058888	0.000002	0.002157
AL139819.1	0.000029	0.032756	0.025045	0.360503	0.001513	0.111523
RP11-720N19.2	0.000000	0.001660	0.000000	0.000698	0.000000	0.000065
FUCA2	0.004044	0.204605	0.000056	0.029634	0.001102	0.097758
SLC25A5	0.006472	0.236088	0.000083	0.033291	0.002679	0.148016
ALDH3B1	0.001755	0.149130	0.000235	0.042956	0.000709	0.074996
CTSA	0.004082	0.204605	0.000117	0.036702	0.000640	0.073304
GNB5	0.000160	0.059991	0.000044	0.029182	0.000343	0.052868
B4GALT1	0.001135	0.126935	0.000144	0.040400	0.000434	0.058490
NLK	0.003517	0.195163	0.000331	0.049012	0.002241	0.135036
EDEM2	0.006720	0.236433	0.000223	0.042956	0.000656	0.073735
RBBP9	0.000249	0.076091	0.000014	0.019528	0.000178	0.038577
FH	0.002566	0.171514	0.000281	0.046302	0.001285	0.105092
PRTFDC1	0.003265	0.185274	0.000187	0.041834	0.001194	0.100931
LAMA1	0.000981	0.123119	0.000073	0.033291	0.000035	0.014491
FNDC3A	0.000799	0.112605	0.000139	0.040400	0.000949	0.087232
FLT1	0.009505	0.266552	0.000288	0.046302	0.000986	0.089379
LAPTM4B	0.002873	0.179111	0.000160	0.041834	0.001176	0.100651
ASAH1	0.004951	0.215000	0.000189	0.041834	0.003732	0.176570
GPI	0.001493	0.139352	0.000141	0.040400	0.000724	0.075610
PON3	0.000138	0.056092	0.000320	0.048495	0.000001	0.002015
PON2	0.000107	0.054628	0.000181	0.041834	0.000124	0.032471
CNTNAP3	0.000397	0.094375	0.000180	0.041834	0.000451	0.059621
MINPP1	0.008687	0.258656	0.000230	0.042956	0.001156	0.100453
SIAE	0.000274	0.076091	0.000022	0.021860	0.000005	0.004597
KIAA1549L	0.000557	0.103086	0.000229	0.042956	0.000004	0.004414
ATP5B	0.004864	0.215000	0.000142	0.040400	0.001569	0.112295
SCNN1A	0.003678	0.199578	0.000165	0.041834	0.001696	0.117375

VEGFA	0.000752	0.108838	0.000023	0.021860	0.000129	0.032479
MRPS27	0.002628	0.173275	0.000090	0.033291	0.000883	0.084849
GBE1	0.002435	0.166353	0.000087	0.033291	0.000366	0.053396
CLIP4	0.001494	0.139352	0.000258	0.045160	0.000653	0.073735
PECR	0.000962	0.123119	0.000157	0.041834	0.000058	0.019898
ID3	0.003923	0.204180	0.000327	0.049012	0.001371	0.107478
RPN2	0.012741	0.298715	0.000080	0.033291	0.001628	0.115184
AVPI1	0.000117	0.055554	0.000230	0.042956	0.000011	0.006872
CHPF	0.005002	0.215000	0.000305	0.046842	0.000753	0.077057
PIGT	0.003043	0.181615	0.000025	0.022917	0.000196	0.039908
EFNB2	0.000263	0.076091	0.000002	0.005137	0.000015	0.007716
HEATR5A	0.005240	0.219309	0.000339	0.049147	0.001554	0.112295
ITFG1	0.006538	0.236356	0.000215	0.042956	0.002613	0.145664
SAT1	0.000550	0.103086	0.000275	0.046302	0.001514	0.111523
TNS4	0.001719	0.149130	0.000084	0.033291	0.003727	0.176570
RFK	0.003120	0.184462	0.000040	0.029182	0.001374	0.107478
ITM2B	0.007714	0.245959	0.000350	0.049504	0.003805	0.178154
CTSV	0.003228	0.185274	0.000290	0.046302	0.000119	0.031837
PPP2R1B	0.000498	0.103086	0.000123	0.037853	0.000434	0.058490
TSPAN3	0.003010	0.181615	0.000112	0.036702	0.001773	0.120082
TOB1	0.006337	0.235304	0.000042	0.029182	0.002453	0.141735
SIK1	0.000606	0.107488	0.000239	0.043088	0.002768	0.150281
PIGK	0.000871	0.117955	0.000018	0.021860	0.000170	0.037648
PTS	0.000924	0.121164	0.000149	0.040780	0.000339	0.052868
AP1S3	0.001885	0.153168	0.000185	0.041834	0.000168	0.037648
USP43	0.000718	0.107690	0.000088	0.033291	0.000008	0.006470
OXA1L	0.000390	0.094375	0.000054	0.029634	0.001026	0.091669
SFXN1	0.000632	0.107690	0.000298	0.046842	0.000459	0.059621
HEY1	0.000162	0.059991	0.000044	0.029182	0.000008	0.006404
SERPINB8	0.004195	0.205281	0.000107	0.036702	0.000965	0.088073
LMBRD1	0.013413	0.306975	0.000235	0.042956	0.003757	0.176570
SLC30A1	0.008443	0.256237	0.000302	0.046842	0.002911	0.156679
EMB	0.005621	0.222630	0.000287	0.046302	0.003475	0.173662
MIR4435-1HG	0.000108	0.054628	0.000190	0.041834	0.001254	0.104376
COQ2	0.000233	0.075168	0.000056	0.029634	0.000189	0.039019
SLC36A4	0.000128	0.055766	0.000056	0.029634	0.000778	0.078421
CA13	0.002549	0.171514	0.000056	0.029634	0.000526	0.065161
LINC00313	0.000175	0.063084	0.000206	0.042274	0.000065	0.021531
LSAMP	0.001274	0.136012	0.000112	0.036702	0.000264	0.046143
PCLO	0.000624	0.107690	0.000198	0.041834	0.000575	0.067734
ADAT2	0.001675	0.147281	0.000199	0.041834	0.002384	0.140941
RP11-317F20.2	0.001409	0.138373	0.000066	0.032126	0.000543	0.066415
PTCHD4	0.001856	0.151830	0.000198	0.041834	0.000341	0.052868
KIAA1456	0.000651	0.107690	0.000275	0.046302	0.000402	0.055650
RP13-870H17.3	0.000221	0.073219	0.000179	0.041834	0.000001	0.001652
DLGAP1-AS2	0.000699	0.107690	0.000114	0.036702	0.000361	0.053396
RP11-388C12.8	0.000749	0.108838	0.000338	0.049147	0.000344	0.052868
RP11-173M1.8	0.000329	0.088221	0.000282	0.046302	0.000575	0.067734

RP11-250B2.6	0.001810	0.150997	0.000347	0.049504	0.000045	0.016636
RP11-78A19.4	0.000509	0.103086	0.000208	0.042274	0.000039	0.015248
KLHL13	0.002670	0.175150	0.000566	0.059949	0.000213	0.040665
CRLF1	0.004110	0.204605	0.022546	0.343812	0.000254	0.045726
KCNG1	0.007702	0.245959	0.016403	0.295745	0.000224	0.041298
DCBLD2	0.000837	0.115703	0.000481	0.056976	0.000020	0.009895
SLC4A4	0.004708	0.212790	0.003520	0.137700	0.000128	0.032479
ARG2	0.000496	0.103086	0.004002	0.148680	0.000039	0.015248
BCKDHB	0.001417	0.138373	0.000481	0.056976	0.000026	0.012253
SEMA6A	0.000689	0.107690	0.000587	0.060898	0.000205	0.039908
SLC7A5	0.000376	0.094375	0.000515	0.057946	0.000033	0.014188
NEFL	0.004438	0.207062	0.000616	0.061050	0.000085	0.025384
ST3GAL4	0.000362	0.094375	0.000833	0.069152	0.000011	0.006872
PCK1	0.006273	0.235304	0.014315	0.277824	0.000143	0.034551
TCN1	0.001092	0.126656	0.002627	0.124392	0.000046	0.016636
SERPINE2	0.001046	0.123119	0.003496	0.137700	0.000276	0.047122
BBS7	0.005217	0.219309	0.000700	0.063849	0.000277	0.047122
BMPR1B	0.002349	0.166353	0.000865	0.069694	0.000081	0.025384
FXYD4	0.003440	0.192588	0.012183	0.259455	0.000051	0.017795
DLG5	0.000277	0.076091	0.000787	0.067189	0.000012	0.007493
SVOPL	0.002050	0.160407	0.009419	0.227360	0.000113	0.031400
SLC16A14	0.003750	0.200135	0.001147	0.081555	0.000226	0.041298
HSPA4L	0.000134	0.056092	0.000580	0.060864	0.000170	0.037648
PDE3A	0.003802	0.201635	0.000807	0.068198	0.000206	0.039908
GCNT1	0.001012	0.123119	0.000647	0.062829	0.000115	0.031400
H1F0	0.000938	0.121787	0.004443	0.158387	0.000218	0.041060
CD55	0.000600	0.107488	0.000590	0.060898	0.000156	0.036484
AJAP1	0.000146	0.057276	0.000410	0.053627	0.000002	0.002157
BPIFA1	0.004096	0.204605	0.084953	0.596494	0.000084	0.025384
SERPINB4	0.001222	0.131591	0.013995	0.274992	0.000015	0.007716
RP13-39P12.2	0.001166	0.127737	0.004755	0.164987	0.000010	0.006872
PAXIP1-AS2	0.004152	0.205063	0.004911	0.168066	0.000149	0.035502
TMEM213	0.003810	0.201635	0.000493	0.057217	0.000204	0.039908
LINC00689	0.004876	0.215000	0.014491	0.279926	0.000097	0.027785
RP11-250B2.3	0.005366	0.219625	0.006030	0.180189	0.000187	0.039019
SLC5A8	0.004864	0.215000	0.002516	0.120475	0.000088	0.025804
RP11-250B2.5	0.004359	0.205601	0.013736	0.271173	0.000258	0.045726

Table 4-5. CRTC1/3-MAML2-associated gene signature

correlation with fusion enriched genes	NAME	SIZE	ES	NES	NOM pval	FDR qval	FWER pval	RANK AT MAX	LEADING EDGE
negatively	TGTTTGY_HNF3_Q6	434	-0.201				1	0	1625	tags=20%,list=13%,signal=23%
negatively	GO_MUSCLE_STRUCTURE_DEVELOPMENT	392	-0.154				1	0	2051	tags=21%,list=17%,signal=25%
negatively	GO_RAS_PROTEIN_SIGNAL_TRANSDUCTION	321	-0.205				1	0	1943	tags=24%,list=16%,signal=27%
negatively	GO_COVALENT_CHROMATIN_MODIFICATION	363	-0.175				1	0	1615	tags=18%,list=13%,signal=20%
negatively	GO_REGULATION_OF_CYTOSKELETON_ORGANIZATION	376	-0.197				1	0	2671	tags=32%,list=22%,signal=39%
negatively	GO_REGULATION_OF_GTPASE_ACTIVITY	344	-0.208				1	0	1831	tags=24%,list=15%,signal=27%
negatively	GO_SUPRAMOLECULAR_FIBER_ORGANIZATION	468	-0.182				1	0	2403	tags=27%,list=19%,signal=32%
negatively	GO_MUSCLE_FILAMENT_SLIDING	26	-0.635	-2.46	0	0.002	0.006	3158	tags=69%,list=25%,signal=93%
negatively	GO_CELL_PROLIFERATION_INVOLVED_IN_KIDNEY_DEVELOPMENT	15	-0.705	-2.35	0	0.003	0.011	1404	tags=40%,list=11%,signal=45%
negatively	GO_REGULATION_OF_SMOOTH_MUSCLE_CELL_DIFFERENTIATION	24	-0.536	-2.06	0	0.049	0.207	2498	tags=46%,list=20%,signal=57%
positively	HALLMARK_OXIDATIVE_PHOSPHORYLATION	184	0.616	2.29	0	0	0	2610	tags=60%,list=21%,signal=75%
positively	GO_ATP_SYNTHESIS_COUPLED_ELECTRON_TRANSPORT	87	0.621	2.19	0	4.4E-04	0.001	2610	tags=63%,list=21%,signal=80%
positively	GO_RESPIRATORY_ELECTRON_TRANSPORT_CHAIN	102	0.595	2.13	0	2.9E-04	0.001	2610	tags=59%,list=21%,signal=74%
positively	GO_OXIDATIVE_PHOSPHORYLATION	106	0.590	2.09	0	1.1E-03	0.005	2686	tags=58%,list=22%,signal=74%
positively	GO_AEROBIC_RESPIRATION	75	0.592	2.06	0	2.3E-03	0.013	2273	tags=55%,list=18%,signal=67%
positively	GO_MITOCHONDRIAL_TRANSLATIONAL_TERMINATION	86	0.580	2.04	0	2.6E-03	0.018	2708	tags=51%,list=22%,signal=65%
positively	GO_CELLULAR_RESPIRATION	164	0.547	2.01	0	3.8E-03	0.029	2630	tags=51%,list=21%,signal=64%
positively	GO_MITOCHONDRIAL_ELECTRON_TRANSPORT_NADH_TO_UBIQUINONE	52	0.612	2.01	0	4.0E-03	0.035	3380	tags=69%,list=27%,signal=95%
positively	GO_INNER_MITOCHONDRIAL_MEMBRANE_ORGANIZATION	24	0.688	1.96	0.0012	7.8E-03	0.074	1972	tags=58%,list=16%,signal=69%
positively	GO_MITOCHONDRIAL_TRANSLATION	125	0.538	1.95	0	8.7E-03	0.093	3340	tags=54%,list=27%,signal=73%
positively	GO_MITOCHONDRIAL_RESPIRATORY_CHAIN_COMPLEX_ASSEMBLY	87	0.555	1.94	0	9.8E-03	0.113	3912	tags=66%,list=32%,signal=95%
positively	GO_ENERGY_DERIVATION_BY_OXIDATION_OF_ORGANIC_COMPOUNDS	217	0.518	1.93	0	1.1E-02	0.139	2630	tags=45%,list=21%,signal=56%
positively	GO_TRANSLATIONAL_TERMINATION	96	0.550	1.93	0	1.0E-02	0.139	2708	tags=48%,list=22%,signal=61%
positively	GO_AEROBIC_ELECTRON_TRANSPORT_CHAIN	18	0.707	1.92	0.0013	1.2E-02	0.166	2610	tags=78%,list=21%,signal=98%
positively	GO_MITOCHONDRIAL_GENE_EXPRESSION	144	0.529	1.91	0	1.3E-02	0.192	3392	tags=53%,list=27%,signal=72%
positively	GO_PYRUVATE_METABOLIC_PROCESS	20	0.692	1.91	0	1.2E-02	0.196	2347	tags=65%,list=19%,signal=80%
positively	GO_ELECTRON_TRANSPORT_CHAIN	151	0.515	1.90	0	1.4E-02	0.237	3729	tags=57%,list=30%,signal=80%
positively	GO_NADH_DEHYDROGENASE_COMPLEX_ASSEMBLY	61	0.570	1.90	0	1.4E-02	0.246	3380	tags=64%,list=27%,signal=87%
positively	GO_DICARBOXYLIC_ACID_METABOLIC_PROCESS	40	0.589	1.86	0.0011	2.4E-02	0.403	3054	tags=63%,list=25%,signal=83%
positively	GO_NEGATIVE_REGULATION_OF_PROTEIN_MATURATION	18	0.683	1.85	0.0013	3.0E-02	0.493	1372	tags=39%,list=11%,signal=44%
positively	GO_TRICARBOXYLIC_ACID_CYCLE	31	0.605	1.83	0.0012	3.7E-02	0.59	2442	tags=58%,list=20%,signal=72%
positively	GO_NEUTRAL_AMINO_ACID_TRANSPORT	20	0.665	1.83	0	3.6E-02	0.598	2557	tags=50%,list=21%,signal=63%
positively	GO_REGULATION_OF_SYSTEMIC_ARTERIAL_BLOOD_PRESSURE_MEDIATED_BY_A_CHEMICAL_SIGNAL	19	0.657	1.81	0.0025	4.5E-02	0.686	1784	tags=47%,list=14%,signal=55%
positively	GO_POLYSACCHARIDE_CATABOLIC_PROCESS	19	0.660	1.81	0	4.9E-02	0.725	1552	tags=42%,list=13%,signal=48%

Table 4-6. Gene set enrichment analysis for CRTC1/3-MAML2-associated gene signature.

Gene	Ordinal Logistic regression		Ordinal Logistic regression log		Linear regression		Linear regression log FPKM	
	P-value	Adjusted P	P-value	Adjusted P	P-value	Adjusted P	P-value	Adjusted P
Y_RNA	0.99485	0.99793	0.00000	0.00000	0.11171	0.36620	0.04222	0.24218
NISCH	0.00344	0.22689	0.00072	0.14085	0.00088	0.13227	0.00013	0.04678
PDE4A	0.00098	0.22689	0.00076	0.14085	0.00021	0.10275	0.00017	0.04933
ZFY	0.00218	0.22689	0.00064	0.14085	0.00027	0.10275	0.00019	0.04933
ACADVL	0.00395	0.22689	0.00083	0.14085	0.00287	0.13439	0.00021	0.04933
ATP2A3	0.00102	0.22689	0.00008	0.14085	0.00060	0.13227	0.00000	0.02674
SEC31B	0.00135	0.22689	0.00081	0.14085	0.00073	0.13227	0.00024	0.04933
CAPN3	0.00118	0.22689	0.00043	0.14085	0.00678	0.14856	0.00019	0.04933
GGA1	0.00066	0.22689	0.00101	0.14085	0.00035	0.12137	0.00016	0.04933
CBX7	0.00101	0.22689	0.00030	0.14085	0.00156	0.13227	0.00007	0.04037
SNRNP70	0.00047	0.22689	0.00015	0.14085	0.00022	0.10275	0.00003	0.02674
PTGDS	0.01200	0.22689	0.00057	0.14085	0.00069	0.13227	0.00011	0.04508
RASSF4	0.00366	0.22689	0.00077	0.14085	0.00090	0.13227	0.00021	0.04933
PHF1	0.00082	0.22689	0.00044	0.14085	0.00024	0.10275	0.00010	0.04412
APBB3	0.00075	0.22689	0.00045	0.14085	0.00024	0.10275	0.00007	0.04037
TMCO6	0.00367	0.22689	0.00092	0.14085	0.00195	0.13344	0.00026	0.04933
C3orf14	0.00418	0.22689	0.00107	0.14085	0.00051	0.13227	0.00016	0.04933
HEMK1	0.00124	0.22689	0.00103	0.14085	0.00041	0.12618	0.00026	0.04933
MUC5B	0.03601	0.26288	0.00055	0.14085	0.09188	0.32954	0.00007	0.04037
FAM53C	0.00084	0.22689	0.00049	0.14085	0.00014	0.10275	0.00008	0.04037
EGR1	0.00571	0.22689	0.00034	0.14085	0.00707	0.14881	0.00007	0.04037
NR4A1	0.00885	0.22689	0.00014	0.14085	0.00156	0.13227	0.00001	0.02674
FOSB	0.03680	0.26379	0.00023	0.14085	0.01105	0.16603	0.00002	0.02674
TMEM175	0.00162	0.22689	0.00064	0.14085	0.00058	0.13227	0.00015	0.04933
TUBGCP6	0.00358	0.22689	0.00075	0.14085	0.00060	0.13227	0.00009	0.04202
SHC2	0.00259	0.22689	0.00059	0.14085	0.00103	0.13227	0.00009	0.04188
SLC27A1	0.00082	0.22689	0.00052	0.14085	0.00020	0.10275	0.00011	0.04508
MYBBP1A	0.00161	0.22689	0.00078	0.14085	0.00058	0.13227	0.00023	0.04933
PAN2	0.00112	0.22689	0.00081	0.14085	0.00054	0.13227	0.00025	0.04933
CYP27A1	0.00155	0.22689	0.00064	0.14085	0.00557	0.14193	0.00020	0.04933
UBE2Q2	0.00040	0.22689	0.00055	0.14085	0.00009	0.10275	0.00019	0.04933
SGSM2	0.00055	0.22689	0.00015	0.14085	0.00008	0.10275	0.00002	0.02674
SCRN2	0.00338	0.22689	0.00061	0.14085	0.00127	0.13227	0.00012	0.04678
CSRNP1	0.00304	0.22689	0.00090	0.14085	0.00560	0.14193	0.00017	0.04933
NICN1	0.00026	0.22689	0.00011	0.14085	0.00002	0.10275	0.00001	0.02674
AKAP6	0.00261	0.22689	0.00094	0.14085	0.00100	0.13227	0.00025	0.04933
RPUSD3	0.00058	0.22689	0.00066	0.14085	0.00015	0.10275	0.00026	0.04933
BTG2	0.00361	0.22689	0.00031	0.14085	0.00114	0.13227	0.00007	0.04037
C5orf45	0.00084	0.22689	0.00042	0.14085	0.00027	0.10275	0.00008	0.04037
SLC25A45	0.00048	0.22689	0.00023	0.14085	0.00016	0.10275	0.00003	0.02674
PIGR	0.01212	0.22689	0.00022	0.14085	0.00878	0.15738	0.00002	0.02674
NDST2	0.00031	0.22689	0.00019	0.14085	0.00007	0.10275	0.00002	0.02674
ENGASE	0.00290	0.22689	0.00041	0.14085	0.00138	0.13227	0.00013	0.04678
LENG8	0.00052	0.22689	0.00011	0.14085	0.00010	0.10275	0.00001	0.02674
INPP5D	0.00268	0.22689	0.00092	0.14085	0.00144	0.13227	0.00023	0.04933
FOS	0.00283	0.22689	0.00034	0.14085	0.00273	0.13439	0.00003	0.02674
MZB1	0.00705	0.22689	0.00102	0.14085	0.00157	0.13227	0.00026	0.04933
MUC7	0.01664	0.23375	0.00126	0.14085	0.10618	0.35657	0.00024	0.04933
GMPPB	0.00113	0.22689	0.00030	0.14085	0.00016	0.10275	0.00002	0.02674
PACS1	0.00233	0.22689	0.00092	0.14085	0.00133	0.13227	0.00025	0.04933
JUN	0.01359	0.22716	0.00095	0.14085	0.01109	0.16603	0.00025	0.04933
EFCAB4A	0.00269	0.22689	0.00027	0.14085	0.00242	0.13439	0.00003	0.03069
PLA2G6	0.00082	0.22689	0.00053	0.14085	0.00025	0.10275	0.00018	0.04933
TPCN1	0.00071	0.22689	0.00048	0.14085	0.00016	0.10275	0.00014	0.04887
SEMA4D	0.00227	0.22689	0.00048	0.14085	0.00151	0.13227	0.00015	0.04933
AKAP17A	0.00092	0.22689	0.00053	0.14085	0.00045	0.13227	0.00017	0.04933
C1orf132	0.00435	0.22689	0.00041	0.14085	0.00096	0.13227	0.00004	0.03449
RXRB	0.00067	0.22689	0.00034	0.14085	0.00060	0.13227	0.00023	0.04933
MZT1	0.00036	0.22689	0.00091	0.14085	0.00004	0.10275	0.00023	0.04933
MIR29A	0.00706	0.22689	0.00094	0.14085	0.00428	0.14082	0.00023	0.04933
RP11-274B21.4	0.00073	0.22689	0.00050	0.14085	0.00041	0.12618	0.00006	0.04037
LINC00852	0.00193	0.22689	0.00062	0.14085	0.00058	0.13227	0.00015	0.04933
LINC-PINT	0.00276	0.22689	0.00132	0.14085	0.00152	0.13227	0.00023	0.04933
RPL3P4	0.06717	0.31089	0.00399	0.15437	0.01204	0.16724	0.00022	0.04933
RP11-274B21.2	0.00045	0.22689	0.00120	0.14085	0.00027	0.10275	0.00010	0.04412
RP11-274B21.3	0.00035	0.22689	0.00019	0.14085	0.00024	0.10275	0.00001	0.02674
STX16-NPEPL1	0.00060	0.22689	0.00026	0.14085	0.00019	0.10275	0.00007	0.04037
CTD-2622I13.3	0.00046	0.22689	0.00028	0.14085	0.00007	0.10275	0.00006	0.04037

Table 4-7.MEC grade-associated gene signature.

Correlation with grade enriched genes	NAME	SIZE	ES	NES	NOM p-val	FDR q-val	FWER p-val	RANK AT MAX
negatively	GO_COMPLEMENT_ACTIVATION	109	-0.69	-2.67	0.0000	0	0	3075
negatively	GO_HUMORAL_IMMUNE_RESPONSE_MEDIATED_BY_CIRCULATING_IMMUNOGLOBULIN	101	-0.69	-2.66	0.0000	0	0	3237
negatively	GO_PHAGOCYTOSIS_RECOGNITION	54	-0.73	-2.54	0.0000	0	0	2692
negatively	GO_B_CELL_RECEPTOR_SIGNALING_PATHWAY	83	-0.66	-2.50	0.0000	0	0	3164
negatively	GO_HUMORAL_IMMUNE_RESPONSE	191	-0.61	-2.47	0.0000	0	0	3237
negatively	GO_B_CELL_MEDIATED_IMMUNITY	142	-0.61	-2.44	0.0000	0	0	3164
negatively	GO_REGULATION_OF_HUMORAL_IMMUNE_RESPONSE	88	-0.65	-2.42	0.0000	0	0	3116
negatively	GO_POSITIVE_REGULATION_OF_B_CELL_ACTIVATION	95	-0.61	-2.32	0.0000	0	0	3384
negatively	GO_MEMBRANE_INVAGINATION	93	-0.59	-2.25	0.0000	0	0	3164
negatively	GO_IMMUNOGLOBULIN_PRODUCTION	116	-0.57	-2.24	0.0000	0	0	3143
negatively	GO_CELL_RECOGNITION	110	-0.57	-2.20	0.0000	0	0	3164
negatively	GO_DEFENSE_RESPONSE_TO_BACTERIUM	155	-0.54	-2.18	0.0000	4.7E-05	0.001	3223
negatively	GO_REGULATION_OF_B_CELL_ACTIVATION	123	-0.55	-2.17	0.0000	4.5E-05	0.001	3528
negatively	GO_ADAPTIVE_IMMUNE_RESPONSE	305	-0.52	-2.16	0.0000	1.3E-04	0.003	3528
negatively	GO_ADAPTIVE_IMMUNE_RESPONSE_BASED_ON_SOMATIC_RECOMBINATION_OF_IMMUNE_RECEPTORS_BUILT_FROM_IMMUNOGLOBULIN_SUPERFAMILY_DOMAINS	228	-0.53	-2.15	0.0000	1.2E-04	0.003	3186
negatively	GO_B_CELL_ACTIVATION	197	-0.52	-2.11	0.0000	5.1E-04	0.013	3174
negatively	HALLMARK_MYOGENESIS	153	-0.52	-2.10	0.0000	5.2E-04	0.014	3757
negatively	GO_LYMPHOCYTE_MEDIATED_IMMUNITY	211	-0.51	-2.08	0.0000	9.0E-04	0.025	3186
negatively	GO_PHAGOCYTOSIS	252	-0.50	-2.03	0.0000	1.5E-03	0.042	3257
negatively	GO_FC_RECEPTOR_MEDIATED_STIMULATORY_SIGNALING_PATHWAY	126	-0.52	-2.03	0.0000	1.5E-03	0.045	2800
negatively	GO_POSITIVE_REGULATION_OF_LYMPHOCYTE_ACTIVATION	210	-0.50	-2.02	0.0000	1.5E-03	0.046	3400
negatively	GO_MATURE_B_CELL_DIFFERENTIATION	18	-0.69	-2.01	0.0000	1.9E-03	0.058	2828
negatively	GO_PRODUCTION_OF_MOLECULAR_MEDIATOR_OF_IMMUNE_RESPONSE	174	-0.50	-2.00	0.0000	2.1E-03	0.066	3186
negatively	GO_POSITIVE_REGULATION_OF_CELL_ACTIVATION	240	-0.48	-1.96	0.0000	3.8E-03	0.119	3400
negatively	GO_RECEPTOR_MEDIATED_ENDOCYTOSIS	230	-0.47	-1.94	0.0000	5.6E-03	0.175	2819
negatively	GO_ANTIBACTERIAL_HUMORAL_RESPONSE	25	-0.63	-1.93	0.0000	5.7E-03	0.183	1284
negatively	GO_LEUKOCYTE_MIGRATION	312	-0.46	-1.92	0.0000	7.6E-03	0.242	3906
negatively	HALLMARK_TNFA_SIGNALING_VIA_NFKB	177	-0.48	-1.91	0.0000	8.6E-03	0.273	3823
negatively	NRSF_01	15	-0.70	-1.91	0.0000	9.0E-03	0.294	1818
negatively	GO_GLYCEROLIPID_CATABOLIC_PROCESS	33	-0.59	-1.90	0.0000	9.4E-03	0.312	2884
negatively	GO_ALPHA_BETA_T_CELL_DIFFERENTIATION	65	-0.52	-1.89	0.0000	1.2E-02	0.391	2659
negatively	GO_NEGATIVE_REGULATION_OF_INTERLEUKIN_6_PRODUCTION	22	-0.63	-1.88	0.0013	1.4E-02	0.443	3604
negatively	GO_RESPONSE_TO_PROSTAGLANDIN_E	17	-0.66	-1.87	0.0013	1.4E-02	0.456	2778
negatively	GO_DENDRITIC_CELL_MIGRATION	16	-0.69	-1.87	0.0025	1.5E-02	0.487	2778
negatively	GO_POSITIVE_T_CELL_SELECTION	17	-0.66	-1.86	0.0039	1.7E-02	0.539	2619
negatively	GO_POSITIVE_REGULATION_OF_LIPID_LOCALIZATION	45	-0.54	-1.86	0.0000	1.8E-02	0.566	2980
negatively	GO_T_CELL_ACTIVATION_INVOLVED_IN_IMMUNE_RESPONSE	54	-0.52	-1.85	0.0000	1.8E-02	0.576	3326
negatively	GO_ENDOCYTOSIS	445	-0.44	-1.85	0.0000	1.8E-02	0.577	2819
negatively	GO_PROTEIN_KINASE_A_SIGNALING	17	-0.65	-1.84	0.0026	2.2E-02	0.671	2289
negatively	GO_ALPHA_BETA_T_CELL_ACTIVATION	79	-0.49	-1.84	0.0000	2.2E-02	0.676	2659
negatively	GO_REGULATION_OF_LYMPHOCYTE_ACTIVATION	298	-0.44	-1.83	0.0000	2.4E-02	0.715	3237
negatively	GO_POSITIVE_REGULATION_OF_REACTIVE_OXYGEN_SPECIES_BIOSYNTHETIC_PROCESS	35	-0.56	-1.83	0.0000	2.4E-02	0.722	2958
negatively	GO_SEQUESTERING_OF_CALCIUM_ION	74	-0.49	-1.82	0.0000	2.5E-02	0.736	3653
negatively	GO_LYMPHOCYTE_ACTIVATION	446	-0.43	-1.82	0.0000	2.5E-02	0.745	3237
negatively	GO_NEUTRAL_LIPID_CATABOLIC_PROCESS	18	-0.63	-1.82	0.0013	2.6E-02	0.76	2580
negatively	GO_FC_EPSILON_RECEPTOR_SIGNALING_PATHWAY	151	-0.45	-1.81	0.0000	2.7E-02	0.791	2800
negatively	GO_ANTIMICROBIAL_HUMORAL_RESPONSE	61	-0.50	-1.81	0.0022	2.7E-02	0.798	2478
negatively	LFA1_Q6	148	-0.46	-1.81	0.0000	2.7E-02	0.802	3843
negatively	GO_RENAL_WATER_HOMEOSTASIS	22	-0.60	-1.81	0.0025	2.7E-02	0.807	3041
negatively	GO_IMMUNE_RESPONSE_REGULATING_CELL_SURFACE_RECEPTOR_SIGNALING_PATHWAY	357	-0.43	-1.81	0.0000	2.7E-02	0.819	2800
negatively	GO_CHEMICAL_SYNAPTIC_TRANSMISSION_POSTSYNAPTIC	34	-0.55	-1.80	0.0024	2.8E-02	0.831	2191
negatively	SRF_01	39	-0.54	-1.80	0.0000	2.9E-02	0.842	1697
negatively	CCAGGGG_MIR331	66	-0.50	-1.80	0.0000	3.0E-02	0.85	4603
negatively	GO_REGULATION_OF_IMMUNE_EFFECTOR_PROCESS	286	-0.44	-1.80	0.0000	3.0E-02	0.855	3922
negatively	GO_CALCIUM_ION_TRANSMEMBRANE_IMPORT_INTO_CYTOSOL	74	-0.48	-1.80	0.0000	3.0E-02	0.857	3653
negatively	GO_B_CELL_PROLIFERATION	46	-0.53	-1.79	0.0000	3.1E-02	0.877	2619
negatively	GO_MONOAMINE_TRANSPORT	22	-0.60	-1.79	0.0051	3.1E-02	0.879	2656
negatively	GO_MACROPHAGE_ACTIVATION	54	-0.51	-1.79	0.0000	3.3E-02	0.904	4411
negatively	GO_LYMPHOCYTE_DIFFERENTIATION	209	-0.44	-1.78	0.0000	3.4E-02	0.915	3400
negatively	GO_POSITIVE_REGULATION_OF_CATION_CHANNEL_ACTIVITY	28	-0.57	-1.78	0.0062	3.4E-02	0.915	3672
negatively	GO_APOPTOTIC_CELL_CLEARANCE	27	-0.56	-1.78	0.0024	3.3E-02	0.917	3421
negatively	GO_INOSITOL_PHOSPHATE_BIOSYNTHETIC_PROCESS	21	-0.61	-1.78	0.0025	3.4E-02	0.925	1821
negatively	GO_RESPONSE_TO_BACTERIUM	379	-0.43	-1.78	0.0000	3.4E-02	0.927	3906
negatively	GO_BICARBONATE_TRANSPORT	21	-0.60	-1.78	0.0000	3.5E-02	0.934	3801
negatively	GO_REGULATION_OF_CELL_ACTIVATION	367	-0.42	-1.77	0.0000	3.6E-02	0.943	3237
negatively	GO_CYTOSOLIC_CALCIUM_ION_TRANSPORT	92	-0.47	-1.77	0.0000	3.8E-02	0.951	3653
negatively	GO_ORGAN_OR_TISSUE_SPECIFIC_IMMUNE_RESPONSE	19	-0.61	-1.77	0.0026	3.8E-02	0.953	2801
negatively	GO_RESPONSE_TO_CHEMOKINE	42	-0.52	-1.76	0.0012	3.9E-02	0.958	3311
negatively	GO_REGULATION_OF_BONE_REMODELING	29	-0.55	-1.76	0.0012	3.9E-02	0.958	1257
negatively	GO_CD4_POSITIVE_ALPHA_BETA_T_CELL_ACTIVATION	53	-0.50	-1.76	0.0023	3.9E-02	0.959	2659
negatively	GO_MUSCLE_FILAMENT_SLIDING	26	-0.57	-1.76	0.0013	4.1E-02	0.965	3282
negatively	GO_T_CELL_SELECTION	26	-0.56	-1.75	0.0024	4.4E-02	0.972	4156
negatively	GO_REGULATION_OF_CATION_CHANNEL_ACTIVITY	74	-0.47	-1.75	0.0000	4.4E-02	0.973	3835
negatively	GO_ANTIGEN_RECEPTOR_MEDIATED_SIGNALING_PATHWAY	221	-0.43	-1.75	0.0000	4.5E-02	0.978	2778
negatively	GO_NEGATIVE_REGULATION_OF_MUSCLE_CELL_APOPTOTIC_PROCESS	26	-0.56	-1.75	0.0025	4.5E-02	0.979	3584
negatively	GO_POSITIVE_REGULATION_OF_LYMPHOCYTE_DIFFERENTIATION	58	-0.48	-1.75	0.0000	4.5E-02	0.98	2659
negatively	GO_SPROUTING_ANGIOGENESIS	84	-0.46	-1.75	0.0000	4.4E-02	0.98	3452
negatively	GO_NEGATIVE_REGULATION_OF_ACTIN_FILAMENT_BUNDLE_ASSEMBLY	16	-0.63	-1.75	0.0026	4.4E-02	0.981	1275
negatively	GO_T_CELL_DIFFERENTIATION_INVOLVED_IN_IMMUNE_RESPONSE	41	-0.51	-1.74	0.0023	4.8E-02	0.986	4411
negatively	GO_REGULATION_OF_CALCIUM_ION_TRANSPORT	122	-0.45	-1.74	0.0000	5.0E-02	0.988	3848
negatively	GO_REGULATION_OF_MACROPHAGE_ACTIVATION	36	-0.52	-1.74	0.0036	4.9E-02	0.988	4411
positively	HALLMARK_E2F_TARGETS	136	0.44	2.53	0.0000	1.5E-03	0.001	2950
positively	HALLMARK_MYC_TARGETS_V1	189	0.44	2.53	0.0000	7.5E-04	0.001	3191
positively	ATAGGAA_MIR202	85	0.43	2.32	0.0000	1.2E-02	0.025	1877

Table 4-8. Gene set enrichment analysis for CRTC1/3-MAML2-associated gene signature.

Sample ID	Gene 1 symbol (5')	Gene 2 symbol (3')	Spanning unique reads	Fusion point for gene 1(5')	Fusion point for gene 2(3')	Predicted effect
81353	CRTC1	MAML2	3	19:18794638:+	11:95826681:-	in-frame
81356	CRTC1	MAML2	4	19:18794638:+	11:95826681:-	in-frame
81357	CRTC1	MAML2	5	19:18794638:+	11:95826681:-	in-frame
81358	CRTC3	MAML2	14	15:91083369:+	11:95826681:-	in-frame
81359	CRTC1	MAML2	3	19:18794638:+	11:95826681:-	in-frame
81366	CRTC1	MAML2	5	19:18794638:+	11:95826681:-	in-frame
81387	CRTC1	MAML2	3	19:18794638:+	11:95826681:-	in-frame
81390	CRTC1	MAML2	4	19:18794638:+	11:95826681:-	in-frame
81560	CRTC1	MAML2	6	19:18794638:+	11:95826681:-	in-frame
85191	CRTC1	MAML2	4	19:18794638:+	11:95826681:-	in-frame
85195	CRTC1	MAML2	3	19:18794638:+	11:95826681:-	in-frame
81390	MAML2	CRTC1	3	11:96074547:-	19:18853720:+	in-frame
81391	MAML2	CRTC1	3	11:96074547:-	19:18853720:+	in-frame
81560	MAML2	CRTC1	3	11:96074547:-	19:18853720:+	in-frame
81352	AC022144.1	NEAT1	5	19:39230088:+	11:65208324:+	
81352	AC024270.2	FABP5	3	15:83713492:+	8:82196772:+	
81352	OR4K8P	OR4K15	3	18:14614138:+	14:20444680:+	
81354	AC022144.1	NEAT1	3	19:39230089:+	11:65208314:+	
81354	AMD1	AC010435.1	4	6:111214819:+	5:43587019:-	
81354	AMD1	AC010435.1	3	6:111215135:+	5:43586697:-	
81354	BHLHE40	NEAT1	8	3:5025513:+	11:65211735:+	
81354	FP475955.3	AC012414.7	3	21:15027024:+	15:21023421:-	
81355	ZNF160	ZNF468	3	19:53589307:-	19:53357305:-	
81356	AC091982.3	GLUL	4	5:151229721:+	1:182355007:-	
81356	AC091982.3	GLUL	4	5:151230346:+	1:182353546:-	
81356	AC091982.3	GLUL	4	5:151230301:+	1:182353591:-	
81356	AC091982.3	GLUL	3	5:151230176:+	1:182353716:-	
81356	AL445305.1	AP3S1	6	1:214656362:-	5:115238606:+	
81356	ARHGEF12	NEAT1	4	11:120334260:+	11:65208344:+	
81356	FABP5	AC104046.1	4	8:82196178:+	15:85958784:+	
81356	GLUL	AC091982.3	3	1:182353594:-	5:151230304:+	
81356	GLUL	AC091982.3	3	1:182354615:-	5:151229920:+	
81356	GLUL	AC091982.3	3	1:182354511:-	5:151230023:+	
81356	GLUL	AC091982.3	3	1:182353751:-	5:151230149:+	
81357	AL450405.1	RPL29	3	6:118320280:+	3:52028098:-	
81357	IMPDH1	AC091807.1	3	7:128036987:-	X:40219071:-	
81357	PKD1	NPIPA1	11	16:2147729:-	16:15035669:+	
81357	RPL29	AL450405.1	4	3:52029455:-	6:118320165:+	
81357	RPL29	AL450405.1	3	3:52028062:-	6:118320141:+	
81357	TPM1	KYNU	5	15:63362489:+	2:143800174:+	
81358	AC091807.1	IMPDH1	4	X:40218582:-	7:128034549:-	
81358	AC091807.1	AC021305.1	7	X:40218582:-	8:124414204:+	
81358	AL109933.2	LPA	4	6:161087260:-	6:161071529:-	
81358	IMPDH1	AC091807.1	3	7:128043764:-	X:40219839:-	
81358	NPIPA1	PKD1	4	16:15023278:+	16:2153627:-	
81358	PKD1	NPIPA1	6	16:2147729:-	16:15035669:+	
81359	AL445305.1	AP3S1	3	1:214656362:-	5:115238606:+	
81359	NEAT1	GPRC5A	4	11:65211769:+	12:13070192:+	
81360	AC022144.1	NEAT1	6	19:39230088:+	11:65208321:+	
81362	AC091807.1	IMPDH1	4	X:40218582:-	7:128034549:-	
81362	IMPDH1	AC091807.1	3	7:128034618:-	X:40218649:-	
81364	AC138409.2	AL021368.3	7	5:34168418:-	6:58238477:-	

81364	AL445305.1	AP3S1	7	1:214656364:-	5:115238607:+
81364	AL450405.1	RPL29	3	6:118320506:+	3:52027881:-
81364	RPL29	AL450405.1	6	3:52027872:-	6:118320501:+
81364	RPL29	AL450405.1	4	3:52028108:-	6:118320149:+
81366	AC022144.1	NEAT1	5	19:39230089:+	11:65208314:+
81367	AC022144.1	NEAT1	3	19:39230088:+	11:65208317:+
81368	AC022144.1	NEAT1	7	19:39230088:+	11:65208322:+
81368	AC073621.1	RPL13	5	17:17286700:-	16:89629443:+
81368	AL445305.1	AP3S1	5	1:214656362:-	5:115238606:+
81368	COL1A1	FOXP1	3	17:48262638:-	3:71005918:-
81368	EBF1	COL1A1	3	5:158123903:-	17:48262637:-
81368	YWHAE	AC092447.2	3	17:1268241:-	7:56636450:-
81375	AC022144.1	NEAT1	3	19:39230088:+	11:65208323:+
81375	AL356585.1	EIF3F	3	Un_gl000212:35601:-	11:8016588:+
81375	EIF3F	AL356585.1	3	11:8016616:+	Un_gl000212:35571:-
81378	AC022144.1	NEAT1	5	19:39230088:+	11:65208319:+
81379	AC009302.1	LLPH	3	2:222809288:+	12:66517662:-
81379	AC118282.2	AC119751.4	4	4:49214489:-	4:49582002:+
81383	IGH2	AC026523.2	4	14:107191071:+	15:98083035:-
81384	SF3A3	AP002765.1	4	1:38455612:-	11:123470134:-
81386	BACH2	NEAT1	7	6:90639864:-	11:65196969:+
81387	RPL13	AC073621.1	3	16:89628134:+	17:17286928:-
81389	AC022144.1	NEAT1	3	19:39230089:+	11:65208314:+
81389	AL450405.1	RPL29	3	6:118320274:+	3:52029466:-
81389	SOD2	AC005332.1	7	6:160101991:-	17:66243500:-
81390	AC022144.1	NEAT1	3	19:39230112:+	11:65208337:+
81390	AL513497.1	PVRIG	4	1:28871710:+	7:99815864:+
81390	CNN3	MALAT1	4	1:95362938:-	11:65270105:+
81391	AC022144.1	NEAT1	4	19:39230088:+	11:65208326:+
81391	AL450405.1	RPL29	8	6:118320274:+	3:52029466:-
81391	BBS5	TRB@	4	2:170361729:+	22:38806453:+
81391	RPL29	AL450405.1	3	3:52028108:-	6:118320149:+
81391	RPL29	AL450405.1	3	3:52027893:-	6:118320495:+
81397	AL139099.5	PRB2	3	14:50053437:+	12:11546053:-
81397	AL445305.1	AP3S1	6	1:214656362:-	5:115238606:+
81397	PRB2	AL139099.5	3	12:11546279:-	14:50053456:+
81397	PRB2	AL139099.5	3	12:11546655:-	14:50053452:+
81397	PRB2	AL139099.5	3	12:11546775:-	14:50053452:+
81397	PRB2	AL139099.5	3	12:11546034:-	14:50053452:+
81397	RPL13	AC073621.1	5	16:89628134:+	17:17286929:-
81560	AC022144.1	NEAT1	3	19:39230089:+	11:65208314:+
81560	AC091807.1	IMPDH1	3	X:40218582:-	7:128034549:-
81560	AC091807.1	AC021305.1	6	X:40218567:-	8:124414219:+
81560	AL161431.1	TIMM50	3	13:109923120:+	19:39984134:+
81560	IMPDH1	AC091807.1	3	7:128036987:-	X:40219071:-
81560	IMPDH1	AC091807.1	3	7:128043764:-	X:40219839:-
81560	MCF2L	NEAT1	4	13:113694724:+	11:65211734:+
81560	NEAT1	GPRC5A	3	11:65211787:+	12:13070188:+
85191	AC022144.1	NEAT1	5	19:39230088:+	11:65208317:+
85192	IGF2BP3	AL023775.1	3	7:23353166:-	6:167114819:-
85194	AC022144.1	NEAT1	4	19:39230088:+	11:65208326:+
85195	AL031282.2	CDK11B	4	1:1635263:-	1:1571843:-
85195	AL132656.2	NUTM2B-AS1	4	10:81426614:+	10:81565878:-
85195	AL450405.1	RPL29	4	6:118320274:+	3:52029466:-
85195	ANXA1	NEAT1	4	9:75783796:+	11:65211742:+

85195	ARHGEF39	AC008993.1	3	9:35675023:-	19:68663:+
85195	IGK@	RBM47	3	2:90072638:-	4:40425672:-
85195	NUTM2B-AS1	AL132656.2	4	10:81568519:-	10:81429244:+
85195	NUTM2B-AS1	AL132656.2	4	10:81568515:-	10:81429244:+
85195	RPL29	AL450405.1	3	3:52028108:-	6:118320149:+

Table 4-9. RNA gene fusion calls made with FusionCatcher.

Study ID	Grade	Fusion	Site
MEC_006	Intermediate	Positive	Buccal
MEC_012	Intermediate	Positive	Submandibular
MEC_017	Intermediate	Positive	Parotid
MEC_038	Low	Positive	Parotid
MEC_123	Low	Positive	Soft Palate
MEC_175	Low	Negative	Parotid
MEC_10N	Normal	Negative	Parotid

Table 4-10. Tumor Localization of samples used in spatial RNAseq

Gene	# CRTC1-MAML2+ Tumors with Sig. Assoc.	Gene	# CRTC1-MAML2+ Tumors with Sig. Assoc.	Gene	# CRTC1-MAML2+ Tumors with Sig. Assoc.
BSG	6	EIF4H	4	GNPTG	3
CD55	6	ERGIC3	4	GPAA1	3
CD9	6	ERP29	4	GPRC5C	3
COX6A1	6	ESYT2	4	GTF2I	3
COX7C	6	FAM3C	4	GUK1	3
CTSD	6	FKBP8	4	HDLBP	3
EEF1G	6	FTH1	4	HEXA	3
EEF2	6	GNAS	4	HMGB1	3
ENO1	6	GPX4	4	HNRNPC	3
GLUL	6	GSTK1	4	HSPA1B	3
ITM2B	6	HINT1	4	HSPA9	3
KRT7	6	HNRNPA2B1	4	HSPE1	3
LGALS3BP	6	HNRNPA3	4	HTRA1	3
PABPC1	6	HSPA5	4	IFITM2	3
POLR2L	6	HSPA8	4	IFITM3	3
PPDPF	6	ID1	4	IGFBP4	3
PSAP	6	ID3	4	IGFBP7	3
S100A6	6	KIAA1522	4	ILF3	3
SAT1	6	LAMTOR4	4	ITGB4	3
SPINT2	6	LAPTM4B	4	ITGB8	3
TAGLN2	6	MAFK	4	JTB	3
TRIM29	6	MORF4L1	4	JUP	3
UBA52	6	MTCH1	4	KLF5	3
UBC	6	MXRA7	4	KRT5	3
UQCRB	6	MYH14	4	KSR1	3
A2M	5	NCL	4	LAMB3	3
ADIRF	5	NDUFA1	4	LAMP2	3
ASAH1	5	NDUFA13	4	LENG8	3
ATP1B1	5	NDUFB7	4	LMO4	3
BAG1	5	NDUFS6	4	LRP10	3
BEX3	5	NFE2L1	4	LSR	3
C19orf33	5	NFIC	4	LTBP3	3
CALR	5	NME2	4	LTBP4	3
CAPN1	5	NONO	4	LUC7L3	3
CD44	5	NPC2	4	METRN	3
CD63	5	NTRK2	4	MGAT4B	3
CD81	5	NUCKS1	4	MRC2	3
CHCHD10	5	NUPR1	4	NARS	3
CIRBP	5	OAZ1	4	NCKAP1	3
COX5A	5	P4HB	4	NCSTN	3
COX5B	5	PBXIP1	4	NDUFB10	3
COX6C	5	PCBP2	4	NDUFB2	3
COX7A2	5	PERP	4	NDUFV1	3
CTNNB1	5	PFN1	4	NFIB	3
DDR1	5	PHPT1	4	NSMF	3
DDX17	5	PIM3	4	NUCB1	3
DDX5	5	PLEKHH1	4	OBSL1	3
DLG5	5	PLXNB2	4	OCIAD1	3
EIF1	5	PPIB	4	PATJ	3
EIF4B	5	PRDX5	4	PCBP1	3
EMC10	5	PROM2	4	PDCD6	3
FAM210B	5	RBM3	4	PDHA1	3
FXYD3	5	RHOA	4	PDIA6	3
GABARAP	5	S100A10	4	PFKP	3
GPI	5	SCNN1A	4	PILRB	3
GSN	5	SIVA1	4	PLEC	3
GSTP1	5	SNRNP70	4	PNRC1	3
HNRNPDL	5	SOD1	4	PON3	3
HSP90AA1	5	SRRM2	4	PPP2CB	3
HSP90AB1	5	SRSF3	4	PRRC2B	3
KRT18	5	SRSF5	4	PTK2	3
KRT8	5	STAT3	4	PTPRF	3
LAMA5	5	TACSTD2	4	RAMP1	3
LAMP1	5	TAF10	4	RASL10B	3
LRPAP1	5	TCN1	4	RBX1	3

MIF	5		TESC	4		RCN2	3
MSLN	5		TLE2	4		RNASEK	3
MYH9	5		TMED2	4		RNF145	3
NDUFS5	5		TSHZ2	4		RNF187	3
NDUFS8	5		TUFM	4		ROMO1	3
NFE2L2	5		UBL5	4		RPN2	3
NR4A1	5		UQCR11	4		RTN4	3
OS9	5		UQCRQ	4		RUNX1	3
PARK7	5		USF2	4		S100A2	3
PEBP1	5		VAMP8	4		SCPEP1	3
PPP1CB	5		VDAC1	4		SDF4	3
PYGB	5		VEGFA	4		SDHA	3
RACK1	5		WDR1	4		SEL1L3	3
RGCC	5		ABHD2	3		SERBP1	3
RHOB	5		ADGRG1	3		SERINC2	3
RSRP1	5		AGRN	3		SLC25A37	3
S100A11	5		ALDH3B1	3		SLC38A1	3
SERF2	5		ANAPC16	3		SMAD2	3
SET	5		ANXA2	3		SMIM4	3
SLC25A3	5		ARHGDIA	3		SON	3
SQSTM1	5		ARPC3	3		SPINT1	3
ST3GAL4	5		ARPP19	3		SRSF11	3
SYNE2	5		ATP6AP1	3		SSR4	3
TNS4	5		ATRAID	3		STRAP	3
TOB1	5		BCAM	3		STT3B	3
TOMM7	5		BCAP31	3		SUMF2	3
UBB	5		BRK1	3		SYNGR2	3
WSB1	5		C4orf3	3		TAX1BP1	3
ACADVL	4		CCNL2	3		TFCP2L1	3
ACTB	4		CHMP1B	3		TFPI2	3
ANXA11	4		CHMP3	3		TM9SF3	3
ANXA5	4		CHPF	3		TMBIM6	3
AP2M1	4		CKB	3		TMED10	3
APLP2	4		CLTB	3		TMED3	3
APP	4		CMPK1	3		TMEM259	3
ARF1	4		CNBP	3		TMEM59	3
ARF4	4		COL18A1	3		TMEM59L	3
ARL6IP4	4		COPE	3		TMEM9	3
ATP1A1	4		COPG1	3		TMSB4X	3
ATP2A2	4		COX8A	3		TPM1	3
AZIN1	4		CRLF1	3		TTC3	3
B2M	4		CST3	3		TUBA4A	3
BCKDHB	4		CSTB	3		TXN	3
BCR	4		CTNNA1	3		TXNL4A	3
C4orf48	4		CTSA	3		UBE2V1	3
CANX	4		CTSZ	3		UNC93B1	3
CAPNS1	4		CUTA	3		UQCRC1	3
CAST	4		CYB5A	3		VDAC2	3
CCNI	4		CYB5R3	3		VIM	3
CD151	4		DCTN1	3		XBP1	3
CDH1	4		DDX39B	3		XRCC5	3
CHCHD2	4		DHCR24	3		YIPF3	3
CLDN7	4		DSP	3		YPEL5	3
CLSTN1	4		DST	3			
CLU	4		DUSP23	3			
COX4I1	4		ECHS1	3			
COX6B1	4		EDF1	3			
CPE	4		EEF1B2	3			
CRIM1	4		EPB41L1	3			
CTNND1	4		ERBB2	3			
CYCS	4		EZR	3			
DUSP4	4		FAM107B	3			
EHF	4		FBLN2	3			
EID1	4		FBXO7	3			
EIF3K	4		FOS	3			
EIF4A1	4		GABARAPL1	3			
EIF4A2	4		GHITM	3			
EIF4G2	4		GNB1	3			

Table 4-11. Genes significantly associated with CRTC1-MAML2 activity signature in fusion-positive tumors.

Genes with a geometric mean R > 0.03 in three or more fusion-positive tumors are reported.

Chapter 5 Summary and Perspectives

5.1 Summary of experimental work

In chapter 1, I discuss the role of high-risk HPV genes in HNSCC gene regulation. This review of the literature describes the way that introducing a handful of oncogenes can affect gene regulation in head and neck cancer, and how this affects disease progression. Throughout the remaining chapters, I study how the introduction of a few tumorigenic structural variants can similarly affect tumorigenesis in other head and neck cancers. In chapter 2, I determine gene regulation differences between two specific head and neck cancers, MEC and HCCC, by RNA sequencing. I identify an *EWSR-ATF1*-specific gene signature, differentially expressed between MEC and HCCC, several genes of which influence an HCCC-specific etiology. In chapter 3, I characterize translocations in MEC cell lines by long read sequencing, defining the CRTC1-MAML2 fusion breakpoint, identifying intermediate breakpoints affecting CRTC1 and MAML2, and discovering novel putative driver mutations in the *TERT* promoter region. These findings elucidate how fusion mutations form in MEC tumors, and how *TERT* mutations drive tumor growth. In chapter 4, I describe how CRTC1-MAML2 affects transcription in MEC primary tumors. I first define gene signatures associated with multiple clinical variables, including *CRTC1-MAML2* fusion status, then use these signatures to uncover cancer phenotypes in MEC tumors, including immune infiltration patterns. I then measure spatially resolved RNA expression patterns, finding a gene signature associated with CRTC1-MAML2-mediated gene

regulation. In this chapter, I will review the literature surrounding MEC tumor genetics and etiology, describing these findings in the context of previous work in the field.

5.2 Perspectives: The Molecular Biology of MEC Salivary Gland Tumors

5.2.1 Introduction

Mucoepidermoid carcinoma (MEC) is one of the most common salivary malignancies, comprising 30%-40% of all salivary gland tumors, and 2-7% of all head and neck cancers (120, 216). The majority of these tumors occur in the major salivary glands, with 50-80% in the parotid gland (184, 217). The rate of patient survival is highly dependent on tumor stage and grade. Five-year disease-free survival of patients with low-grade, intermediate-grade, and high-grade MEC tumors is 92-100%, 90-100%, and 37-67%, respectively (121, 218, 219). Meanwhile, disease-free survival of ranges from 91.7% for T1 stage patients to 12.5% for T4 patients (121). Of note, MEC does not share risk factors common in other oral cancers, including HPV infection and smoking (220, 221), and instead is most associated with exposure to environmental carcinogens, such as radiation exposure (222, 223). The standard of care for MEC patients is surgical resection, followed by radiotherapy in high grade tumors or tumors with post-resection positive margins, and no targeted therapies are in widespread use (224). Therefore, there is active need for the development of targeted therapies for MEC patients, especially for those with high-grade tumors.

5.2.2 MEC Intra-Tumor Heterogeneity

Mucoepidermoid carcinoma is histologically characterized by the presence of mucosal, epidermoid, and intermediate cells (225, 226). Several studies have shown cell type-specific expression of cancer-relevant proteins in each cell type, though the role of heterogeneous cell

types in MEC etiology is not currently known (227-229). All three cell types are thought to differentiate from cancer stem cells, though the specific mechanism of differentiation has not been described. It has been postulated that intermediate cells are a progenitor cell, differentiating into mucosal and epidermoid cells, though evidence for this mechanism is sparce (230). Cancer stem cells have been detected in MEC tumors and tumor-derived cell lines, marked with a $CD44^{high}ALDH^{high}$ phenotype and increased tumorigenic potential (231). CD44 expression is associated with high-grade tumors, characterized by aggressive growth, and increased ALDH protein levels are associated with tumor stage (232-234). These results support a cancer stem cell hypothesis, that a small proportion of stem cells maintain tumor growth and survival through self-renewal and differentiation (231). This stem cell tumorigenic activity may be regulated by mTOR signaling and epigenetic regulation (235, 236). Taken together, these data suggest that MEC growth is mediated by CSCs that differentiate into mucosal, epidermoid, and intermediate cells, leading to a tumor comprised of three histologically and molecularly different cell types.

5.2.3 Genetics of MEC tumors

Somatic mutation formation is a hallmark characteristic of cancer (237). In many cancers, driver mutations induce the development of cancer phenotypes, such as growth and metastasis. Therefore, identifying genes and pathways that reproducibly contain somatic mutations plays a critical role in determining tumor etiology.

The CRTC1-MAML2 Fusion Gene

In MEC, the most detected genetic alteration is a translocation between chromosome 11 and chromosome 19, resulting in formation of the CRTC1-MAML2 fusion protein (74, 238). The CRTC1-MAML2 fusion is found in 38-88% of all MEC tumors (124, 127, 129, 239-242).

Fusion-positive tumors are associated with a lower tumor grade and improved overall survival (127, 129, 239-242). The CRTC1-MAML2 fusion contains the first exon of *CRTC1*, containing a CREB binding domain, and exons 2-5 of *MAML2*, containing a transcription activation domain (71, 74). While wild type CRTC1 and MAML2 are both transcription factors in their own right, the CRTC1-MAML2 fusion gains transcription regulation function unique from both constituent genes. Initial studies demonstrate that transduction with CRTC1-MAML2, but not CRTC1 or MAML2, is sufficient to induce neoplastic transformation in an immortalized cell line (70, 74). This transformation is due to transcriptional regulation of several cancer-associated pathways. CRTC1-MAML2 regulates transcription in MEC tumors by several mechanisms. CRTC1-MAML2 directly activates CREB via the CRTC1 CREB binding domain (74, 132). Unlike CRTC1, CRTC1-MAML2 is localized to the nucleus, and the fusion constitutively activates CREB. In addition, CRTC1-MAML2-mediated transformation is decreased by CREB inhibition, suggesting that CRTC1-MAML2 plays a role in MEC tumorigenesis.

CRTC1-MAML2 regulates transcription via interactions with other transcription factors, in addition to CREB. Amelio *et al*. first discovered activation of MYC regulated genes by CRTC1-MAML2, but not CRTC1 or MAML2 (73). They further determine that CRTC1-MAML2 directly interacts with MYC, activating MYC-regulated genes. Importantly, they also demonstrate that the previously described CRTC1-MAML2-mediated transformed colony formation is abrogated by transduction with MYC with a loss-of-function mutation.

Chen J, *et al* is the first study to measure differential gene expression in fusion-positive MEC derived cell lines (71). By comparing RNA expression of fusion-positive to fusion-knockdown samples, they identify a CRTC1-MAML2-associated gene signature. The genes inhibited by CRTC1-MAML2 knockdown are activated by expected transcription factors, such

as CREB and MYC, but also previously unassociated transcription factors, such as NFkB and HIF1A. These genes are also involved in cancer phenotypes; this gene signature is enriched in pathways such as cell growth, cell survival, and cell-to-cell signaling. *C6orf176*, encoding LINC00473, is the gene most highly repressed by CRTC1-MAML2 knockdown. In a following study, the same group determines that LINC00473 is necessary for CRTC1-MAML2-mediated growth (175). CRTC1-MAML2 directly induces LINC00473 expression in a CREB-dependent manner. In addition, LINC00473 knockdown leads to reduced fusion-positive cell line and xenograft growth, and increased apoptosis, as well as downregulation of genes associated with cell survival and growth. LINC00473 directly interacts with the NONO protein, a coactivator necessary for CREB activation in noncancerous cells. Therefore, the authors propose that LINC00473 enhances CRTC1-MAML2-mediated CREB activation by interacting with the CRE site.

By directly performing RNA sequencing on MEC fusion-positive and fusion negative tumors, Musicant *et al* reveal a fusion-specific gene expression signature, which is enriched for IGF-1 signaling associated genes (72). However, the *IGF1* promoter lacks a CRE site. Instead, the authors demonstrate that CRTC1-MAML2 instead induces expression of the splice variant PGC-1a4, which coactivates PPARγ, which directly induces IGF-1 expression.

Through the transcriptional regulation of multiple pathways, CRTC2-MAML2 is the primary driver mutation in fusion-positive MEC tumors, as demonstrated by Chen *et al* (243). The authors first that CRTC1-MAML2 knockdown is sufficient to fully stop MEC xenograft growth. Furthermore, expression of a CRTC1-MAML2 fusion transgene in healthy mouse salivary glands is sufficient to generate murine MEC tumors, which are phenotypically similar to human MEC tumors, containing mucosal, epidermoid, and intermediate cells. They then treat

CRTC1-MAML2-expressing mice with CDK4/6 and EGFR inhibitor combination therapy, which leads to decreased tumor size, suggesting interesting future directions for both this combination therapy and this mouse model.

In chapter 4, I perform RNA sequencing to identify a MAML2-associated gene signature. This one of only two studies available that compare gene expression by MAML2 fusion status in MEC primary tumors, and the largest cohort to date. I find that fusion-associated genes are enriched for gene sets involved in cellular respiration, including oxidative phosphorylation, electron transport chain, and aerobic respiration. Of note, in mouse skeletal muscle, oxidative phosphorylation gene sets are directly regulated by PGC-1a, a downstream regulator of CRTC1-MAML2-associated transcription in the findings of Musicant *et al (72, 244).* In addition, we find that CRTC1-MAML2 positive tumors are significantly CD8+ T cell-depleted, compared to fusion-negative tumors. I also use spatial RNA sequencing to show that expression of this fusion-associated signature varies within each tumor, suggesting an interaction between CRTC1-MAML2 and the MEC tumor microenvironment, which will be important for future research.

Taken together, these data demonstrate that CRTC1-MAML2 is sufficient to cause tumorigenesis and neoplastic transformation. This fusion directly regulates large-scale changes in gene transcription within the MEC tumor, through coactivating with the transcription factors CREB, MYC, and PPARγ.

Other Somatic Mutations in MEC Tumors

The t(11;19) translocation, leading to the CRTC1-MAML2 fusion gene, is the best studied driver mutation in MEC. However, other somatic mutations have also been identified. These mutations may contribute to cancer phenotypes in CRC1-MAML2 fusion-positive MEC

tumors or contribute to pathogenesis in fusion-negative tumors. Here, I review those limited genetic studies.

While CRTC1-MAML2 is the most prevalent MEC-associated gene fusion, CRTC3-MAML2 fusions have also been detected (98, 130, 245). These studies each demonstrate that CRTC1-MAML2 and CRTC3-MAML2 positive MEC tumors are histologically similar, but the fusions are mutually exclusive.

In addition to MAML2 fusions, other somatic mutations contribute to cancer etiology, in both fusion positive and fusion negative tumors. A variety of high-throughput methods have been used to successfully detect single nucleotide variant (SNV), copy number alteration (CNA), and structural variation mutations throughout the genome. For example, using a 261,563 SNP genotyping microarray (246), one study uses the relative signal of these SNPs to identify copy number variants. The most reproducible CNA mutations overlapped with *MAPT* (50% patients with copy number gain/40% copy number loss), *APEX1* (35% gain/45% loss) *POTEB* (45%/25%), *OR4N4* (40%/25%), and *TP53TG3* (5%/25%). Interestingly, CRTC1-MAML2 positive tumors have lower rate of copy number alterations, compared to fusion-negative tumors (246-248). Another study used an array custom genomic hybridization (aCGH), a microarray that can detect copy number imbalances in 99,000 human sequences (249). They detect a chr9p21.3 (*CDKN2A*) deletion in 4/15 patients, which is only found in patients with worse overall survival. This *CDKN2A* loss was reproduced in a different copy number assay, which also described copy number losses in 18q12.2-qter (*DCC, SMAD4, GALR1*) and copy number gains in 8q24.3 (*MAFA*) and 8q11.1-q12.2 (*LYN, MOS, PLAG1*)(247).

Following in the footsteps of these array-based study, a few published studies have used next generation sequencing to identify somatic mutations. Unlike microarrays, whole exome

sequencing and whole genome sequencing (WES and WGS) allow for direct sequencing of the entire exome or genome, facilitating detection of rare or novel variants previously not included in the preset probes of a microarray. The first WES study performed on MEC tumors sequences 18 tumors, plus matched normal controls, identifying a mean and standard deviation of 43 ± 63.5 mutations per tumor (124). Specifically, the authors identify somatic SNVs in *TP53, IRAK1, MAP3K9, ITGAL, ERGG4, OTOGL, KMT2C* and *OBSCN* (2+ tumors/18 for all genes listed). This study also bioinformatically analyzes copy number variations in these 18 tumors, 32/33 of which were tumor specific. A second study uses a more targeted method, comprehensive genomic screening (CGS). CGS is a capture sequencing-based technology and is an FDA-approved test for identifying cancer driver mutations (250). In this study, CGS was used to detect SNVs, but not copy number alterations or translocation events, in a pan-salivary gland cancer cohort, including 57 MEC tumors (251). They detect a tumor mutational burden of 4 mutations per tumor, most commonly affecting *CDKN2A*, *TP53*, *CDKN2B*, *BAP1*, *PIK3CA*, *ERBB2*, *BRCA2*, *HRAS*, and *FGFR1* (251, 252). Other targeted sequencing on pan-salivary gland cancer cohorts have too small of a MEC sample size (N=5, 8 respectively) to identify genes that are mutated in more than one patient, except for *TP53* (253, 254).

The most exhaustive study of gene mutations in MEC used whole genome sequencing (WGS) to sequence 18 primary MEC tumors, as part of a common salivary gland cancer cohort. Unlike WES and capture sequencing, WGS allows for detection of intronic and intergenic variants. First, the authors identify missense SNVs in *BAP1, LRFN1, CIC, MUC16, ATN1, CARD154, CCDC58,* and *TTN*. In addition to these findings, the authors apply a variety of bioinformatics tools to understand more genetic information from these tumors. For example, they use a mutational signature analysis to determine the relative contribution of thirty well-

defined patterns of single nucleotide changes. Several of these signatures are associated with known mutational processes, suggesting potential mutagenic mechanisms for MEC tumor SNV formation. Specifically, MEC tumors have broad enrichment for multiple signatures associated with DNA repair pathways: most notably signatures 3 (homologous recombination), 6, 15, 20 (mismatch repair), and 24 (nucleotide excision repair). In chapter 3, a similar method reproduces signature 3 enrichment, while also identifying enrichment of signatures 5 and 12, associated with unknown etiologies. In addition, the CRTC1-MAML2 fusion could be detected in 50% of patients, using bioinformatics-based analysis of WGS data, rather than the orthogonal and time-consuming methods of previous studies, such as FISH. This method is then used to identify additional fusions, including HFM1-RYR2, CACNA1B-NBPF10, EWSR1-ATF1, and FSIP1-BAZ2A. However, these fusions were not verified by secondary methods, such as rtPCR, and the way they affect cancer phenotypes is not currently known. With additional repetition and validation, direct tumor sequencing coupled with bioinformatics analysis strategies has the potential to identify novel driver mutations.

In chapter 3, I perform two sequencing methods to identify MEC tumor mutation. By long read sequencing, we could identify fully novel *TERT* promoter translocations, and reproduce the CRTC1-MAML2 fusion. In addition, by WES of primary tumors, we identify additional *TERT* copy number alterations. Interestingly, many of the SNVs, CNAs, and fusions we identify, affecting genes including, *ATM*, *CTBP2*, *MAML1*, *MUC12*, *MUC3a*, and *MUC5B*, have not been previously described in the literature, suggesting that further studies with increased sample sizes are needed to describe the full range and prevalence of somatic mutations in MEC tumors.

The remaining type of genetic research directly genotype genes with well-studied effects on cancer phenotypes. For example, using Sanger sequencing and chromogenic *in-situ* hybridization (CISH), Nakano *et al.* detect *MAML2* fusions (15/31 tumors), *HER2* amplification (4/31 tumors), and *EGFR* copy number increase (4/31 tumors) in MEC tumors (255). HER2 protein overexpression was highly correlated with gene amplification; however, EGFR protein overexpression occurs in the epidermoid and intermediate cells of 18 of 31 tumors, often unrelated to copy number status. Another strategy uses a custom 123-SNV microarray panel targeting commonly mutated genes to detect a *KRAS* A59T mutation in 1 of 30 tumors, but no other variants (256). The results of this study emphasize the increased power of unbiased mutation detection assays, such as WES. However, these low-throughput techniques successfully identify mutations in genes with available targeted therapies, with the end goal of identifying potential therapeutic targets. Multiple clinical trials have tested the efficacy of therapeutics that inhibit HER2 or EGFR in general salivary gland cancer cohorts, though these studies are limited due to low MEC patient enrollment (reviewed (224)). Therefore, these genotyping studies inform future experiments to identify treatments for MEC patients.

In summary, unbiased and targeted studies have identified many SNV, copy number, and translocation mutations in MEC tumors. While some of these mutations, such as CRTC1-MAML2, CRTC3-MAML2, and *TP53* SNVs, are reproduced in many samples, several tumor-specific variants have also been described. Further study is required determine the mutation rate of less commonly mutated genes, and to determine how many of these mutations affect tumor progression.

5.2.4 Improved genetics and bioinformatics methods advance somatic mutation detection

The development of improved genetics and bioinformatics technologies has been integral to determining the etiology of MEC tumors. The genome translocation event that leads to the CRTC1-MAML2 was first described by tumor karyotyping in 1994 (238). This study describes a t(11;19)(q14-21;p12) rearrangement, but cannot determine the genes or cancer phenotypes affected by the translocation. The CRTC1-MAML2 fusion was then identified from this translocation, using FISH visualization, followed by BAC cloning and Sanger sequencing (70). Using this method, Tonon *et al.* could identify the gene fusion, directly sequence it, and characterize the wild type MAML2 gene for the first time. Future studies demonstrated that this fusion affects tumorigenesis via CREB activation. Detecting a CREB-mediated effect was facilitated via RNA expression microarray, which was used to demonstrate an enrichment of CREB-regulated genes in fusion-positive cell lines, which could then be followed up by CREB inhibition studies (74). The increased of genotyping microarray, then next-generation sequencing facilitated the further detection of CRTC1-MAML2-independent mutations and pathways involved in MEC etiology, in both fusion-positive and negative tumors (124, 246). However, the CRTC1-MAML2 fusion is not easily detectable by whole exome sequencing, requiring separate detection by FISH assay, similar to methods performed in 2003 (124). Improving bioinformatics techniques allow for more accurate and direct detection of structural variants. For example, one recent study directly sequences the fusion junction site via whole-genome sequencing, then uses RNA splice site identification methods to identify DNA translocation junctions (257). New developments in long read-sequencing methods will improve this method further, allowing for more sensitive detection of structural variants, including gene fusions (141, 258). Using this method, our group has been able to detect novel intermediate translocations in CRTC1-MAML2-positive MEC tumor-derived cell lines, as well as a separate *TERT* translocation with a putative

role in tumor growth (chapter 3). In addition, recent sequencing methods can be used to better elucidate tumor phenotypes. For example, advances in spatial RNA sequencing and single-cell RNA sequencing will elucidate cancer phenotypes throughout the tumor microenvironment, as discussed in more detail in chapter 4. As genetics and bioinformatics methods continue to improve, improved detection of somatic driver mutations, more accurately characterizing the genetics and molecular pathology of MEC etiology.

5.2.5 Next steps

In this thesis, I investigate somatic mutations in MEC tumors, and the effects on transcriptional regulation and disease etiology. In chapter 2, I discover a gene expression signature differentiating between MEC and HCCC tumors. Further study is necessary to determine if the genes that we highlight could be used as a diagnostic biomarker, differentiating between the visually similar MEC and HCCC tumors. However, this research is most interesting because it identifies differences in cancer phenotypes and molecular etiology between HCCC and MEC, suggesting the need for HCCC and MEC specific treatment strategies.

In the third chapter, I use new long-read sequencing technologies to identify the pattern of genomic rearrangement in a MEC cell line. Using this method, we identify translocation events, including a novel *TERT* rearrangement. While we determine that telomerase activity is necessary for cell growth in a TERT-rearranged cell line, these findings suggest that the role of telomerase in MEC etiology is an interesting topic for future study. In addition, other salivary gland cancers, including adenoid cystic carcinoma, mammary analogue secretory carcinoma, and HCCC, are characterized by putative driver translocation mutations (259). Given our success, this Nanopore sequencing-based method could be similarly utilized to identify novel genomic rearrangement in other salivary malignancies.

In chapter 4, I describe gene regulation patterns associated with CRTC1-MAML2 expression, using both bulk and spatial RNA sequencing. Through these methods, we identified several genes that might play a role in MEC pathogenesis and growth. However, further studies are necessary to determine how these genes interact with each other.

Each of these future studies seeks to better understand the formation and growth of MEC tumors, building on the findings of the work presented in this thesis, and described in previous studies. This improved understanding is necessary to identify potential targeted therapies that will target MEC growth, with the goal of improving outcomes for MEC patients.

Bibliography

1. Walboomers JM, Jacobs MV, Manos MM, Bosch FX, Kummer JA, Shah KV, Snijders PJ, Peto J, Meijer CJ, Munoz N. Human papillomavirus is a necessary cause of invasive cervical cancer worldwide. J Pathol. 1999;189(1):12-9. doi: 10.1002/(SICI)1096-9896(199909)189:1<12::AID-PATH431>3.0.CO;2-F. PubMed PMID: 10451482.

2. Michmerhuizen NL, Birkeland AC, Bradford CR, Brenner JC. Genetic determinants in head and neck squamous cell carcinoma and their influence on global personalized medicine. Genes Cancer. 2016;7(5-6):182-200. doi: 10.18632/genesandcancer.110. PubMed PMID: 27551333; PMCID: PMC4979591.

3. Gingerich MA, Smith JD, Michmerhuizen NL, Ludwig M, Devenport S, Matovina C, Brenner C, Chinn SB. Comprehensive review of genetic factors contributing to head and neck squamous cell carcinoma development in low-risk, nontraditional patients. Head Neck. 2018;40(5):943-54. doi: 10.1002/hed.25057. PubMed PMID: 29427520; PMCID: PMC5912962.

4. Tillman BN, Yanik M, Birkeland AC, Liu CJ, Hovelson DH, Cani AK, Palanisamy N, Carskadon S, Carey TE, Bradford CR, Tomlins SA, McHugh JB, Spector ME, Brenner J. Fibroblast growth factor family aberrations as a putative driver of head and neck squamous cell carcinoma in an epidemiologically low-risk patient as defined by targeted sequencing. Head Neck. 2016;38 Suppl 1:E1646-52. doi: 10.1002/hed.24292. PubMed PMID: 26849095; PMCID: PMC4844767.

5. Marur S, D'Souza G, Westra WH, Forastiere AA. HPV-associated head and neck cancer: a virus-related cancer epidemic. Lancet Oncol. 2010;11(8):781-9. doi: 10.1016/S1470-2045(10)70017-6. PubMed PMID: 20451455; PMCID: PMC5242182.

6. D'Souza G, Kreimer AR, Viscidi R, Pawlita M, Fakhry C, Koch WM, Westra WH, Gillison ML. Case-control study of human papillomavirus and oropharyngeal cancer. N Engl J Med. 2007;356(19):1944-56. doi: 10.1056/NEJMoa065497. PubMed PMID: 17494927.

7. Braaten KP, Laufer MR. Human Papillomavirus (HPV), HPV-Related Disease, and the HPV Vaccine. Rev Obstet Gynecol. 2008;1(1):2-10. PubMed PMID: 18701931; PMCID: PMC2492590.

8. Nulton TJ, Olex AL, Dozmorov M, Morgan IM, Windle B. Analysis of The Cancer Genome Atlas sequencing data reveals novel properties of the human papillomavirus 16 genome in head and neck squamous cell carcinoma. Oncotarget. 2017;8(11):17684-99. doi: 10.18632/oncotarget.15179. PubMed PMID: 28187443; PMCID: PMC5392278.

9. Louvanto K, Rautava J, Willberg J, Wideman L, Syrjanen K, Grenman S, Syrjanen S. Genotype-specific incidence and clearance of human papillomavirus in oral mucosa of women: a six-year follow-up study. PLoS One. 2013;8(1):e53413. doi: 10.1371/journal.pone.0053413. PubMed PMID: 23301068; PMCID: PMC3536668.

10. Ou D, Adam J, Garberis I, Blanchard P, Nguyen F, Levy A, Casiraghi O, Gorphe P, Breuskin I, Janot F, Temam S, Scoazec JY, Deutsch E, Tao Y. Influence of tumor-associated macrophages and HLA class I expression according to HPV status in head and neck cancer patients receiving chemo/bioradiotherapy. Radiother Oncol. 2019;130:89-96. doi: 10.1016/j.radonc.2018.08.013. PubMed PMID: 30172455.

11. Campo MS, Graham SV, Cortese MS, Ashrafi GH, Araibi EH, Dornan ES, Miners K, Nunes C, Man S. HPV-16 E5 down-regulates expression of surface HLA class I and reduces recognition by CD8 T cells. Virology. 2010;407(1):137-42. doi: 10.1016/j.virol.2010.07.044. PubMed PMID: 20813390.

12. Bottley G, Watherston OG, Hiew YL, Norrild B, Cook GP, Blair GE. High-risk human papillomavirus E7 expression reduces cell-surface MHC class I molecules and increases susceptibility to natural killer cells. Oncogene. 2008;27(12):1794-9. doi: 10.1038/sj.onc.1210798. PubMed PMID: 17828295.

13. Partlova S, Boucek J, Kloudova K, Lukesova E, Zabrodsky M, Grega M, Fucikova J, Truxova I, Tachezy R, Spisek R, Fialova A. Distinct patterns of intratumoral immune cell infiltrates in patients with HPV-associated compared to non-virally induced head and neck squamous cell carcinoma. Oncoimmunology. 2015;4(1):e965570. doi: 10.4161/21624011.2014.965570. PubMed PMID: 25949860; PMCID: PMC4368144.

14. Krishna S, Ulrich P, Wilson E, Parikh F, Narang P, Yang S, Read AK, Kim-Schulze S, Park JG, Posner M, Wilson Sayres MA, Sikora A, Anderson KS. Human papilloma virus specific immunogenicity and dysfunction of CD8+ T cells in head and neck cancer. Cancer Res. 2018. doi: 10.1158/0008-5472.CAN-18-0163. PubMed PMID: 30154146.

15. Russell S, Angell T, Lechner M, Liebertz D, Correa A, Sinha U, Kokot N, Epstein A. Immune cell infiltration patterns and survival in head and neck squamous cell carcinoma. Head Neck Oncol. 2013;5(3):24. PubMed PMID: 24723971; PMCID: PMC3979926.

16. Hoesli R, Birkeland AC, Rosko AJ, Issa M, Chow KL, Michmerhuizen NL, Mann JE, Chinn SB, Shuman AG, Prince ME, Wolf GT, Bradford CR, McHugh JB, Brenner JC, Spector ME. Proportion of CD4 and CD8 tumor infiltrating lymphocytes predicts survival in persistent/recurrent laryngeal squamous cell carcinoma. Oral Oncol. 2018;77:83-9. doi: 10.1016/j.oraloncology.2017.12.003. PubMed PMID: 29362129; PMCID: PMC5897103.

17. Mann JE, Smith JD, Birkeland AC, Bellile E, Swiecicki P, Mierzwa M, Chinn SB, Shuman AG, Malloy KM, Casper KA, McLean SA, Moyer JS, Wolf GT, Bradford CR, Prince ME, Carey TE, McHugh JB, Spector ME, Brenner JC. Analysis of tumor-infiltrating CD103 resident memory T-cell content in recurrent laryngeal squamous cell carcinoma. Cancer Immunol Immunother. 2019;68(2):213-20. doi: 10.1007/s00262-018-2256-3. PubMed PMID: 30361882; PMCID: PMC6375747.

18. Lechner A, Schlosser HA, Thelen M, Wennhold K, Rothschild SI, Gilles R, Quaas A, Siefer OG, Huebbers CU, Cukuroglu E, Goke J, Hillmer A, Gathof B, Meyer MF, Klussmann JP, Shimabukuro-Vornhagen A, Theurich S, Beutner D, von Bergwelt-Baildon M. Tumor-associated B cells and humoral immune response in head and neck squamous cell carcinoma. Oncoimmunology. 2019;8(3):1535293. doi: 10.1080/2162402X.2018.1535293. PubMed PMID: 30723574; PMCID: PMC6350680.

19. Adiko AC, Babdor J, Gutierrez-Martinez E, Guermonprez P, Saveanu L. Intracellular Transport Routes for MHC I and Their Relevance for Antigen Cross-Presentation. Front Immunol. 2015;6:335. doi: 10.3389/fimmu.2015.00335. PubMed PMID: 26191062; PMCID: PMC4489332.

20. Doherty PC, Zinkernagel RM. Enhanced immunological surveillance in mice heterozygous at the H-2 gene complex. Nature. 1975;256(5512):50-2. PubMed PMID: 1079575.

21. Gruener M, Bravo IG, Momburg F, Alonso A, Tomakidi P. The E5 protein of the human papillomavirus type 16 down-regulates HLA-I surface expression in calnexin-expressing but not in calnexin-deficient cells. Virol J. 2007;4:116. doi: 10.1186/1743-422X-4-116. PubMed PMID: 17971213; PMCID: PMC2164959.

22. Hackl H, Charoentong P, Finotello F, Trajanoski Z. Computational genomics tools for dissecting tumour-immune cell interactions. Nat Rev Genet. 2016;17(8):441-58. doi: 10.1038/nrg.2016.67. PubMed PMID: 27376489.

23. Hutchison S, Pritchard AL. Identifying neoantigens for use in immunotherapy. Mamm Genome. 2018;29(11-12):714-30. doi: 10.1007/s00335-018-9771-6. PubMed PMID: 30167844; PMCID: PMC6267674.

24. Lee CH, Yelensky R, Jooss K, Chan TA. Update on Tumor Neoantigens and Their Utility: Why It Is Good to Be Different. Trends Immunol. 2018;39(7):536-48. doi: 10.1016/j.it.2018.04.005. PubMed PMID: 29751996; PMCID: PMC7954132.

25. Lundegaard C, Hoof I, Lund O, Nielsen M. State of the art and challenges in sequence based T-cell epitope prediction. Immunome Res. 2010;6 Suppl 2:S3. doi: 10.1186/1745-7580-6-S2-S3. PubMed PMID: 21067545; PMCID: PMC2981877.

26. Soria-Guerra RE, Nieto-Gomez R, Govea-Alonso DO, Rosales-Mendoza S. An overview of bioinformatics tools for epitope prediction: implications on vaccine development. J Biomed Inform. 2015;53:405-14. doi: 10.1016/j.jbi.2014.11.003. PubMed PMID: 25464113.

27. Snyder A, Chan TA. Immunogenic peptide discovery in cancer genomes. Curr Opin Genet Dev. 2015;30:7-16. doi: 10.1016/j.gde.2014.12.003. PubMed PMID: 25588790; PMCID: PMC6657809.

28. Andreatta M, Nielsen M. Gapped sequence alignment using artificial neural networks: application to the MHC class I system. Bioinformatics. 2016;32(4):511-7. doi: 10.1093/bioinformatics/btv639. PubMed PMID: 26515819; PMCID: PMC6402319.

29. Nielsen M, Lundegaard C, Lund O, Kesmir C. The role of the proteasome in generating cytotoxic T-cell epitopes: insights obtained from improved predictions of proteasomal cleavage. Immunogenetics. 2005;57(1-2):33-41. doi: 10.1007/s00251-005-0781-7. PubMed PMID: 15744535.

30. Bhasin M, Raghava GP. Pcleavage: an SVM based method for prediction of constitutive proteasome and immunoproteasome cleavage sites in antigenic sequences. Nucleic Acids Res. 2005;33(Web Server issue):W202-7. doi: 10.1093/nar/gki587. PubMed PMID: 15988831; PMCID: PMC1160263.

31. Zhang GL, Petrovsky N, Kwoh CK, August JT, Brusic V. PRED(TAP): a system for prediction of peptide binding to the human transporter associated with antigen processing. Immunome Res. 2006;2:3. doi: 10.1186/1745-7580-2-3. PubMed PMID: 16719926; PMCID: PMC1524936.

32. Trolle T, Nielsen M. NetTepi: an integrated method for the prediction of T cell epitopes. Immunogenetics. 2014;66(7-8):449-56. doi: 10.1007/s00251-014-0779-0. PubMed PMID: 24863339.

33. Schenck RO, Lakatos E, Gatenbee C, Graham TA, Anderson ARA. NeoPredPipe: high-throughput neoantigen prediction and recognition potential pipeline. BMC Bioinformatics. 2019;20(1):264. doi: 10.1186/s12859-019-2876-4. PubMed PMID: 31117948; PMCID: PMC6532147.

34. Hundal J, Carreno BM, Petti AA, Linette GP, Griffith OL, Mardis ER, Griffith M. pVAC-Seq: A genome-guided in silico approach to identifying tumor neoantigens. Genome Med. 2016;8(1):11. doi: 10.1186/s13073-016-0264-5. PubMed PMID: 26825632; PMCID: PMC4733280.

35. Bais P, Namburi S, Gatti DM, Zhang X, Chuang JH. CloudNeo: a cloud pipeline for identifying patient-specific tumor neoantigens. Bioinformatics. 2017;33(19):3110-2. doi: 10.1093/bioinformatics/btx375. PubMed PMID: 28605406; PMCID: PMC5870764.

36. Tappeiner E, Finotello F, Charoentong P, Mayer C, Rieder D, Trajanoski Z. TIminer: NGS data mining pipeline for cancer immunology and immunotherapy. Bioinformatics. 2017;33(19):3140-1. doi: 10.1093/bioinformatics/btx377. PubMed PMID: 28633385; PMCID: PMC5870678.

37. Wang TY, Wang L, Alam SK, Hoeppner LH, Yang R. ScanNeo: identifying indel-derived neoantigens using RNA-Seq data. Bioinformatics. 2019;35(20):4159-61. doi: 10.1093/bioinformatics/btz193. PubMed PMID: 30887025.

38. Zhang J, Mardis ER, Maher CA. INTEGRATE-neo: a pipeline for personalized gene fusion neoantigen discovery. Bioinformatics. 2017;33(4):555-7. doi: 10.1093/bioinformatics/btw674. PubMed PMID: 27797777; PMCID: PMC5408800.

39. Chen YP, Wang YQ, Lv JW, Li YQ, Chua MLK, Le QT, Lee N, Colevas AD, Seiwert T, Hayes DN, Riaz N, Vermorken JB, O'Sullivan B, He QM, Yang XJ, Tang LL, Mao YP, Sun Y,

Liu N, Ma J. Identification and validation of novel microenvironment-based immune molecular subgroups of head and neck squamous cell carcinoma: implications for immunotherapy. Ann Oncol. 2019;30(1):68-75. doi: 10.1093/annonc/mdy470. PubMed PMID: 30407504.

40. Karosiene E, Lundegaard C, Lund O, Nielsen M. NetMHCcons: a consensus method for the major histocompatibility complex class I predictions. Immunogenetics. 2012;64(3):177-86. doi: 10.1007/s00251-011-0579-8. PubMed PMID: 22009319.

41. Ott PA, Hu Z, Keskin DB, Shukla SA, Sun J, Bozym DJ, Zhang W, Luoma A, Giobbie-Hurder A, Peter L, Chen C, Olive O, Carter TA, Li S, Lieb DJ, Eisenhaure T, Gjini E, Stevens J, Lane WJ, Javeri I, Nellaiappan K, Salazar AM, Daley H, Seaman M, Buchbinder EI, Yoon CH, Harden M, Lennon N, Gabriel S, Rodig SJ, Barouch DH, Aster JC, Getz G, Wucherpfennig K, Neuberg D, Ritz J, Lander ES, Fritsch EF, Hacohen N, Wu CJ. An immunogenic personal neoantigen vaccine for patients with melanoma. Nature. 2017;547(7662):217-21. doi: 10.1038/nature22991. PubMed PMID: 28678778; PMCID: PMC5577644.

42. Sahin U, Derhovanessian E, Miller M, Kloke BP, Simon P, Lower M, Bukur V, Tadmor AD, Luxemburger U, Schrors B, Omokoko T, Vormehr M, Albrecht C, Paruzynski A, Kuhn AN, Buck J, Heesch S, Schreeb KH, Muller F, Ortseifer I, Vogler I, Godehardt E, Attig S, Rae R, Breitkreuz A, Tolliver C, Suchan M, Martic G, Hohberger A, Sorn P, Diekmann J, Ciesla J, Waksmann O, Bruck AK, Witt M, Zillgen M, Rothermel A, Kasemann B, Langer D, Bolte S, Diken M, Kreiter S, Nemecek R, Gebhardt C, Grabbe S, Holler C, Utikal J, Huber C, Loquai C, Tureci O. Personalized RNA mutanome vaccines mobilize poly-specific therapeutic immunity against cancer. Nature. 2017;547(7662):222-6. doi: 10.1038/nature23003. PubMed PMID: 28678784.

43. Chu Y, Liu Q, Wei J, Liu B. Personalized cancer neoantigen vaccines come of age. Theranostics. 2018;8(15):4238-46. doi: 10.7150/thno.24387. PubMed PMID: 30128050; PMCID: PMC6096398.

44. Birkeland AC, Yanik M, Tillman BN, Scott MV, Foltin SK, Mann JE, Michmerhuizen NL, Ludwig ML, Sandelski MM, Komarck CM, Carey TE, Prince ME, Bradford CR, McHugh JB, Spector ME, Brenner JC. Identification of Targetable ERBB2 Aberrations in Head and Neck Squamous Cell Carcinoma. JAMA Otolaryngol Head Neck Surg. 2016;142(6):559-67. doi: 10.1001/jamaoto.2016.0335. PubMed PMID: 27077364; PMCID: PMC4911238.

45. Giefing M, Wierzbicka M, Szyfter K, Brenner JC, Braakhuis BJ, Brakenhoff RH, Bradford CR, Sorensen JA, Rinaldo A, Rodrigo JP, Takes RP, Ferlito A. Moving towards personalised therapy in head and neck squamous cell carcinoma through analysis of next generation sequencing data. Eur J Cancer. 2016;55:147-57. doi: 10.1016/j.ejca.2015.10.070. PubMed PMID: 26851381; PMCID: PMC4761501.

46. Lawrence MS, Stojanov P, Polak P, Kryukov GV, Cibulskis K, Sivachenko A, Carter SL, Stewart C, Mermel CH, Roberts SA, Kiezun A, Hammerman PS, McKenna A, Drier Y, Zou L, Ramos AH, Pugh TJ, Stransky N, Helman E, Kim J, Sougnez C, Ambrogio L, Nickerson E, Shefler E, Cortes ML, Auclair D, Saksena G, Voet D, Noble M, DiCara D, Lin P, Lichtenstein L, Heiman DI, Fennell T, Imielinski M, Hernandez B, Hodis E, Baca S, Dulak AM, Lohr J, Landau

DA, Wu CJ, Melendez-Zajgla J, Hidalgo-Miranda A, Koren A, McCarroll SA, Mora J, Crompton B, Onofrio R, Parkin M, Winckler W, Ardlie K, Gabriel SB, Roberts CWM, Biegel JA, Stegmaier K, Bass AJ, Garraway LA, Meyerson M, Golub TR, Gordenin DA, Sunyaev S, Lander ES, Getz G. Mutational heterogeneity in cancer and the search for new cancer-associated genes. Nature. 2013;499(7457):214-8. doi: 10.1038/nature12213. PubMed PMID: 23770567; PMCID: PMC3919509.

47. Rooney MS, Shukla SA, Wu CJ, Getz G, Hacohen N. Molecular and genetic properties of tumors associated with local immune cytolytic activity. Cell. 2015;160(1-2):48-61. doi: 10.1016/j.cell.2014.12.033. PubMed PMID: 25594174; PMCID: PMC4856474.

48. Ren L, Leisegang M, Deng B, Matsuda T, Kiyotani K, Kato T, Harada M, Park JH, Saloura V, Seiwert T, Vokes E, Agrawal N, Nakamura Y. Identification of neoantigen-specific T cells and their targets: implications for immunotherapy of head and neck squamous cell carcinoma. Oncoimmunology. 2019;8(4):e1568813. doi: 10.1080/2162402X.2019.1568813. PubMed PMID: 30906664; PMCID: PMC6422382.

49. Yang W, Lee KW, Srivastava RM, Kuo F, Krishna C, Chowell D, Makarov V, Hoen D, Dalin MG, Wexler L, Ghossein R, Katabi N, Nadeem Z, Cohen MA, Tian SK, Robine N, Arora K, Geiger H, Agius P, Bouvier N, Huberman K, Vanness K, Havel JJ, Sims JS, Samstein RM, Mandal R, Tepe J, Ganly I, Ho AL, Riaz N, Wong RJ, Shukla N, Chan TA, Morris LGT. Immunogenic neoantigens derived from gene fusions stimulate T cell responses. Nat Med. 2019;25(5):767-75. doi: 10.1038/s41591-019-0434-2. PubMed PMID: 31011208; PMCID: PMC6558662.

50. Davidson EJ, Davidson JA, Sterling JC, Baldwin PJ, Kitchener HC, Stern PL. Association between human leukocyte antigen polymorphism and human papillomavirus 16-positive vulval intraepithelial neoplasia in British women. Cancer Res. 2003;63(2):400-3. PubMed PMID: 12543794.

51. Qiu X, Zhang F, Chen D, Azad AK, Zhang L, Yuan Y, Jiang Z, Liu W, Tan Y, Tao N. HLA-B*07 is a high risk allele for familial cervical cancer. Asian Pac J Cancer Prev. 2011;12(10):2597-600. PubMed PMID: 22320938.

52. Hildesheim A, Schiffman M, Scott DR, Marti D, Kissner T, Sherman ME, Glass AG, Manos MM, Lorincz AT, Kurman RJ, Buckland J, Rush BB, Carrington M. Human leukocyte antigen class I/II alleles and development of human papillomavirus-related cervical neoplasia: results from a case-control study conducted in the United States. Cancer Epidemiol Biomarkers Prev. 1998;7(11):1035-41. PubMed PMID: 9829713.

53. Grandis JR, Falkner DM, Melhem MF, Gooding WE, Drenning SD, Morel PA. Human leukocyte antigen class I allelic and haplotype loss in squamous cell carcinoma of the head and neck: clinical and immunogenetic consequences. Clin Cancer Res. 2000;6(7):2794-802. PubMed PMID: 10914726.

54. Nasman A, Andersson E, Nordfors C, Grun N, Johansson H, Munck-Wikland E, Massucci G, Dalianis T, Ramqvist T. MHC class I expression in HPV positive and negative

tonsillar squamous cell carcinoma in correlation to clinical outcome. Int J Cancer. 2013;132(1):72-81. doi: 10.1002/ijc.27635. PubMed PMID: 22592660.

55. Shukla SA, Rooney MS, Rajasagi M, Tiao G, Dixon PM, Lawrence MS, Stevens J, Lane WJ, Dellagatta JL, Steelman S, Sougnez C, Cibulskis K, Kiezun A, Hacohen N, Brusic V, Wu CJ, Getz G. Comprehensive analysis of cancer-associated somatic mutations in class I HLA genes. Nat Biotechnol. 2015;33(11):1152-8. doi: 10.1038/nbt.3344. PubMed PMID: 26372948; PMCID: PMC4747795.

56. Ludwig ML, Kulkarni A, Birkeland AC, Michmerhuizen NL, Foltin SK, Mann JE, Hoesli RC, Devenport SN, Jewell BM, Shuman AG, Spector ME, Carey TE, Jiang H, Brenner JC. The genomic landscape of UM-SCC oral cavity squamous cell carcinoma cell lines. Oral Oncol. 2018;87:144-51. doi: 10.1016/j.oraloncology.2018.10.031. PubMed PMID: 30527230; PMCID: PMC6349383.

57. Mann JE, Kulkarni A, Birkeland AC, Kafelghazal J, Eisenberg J, Jewell BM, Ludwig ML, Spector ME, Jiang H, Carey TE, Brenner JC. The molecular landscape of the University of Michigan laryngeal squamous cell carcinoma cell line panel. Head Neck. 2019;41(9):3114-24. doi: 10.1002/hed.25803. PubMed PMID: 31090975; PMCID: PMC6692232.

58. Fairfax BP, Makino S, Radhakrishnan J, Plant K, Leslie S, Dilthey A, Ellis P, Langford C, Vannberg FO, Knight JC. Genetics of gene expression in primary immune cells identifies cell type-specific master regulators and roles of HLA alleles. Nat Genet. 2012;44(5):502-10. doi: 10.1038/ng.2205. PubMed PMID: 22446964; PMCID: PMC3437404.

59. Robinson J, Soormally AR, Hayhurst JD, Marsh SGE. The IPD-IMGT/HLA Database - New developments in reporting HLA variation. Hum Immunol. 2016;77(3):233-7. doi: 10.1016/j.humimm.2016.01.020. PubMed PMID: 26826444.

60. Gensterblum-Miller E, Wu W, Sawalha AH. Novel Transcriptional Activity and Extensive Allelic Imbalance in the Human MHC Region. J Immunol. 2018;200(4):1496-503. doi: 10.4049/jimmunol.1701061. PubMed PMID: 29311362; PMCID: PMC5823012.

61. Hafner N, Driesch C, Gajda M, Jansen L, Kirchmayr R, Runnebaum IB, Durst M. Integration of the HPV16 genome does not invariably result in high levels of viral oncogene transcripts. Oncogene. 2008;27(11):1610-7. doi: 10.1038/sj.onc.1210791. PubMed PMID: 17828299.

62. Scheffner M, Werness BA, Huibregtse JM, Levine AJ, Howley PM. The E6 oncoprotein encoded by human papillomavirus types 16 and 18 promotes the degradation of p53. Cell. 1990;63(6):1129-36. PubMed PMID: 2175676.

63. van Houten VM, Snijders PJ, van den Brekel MW, Kummer JA, Meijer CJ, van Leeuwen B, Denkers F, Smeele LE, Snow GB, Brakenhoff RH. Biological evidence that human papillomaviruses are etiologically involved in a subgroup of head and neck squamous cell carcinomas. Int J Cancer. 2001;93(2):232-5. doi: 10.1002/ijc.1313. PubMed PMID: 11410871.

64. Dyson N, Guida P, Munger K, Harlow E. Homologous sequences in adenovirus E1A and human papillomavirus E7 proteins mediate interaction with the same set of cellular proteins. J Virol. 1992;66(12):6893-902. PubMed PMID: 1331501; PMCID: PMC240306.

65. Heller C, Weisser T, Mueller-Schickert A, Rufer E, Hoh A, Leonhardt RM, Knittler MR. Identification of key amino acid residues that determine the ability of high risk HPV16-E7 to dysregulate major histocompatibility complex class I expression. J Biol Chem. 2011;286(13):10983-97. doi: 10.1074/jbc.M110.199190. PubMed PMID: 21321113; PMCID: PMC3064153.

66. Georgopoulos NT, Proffitt JL, Blair GE. Transcriptional regulation of the major histocompatibility complex (MHC) class I heavy chain, TAP1 and LMP2 genes by the human papillomavirus (HPV) type 6b, 16 and 18 E7 oncoproteins. Oncogene. 2000;19(42):4930-5. doi: 10.1038/sj.onc.1203860. PubMed PMID: 11039910.

67. Li H, Zhan T, Li C, Liu M, Wang QK. Repression of MHC class I transcription by HPV16E7 through interaction with a putative RXRbeta motif and NF-kappaB cytoplasmic sequestration. Biochem Biophys Res Commun. 2009;388(2):383-8. doi: 10.1016/j.bbrc.2009.08.019. PubMed PMID: 19665994.

68. Vita R, Mahajan S, Overton JA, Dhanda SK, Martini S, Cantrell JR, Wheeler DK, Sette A, Peters B. The Immune Epitope Database (IEDB): 2018 update. Nucleic Acids Res. 2019;47(D1):D339-D43. doi: 10.1093/nar/gky1006. PubMed PMID: 30357391; PMCID: PMC6324067.

69. Simova J, Pollakova V, Indrova M, Mikyskova R, Bieblova J, Stepanek I, Bubenik J, Reinis M. Immunotherapy augments the effect of 5-azacytidine on HPV16-associated tumours with different MHC class I-expression status. Br J Cancer. 2011;105(10):1533-41. doi: 10.1038/bjc.2011.428. PubMed PMID: 22015556; PMCID: PMC3242529.

70. Tonon G, Modi S, Wu L, Kubo A, Coxon AB, Komiya T, O'Neil K, Stover K, El-Naggar A, Griffin JD, Kirsch IR, Kaye FJ. t(11;19)(q21;p13) translocation in mucoepidermoid carcinoma creates a novel fusion product that disrupts a Notch signaling pathway. Nat Genet. 2003;33(2):208-13. doi: 10.1038/ng1083. PubMed PMID: 12539049.

71. Chen J, Li JL, Chen Z, Griffin JD, Wu L. Gene expression profiling analysis of CRTC1-MAML2 fusion oncogene-induced transcriptional program in human mucoepidermoid carcinoma cells. BMC Cancer. 2015;15:803. Epub 2015/10/28. doi: 10.1186/s12885-015-1827-3. PubMed PMID: 26503699; PMCID: PMC4624166.

72. Musicant AM, Parag-Sharma K, Gong W, Sengupta M, Chatterjee A, Henry EC, Tsai YH, Hayward MC, Sheth S, Betancourt R, Hackman TG, Padilla RJ, Parker JS, Giudice J, Flaveny CA, Hayes DN, Amelio AL. CRTC1/MAML2 directs a PGC-1alpha-IGF-1 circuit that confers vulnerability to PPARgamma inhibition. Cell reports. 2021;34(8):108768. Epub 2021/02/25. doi: 10.1016/j.celrep.2021.108768. PubMed PMID: 33626346; PMCID: PMC7955229.

73. Amelio AL, Fallahi M, Schaub FX, Zhang M, Lawani MB, Alperstein AS, Southern MR, Young BM, Wu L, Zajac-Kaye M, Kaye FJ, Cleveland JL, Conkright MD. CRTC1/MAML2 gain-of-function interactions with MYC create a gene signature predictive of cancers with CREB-MYC involvement. Proc Natl Acad Sci U S A. 2014;111(32):E3260-8. doi: 10.1073/pnas.1319176111. PubMed PMID: 25071166; PMCID: PMC4136562.

74. Wu L, Liu J, Gao P, Nakamura M, Cao Y, Shen H, Griffin JD. Transforming activity of MECT1-MAML2 fusion oncoprotein is mediated by constitutive CREB activation. EMBO J. 2005;24(13):2391-402. doi: 10.1038/sj.emboj.7600719. PubMed PMID: 15961999; PMCID: PMC1173159.

75. Chattopadhyay K. A comprehensive review on host genetic susceptibility to human papillomavirus infection and progression to cervical cancer. Indian J Hum Genet. 2011;17(3):132-44. doi: 10.4103/0971-6866.92087. PubMed PMID: 22345983; PMCID: PMC3276980.

76. Girdlestone J, Isamat M, Gewert D, Milstein C. Transcriptional regulation of HLA-A and -B: differential binding of members of the Rel and IRF families of transcription factors. Proc Natl Acad Sci U S A. 1993;90(24):11568-72. doi: 10.1073/pnas.90.24.11568. PubMed PMID: 8265591; PMCID: PMC48025.

77. Meissner TB, Liu YJ, Lee KH, Li A, Biswas A, van Eggermond MC, van den Elsen PJ, Kobayashi KS. NLRC5 cooperates with the RFX transcription factor complex to induce MHC class I gene expression. J Immunol. 2012;188(10):4951-8. doi: 10.4049/jimmunol.1103160. PubMed PMID: 22490869; PMCID: PMC3345046.

78. Gobin SJ, Keijsers V, van Zutphen M, van den Elsen PJ. The role of enhancer A in the locus-specific transactivation of classical and nonclassical HLA class I genes by nuclear factor kappa B. J Immunol. 1998;161(5):2276-83. PubMed PMID: 9725221.

79. Ladasky JJ, Boyle S, Seth M, Li H, Pentcheva T, Abe F, Steinberg SJ, Edidin M. Bap31 enhances the endoplasmic reticulum export and quality control of human class I MHC molecules. J Immunol. 2006;177(9):6172-81. PubMed PMID: 17056546; PMCID: PMC1978250.

80. Hewitt EW. The MHC class I antigen presentation pathway: strategies for viral immune evasion. Immunology. 2003;110(2):163-9. doi: 10.1046/j.1365-2567.2003.01738.x. PubMed PMID: 14511229; PMCID: PMC1783040.

81. Rizvi SM, Del Cid N, Lybarger L, Raghavan M. Distinct functions for the glycans of tapasin and heavy chains in the assembly of MHC class I molecules. J Immunol. 2011;186(4):2309-20. doi: 10.4049/jimmunol.1002959. PubMed PMID: 21263072; PMCID: PMC3057457.

82. Jeffus SK, Gardner JM, Steliga MA, Shah AA, Stelow EB, Arnaoutakis K. Hyalinizing Clear Cell Carcinoma of the Lung: Case Report and Review of the Literature. American journal of clinical pathology. 2017;148(1):73-80. doi: 10.1093/ajcp/aqx048. PubMed PMID: 28927164.

83. Gushi E, Seki U, Orhan K. Hyalinizing Clear Cell Carcinoma of Maxilla. Journal of clinical and diagnostic research : JCDR. 2017;11(7):ZL01-ZL2. doi: 10.7860/JCDR/2017/29193.10228. PubMed PMID: 28893056; PMCID: 5583847.

84. Burgess B, Ananthanarayanan V, Charous S. Hyalinizing Clear Cell Carcinoma of the Tonsil: A Case Report. Head and neck pathology. 2017;11(4):580-3. doi: 10.1007/s12105-017-0822-1. PubMed PMID: 28509953; PMCID: 5677072.

85. Shahi M, Dolan M, Murugan P. Hyalinizing Clear Cell Carcinoma of the Bronchus. Head and neck pathology. 2017;11(4):575-9. doi: 10.1007/s12105-017-0820-3. PubMed PMID: 28508996; PMCID: 5677074.

86. Fukuda A, Tagami Y, Takasawa A, Sugita S, Kuramoto R, Imai S, Hasegawa T, Iizuka K. Nasopharyngeal hyalinizing clear cell carcinoma with EWSR1 rearrangements diagnosed by fluorescence in situ hybridization. Auris, nasus, larynx. 2015;42(5):412-5. doi: 10.1016/j.anl.2015.02.015. PubMed PMID: 25805066.

87. Tang SK, Wan SK, Chan JK. Hyalinizing clear cell carcinoma of salivary gland: report of a case with multiple recurrences over 12 years. The American journal of surgical pathology. 1995;19(2):240-1. PubMed PMID: 7530410.

88. Batsakis JG, el-Naggar AK, Luna MA. Hyalinizing clear cell carcinoma of salivary origin. The Annals of otology, rhinology, and laryngology. 1994;103(9):746-8. doi: 10.1177/000348949410300917. PubMed PMID: 8085740.

89. Milchgrub S, Gnepp DR, Vuitch F, Delgado R, Albores-Saavedra J. Hyalinizing clear cell carcinoma of salivary gland. The American journal of surgical pathology. 1994;18(1):74-82. PubMed PMID: 7506496.

90. Yang XH, Liu L, Shi YY, Hu YJ, Hu QG, Zhang P. Hyalinizing clear cell carcinoma of salivary gland origin in the head and neck: clinical and histopathological analysis. International journal of oral and maxillofacial surgery. 2018;47(6):692-8. doi: 10.1016/j.ijom.2017.11.004. PubMed PMID: 29217083.

91. Roby BB, Pambuccian SE, Khariwala SS. Pathology quiz case 2. Hyalinizing clear cell carcinoma. Archives of otolaryngology--head & neck surgery. 2012;138(2):207. doi: 10.1001/archoto.2011.1245a. PubMed PMID: 22351872; PMCID: 5548609.

92. O'Sullivan-Mejia ED, Massey HD, Faquin WC, Powers CN. Hyalinizing clear cell carcinoma: report of eight cases and a review of literature. Head and neck pathology. 2009;3(3):179-85. doi: 10.1007/s12105-009-0124-3. PubMed PMID: 20596970; PMCID: 2811632.

93. Lai G, Nemolato S, Lecca S, Parodo G, Medda C, Faa G. The role of immunohistochemistry in the diagnosis of hyalinizing clear cell carcinoma of the minor salivary gland: a case report. European journal of histochemistry : EJH. 2008;52(4):251-4. PubMed PMID: 19109100.

94. Milchgrub S, Vuitch F, Saboorian MH, Hameed A, Wu H, Albores-Saavedra J. Hyalinizing clear-cell carcinoma of salivary glands in fine-needle aspiration. Diagnostic cytopathology. 2000;23(5):333-7. PubMed PMID: 11074629.

95. Antonescu CR, Katabi N, Zhang L, Sung YS, Seethala RR, Jordan RC, Perez-Ordonez B, Have C, Asa SL, Leong IT, Bradley G, Klieb H, Weinreb I. EWSR1-ATF1 fusion is a novel and consistent finding in hyalinizing clear-cell carcinoma of salivary gland. Genes, chromosomes & cancer. 2011;50(7):559-70. doi: 10.1002/gcc.20881. PubMed PMID: 21484932.

96. Chapman E, Skalova A, Ptakova N, Martinek P, Goytain A, Tucker T, Xiong W, Leader M, Kudlow BA, Haimes JD, Hayes MM, Bohus P, Miesbauerova M, Lee CH, Ng TL. Molecular Profiling of Hyalinizing Clear Cell Carcinomas Revealed a Subset of Tumors Harboring a Novel EWSR1-CREM Fusion: Report of 3 Cases. The American journal of surgical pathology. 2018;42(9):1182-9. doi: 10.1097/PAS.0000000000001114. PubMed PMID: 29975250.

97. Birkeland AC, Burgin SJ, Yanik M, Scott MV, Bradford CR, McHugh JB, McLean SA, Sullivan SE, Nor JE, McKean EL, Brenner JC. Pathogenetic Analysis of Sinonasal Teratocarcinosarcomas Reveal Actionable beta-catenin Overexpression and a beta-catenin Mutation. Journal of neurological surgery Part B, Skull base. 2017;78(4):346-52. doi: 10.1055/s-0037-1601320. PubMed PMID: 28725522; PMCID: 5515660.

98. Birkeland AC, Foltin SK, Michmerhuizen NL, Hoesli RC, Rosko AJ, Byrd S, Yanik M, Nor JE, Bradford CR, Prince ME, Carey TE, McHugh JB, Spector ME, Brenner JC. Correlation of Crtc1/3-Maml2 fusion status, grade and survival in mucoepidermoid carcinoma. Oral Oncol. 2017;68:5-8. doi: 10.1016/j.oraloncology.2017.02.025. PubMed PMID: 28438292; PMCID: PMC5433350.

99. Birkeland AC, Yanik M, Tillman BN, Scott MV, Foltin SK, Mann JE, Michmerhuizen NL, Ludwig ML, Sandelski MM, Komarck CM, Carey TE, Prince ME, Bradford CR, McHugh JB, Spector ME, Brenner JC. Identification of Targetable ERBB2 Aberrations in Head and Neck Squamous Cell Carcinoma. JAMA otolaryngology-- head & neck surgery. 2016. doi: 10.1001/jamaoto.2016.0335. PubMed PMID: 27077364.

100. Smith J, Kulkarni A, Birkeland AC, McHugh JB, Brenner JC. Whole-Exome Sequencing of Sinonasal Small Cell Carcinoma Arising within a Papillary Schneiderian Carcinoma In Situ. Otolaryngology--head and neck surgery : official journal of American Academy of Otolaryngology-Head and Neck Surgery. 2018:194599818774004. doi: 10.1177/0194599818774004. PubMed PMID: 29734873.

101. Kautto EA, Bonneville R, Miya J, Yu L, Krook MA, Reeser JW, Roychowdhury S. Performance evaluation for rapid detection of pan-cancer microsatellite instability with MANTIS. Oncotarget. 2017;8(5):7452-63. doi: 10.18632/oncotarget.13918. PubMed PMID: 27980218; PMCID: 5352334.

102. Chandrani P, Kulkarni V, Iyer P, Upadhyay P, Chaubal R, Das P, Mulherkar R, Singh R, Dutt A. NGS-based approach to determine the presence of HPV and their sites of integration in human cancer genome. British journal of cancer. 2015;112(12):1958-65. doi: 10.1038/bjc.2015.121. PubMed PMID: 25973533; PMCID: 4580395.

103. Swiecicki PL, Durm G, Bellile E, Bhangale A, Brenner JC, Worden FP. A multi-center phase II trial evaluating the efficacy of palbociclib in combination with carboplatin for the treatment of unresectable recurrent or metastatic head and neck squamous cell carcinoma. Invest New Drugs. 2020. Epub 2020/01/26. doi: 10.1007/s10637-020-00898-2. PubMed PMID: 31981071.

104. Mann JE, Smith JD, Birkeland AC, Bellile E, Swiecicki P, Mierzwa M, Chinn SB, Shuman AG, Malloy KM, Casper KA, McLean SA, Moyer JS, Wolf GT, Bradford CR, Prince ME, Carey TE, McHugh JB, Spector ME, Brenner JC. Analysis of tumor-infiltrating CD103 resident memory T-cell content in recurrent laryngeal squamous cell carcinoma. Cancer immunology, immunotherapy : CII. 2018. doi: 10.1007/s00262-018-2256-3. PubMed PMID: 30361882.

105. Liu J, Pan S, Hsieh MH, Ng N, Sun F, Wang T, Kasibhatla S, Schuller AG, Li AG, Cheng D, Li J, Tompkins C, Pferdekamper A, Steffy A, Cheng J, Kowal C, Phung V, Guo G, Wang Y, Graham MP, Flynn S, Brenner JC, Li C, Villarroel MC, Schultz PG, Wu X, McNamara P, Sellers WR, Petruzzelli L, Boral AL, Seidel HM, McLaughlin ME, Che J, Carey TE, Vanasse G, Harris JL. Targeting Wnt-driven cancer through the inhibition of Porcupine by LGK974. Proceedings of the National Academy of Sciences of the United States of America. 2013;110(50):20224-9. doi: 10.1073/pnas.1314239110. PubMed PMID: 24277854; PMCID: 3864356.

106. Camidge DR, Dziadziuszko R, Peters S, Mok T, Noe J, Nowicka M, Gadgeel SM, Cheema P, Pavlakis N, de Marinis F, Cho BC, Zhang L, Moro-Sibilot D, Liu T, Bordogna W, Balas B, Müller B, Shaw AT. Updated Efficacy and Safety Data and Impact of the EML4-ALK Fusion Variant on the Efficacy of Alectinib in Untreated ALK-Positive Advanced Non-Small Cell Lung Cancer in the Global Phase III ALEX Study. Journal of thoracic oncology : official publication of the International Association for the Study of Lung Cancer. 2019;14(7):1233-43. Epub 2019/03/25. doi: 10.1016/j.jtho.2019.03.007. PubMed PMID: 30902613.

107. Haimes JD, Stewart CJR, Kudlow BA, Culver BP, Meng B, Koay E, Whitehouse A, Cope N, Lee JC, Ng T, McCluggage WG, Lee CH. Uterine Inflammatory Myofibroblastic Tumors Frequently Harbor ALK Fusions With IGFBP5 and THBS1. The American journal of surgical pathology. 2017;41(6):773-80. Epub 2017/05/11. doi: 10.1097/pas.0000000000000801. PubMed PMID: 28490045.

108. Soda M, Choi YL, Enomoto M, Takada S, Yamashita Y, Ishikawa S, Fujiwara S, Watanabe H, Kurashina K, Hatanaka H, Bando M, Ohno S, Ishikawa Y, Aburatani H, Niki T, Sohara Y, Sugiyama Y, Mano H. Identification of the transforming EML4-ALK fusion gene in non-small-cell lung cancer. Nature. 2007;448(7153):561-6. Epub 2007/07/13. doi: 10.1038/nature05945. PubMed PMID: 17625570.

109. Slagsvold T, Marchese A, Brech A, Stenmark H. CISK attenuates degradation of the chemokine receptor CXCR4 via the ubiquitin ligase AIP4. The EMBO journal. 2006;25(16):3738-49. doi: 10.1038/sj.emboj.7601267. PubMed PMID: 16888620; PMCID: 1553197.

110. Charville GW, Wang WL, Ingram DR, Roy A, Thomas D, Patel RM, Hornick JL, van de Rijn M, Lazar AJ. EWSR1 fusion proteins mediate PAX7 expression in Ewing sarcoma. Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc. 2017;30(9):1312-20. doi: 10.1038/modpathol.2017.49. PubMed PMID: 28643791.

111. Camoes MJ, Paulo P, Ribeiro FR, Barros-Silva JD, Almeida M, Costa VL, Cerveira N, Skotheim RI, Lothe RA, Henrique R, Jeronimo C, Teixeira MR. Potential downstream target genes of aberrant ETS transcription factors are differentially affected in Ewing's sarcoma and prostate carcinoma. PloS one. 2012;7(11):e49819. doi: 10.1371/journal.pone.0049819. PubMed PMID: 23185447; PMCID: 3501462.

112. Qu X, Wu Z, Dong W, Zhang T, Wang L, Pang Z, Ma W, Du J. Update of IGF-1 receptor inhibitor (ganitumab, dalotuzumab, cixutumumab, teprotumumab and figitumumab) effects on cancer therapy. Oncotarget. 2017;8(17):29501-18. doi: 10.18632/oncotarget.15704. PubMed PMID: 28427155; PMCID: 5438747.

113. Yee D. Anti-insulin-like growth factor therapy in breast cancer. Journal of molecular endocrinology. 2018;61(1):T61-T8. doi: 10.1530/JME-17-0261. PubMed PMID: 29378771.

114. Eckert F, Schilbach K, Klumpp L, Bardoscia L, Sezgin EC, Schwab M, Zips D, Huber SM. Potential Role of CXCR4 Targeting in the Context of Radiotherapy and Immunotherapy of Cancer. Frontiers in immunology. 2018;9:3018. doi: 10.3389/fimmu.2018.03018. PubMed PMID: 30622535; PMCID: 6308162.

115. Xu C, Zhao H, Chen H, Yao Q. CXCR4 in breast cancer: oncogenic role and therapeutic targeting. Drug design, development and therapy. 2015;9:4953-64. doi: 10.2147/DDDT.S84932. PubMed PMID: 26356032; PMCID: 4560524.

116. Okuyama Kishima M, de Oliveira CE, Banin-Hirata BK, Losi-Guembarovski R, Brajao de Oliveira K, Amarante MK, Watanabe MA. Immunohistochemical expression of CXCR4 on breast cancer and its clinical significance. Analytical cellular pathology. 2015;2015:891020. doi: 10.1155/2015/891020. PubMed PMID: 26161302; PMCID: 4486754.

117. Dombrowski ND, Wolter NE, Irace AL, Cunningham MJ, Mack JW, Marcus KJ, Vargas SO, Perez-Atayde AR, Robson CD, Rahbar R. Mucoepidermoid carcinoma of the head and neck in children. Int J Pediatr Otorhinolaryngol. 2019;120:93-9. doi: 10.1016/j.ijporl.2019.02.020. PubMed PMID: 30772619.

118. Granic M, Suton P, Mueller D, Cvrljevic I, Luksic I. Prognostic factors in head and neck mucoepidermoid carcinoma: experience at a single institution based on 64 consecutive patients over a 28-year period. Int J Oral Maxillofac Surg. 2018;47(3):283-8. doi: 10.1016/j.ijom.2017.09.005. PubMed PMID: 28969884.

119. Nance MA, Seethala RR, Wang Y, Chiosea SI, Myers EN, Johnson JT, Lai SY. Treatment and survival outcomes based on histologic grading in patients with head and neck mucoepidermoid carcinoma. Cancer. 2008;113(8):2082-9. Epub 2008/08/23. doi: 10.1002/cncr.23825. PubMed PMID: 18720358; PMCID: PMC2746751.

120. Spiro RH. Salivary neoplasms: overview of a 35-year experience with 2,807 patients. Head Neck Surg. 1986;8(3):177-84. doi: 10.1002/hed.2890080309. PubMed PMID: 3744850.

121. McHugh CH, Roberts DB, El-Naggar AK, Hanna EY, Garden AS, Kies MS, Weber RS, Kupferman ME. Prognostic factors in mucoepidermoid carcinoma of the salivary glands. Cancer. 2012;118(16):3928-36. doi: 10.1002/cncr.26697. PubMed PMID: 22180391.

122. Byrd SA, Spector ME, Carey TE, Bradford CR, McHugh JB. Predictors of recurrence and survival for head and neck mucoepidermoid carcinoma. Otolaryngol Head Neck Surg. 2013;149(3):402-8. Epub 2013/05/23. doi: 10.1177/0194599813489659. PubMed PMID: 23695589; PMCID: PMC4106041.

123. Chen Z, Chen J, Gu Y, Hu C, Li JL, Lin S, Shen H, Cao C, Gao R, Li J, Ha PK, Kaye FJ, Griffin JD, Wu L. Aberrantly activated AREG-EGFR signaling is required for the growth and survival of CRTC1-MAML2 fusion-positive mucoepidermoid carcinoma cells. Oncogene. 2014;33(29):3869-77. doi: 10.1038/onc.2013.348. PubMed PMID: 23975434.

124. Kang H, Tan M, Bishop JA, Jones S, Sausen M, Ha PK, Agrawal N. Whole-Exome Sequencing of Salivary Gland Mucoepidermoid Carcinoma. Clinical cancer research : an official journal of the American Association for Cancer Research. 2017;23(1):283-8. Epub 2016/06/25. doi: 10.1158/1078-0432.Ccr-16-0720. PubMed PMID: 27340278; PMCID: PMC5182193.

125. Shinomiya H, Ito Y, Kubo M, Yonezawa K, Otsuki N, Iwae S, Inagaki H, Nibu KI. Expression of amphiregulin in mucoepidermoid carcinoma of the major salivary glands: a molecular and clinicopathological study. Hum Pathol. 2016;57:37-44. doi: 10.1016/j.humpath.2016.06.016. PubMed PMID: 27393417.

126. Yan K, Yesensky J, Hasina R, Agrawal N. Genomics of mucoepidermoid and adenoid cystic carcinomas. Laryngoscope investigative otolaryngology. 2018;3(1):56-61. Epub 2018/03/02. doi: 10.1002/lio2.139. PubMed PMID: 29492469; PMCID: PMC5824110.

127. Behboudi A, Enlund F, Winnes M, Andren Y, Nordkvist A, Leivo I, Flaberg E, Szekely L, Makitie A, Grenman R, Mark J, Stenman G. Molecular classification of mucoepidermoid carcinomas-prognostic significance of the MECT1-MAML2 fusion oncogene. Genes Chromosomes Cancer. 2006;45(5):470-81. doi: 10.1002/gcc.20306. PubMed PMID: 16444749.

128. Komiya T, Park Y, Modi S, Coxon AB, Oh H, Kaye FJ. Sustained expression of Mect1-Maml2 is essential for tumor cell growth in salivary gland cancers carrying the t(11;19) translocation. Oncogene. 2006;25(45):6128-32. Epub 2006/05/03. doi: 10.1038/sj.onc.1209627. PubMed PMID: 16652146.

129. Okumura Y, Miyabe S, Nakayama T, Fujiyoshi Y, Hattori H, Shimozato K, Inagaki H. Impact of CRTC1/3-MAML2 fusions on histological classification and prognosis of mucoepidermoid carcinoma. Histopathology. 2011;59(1):90-7. doi: 10.1111/j.1365-2559.2011.03890.x. PubMed PMID: 21668476.

130. Nakayama T, Miyabe S, Okabe M, Sakuma H, Ijichi K, Hasegawa Y, Nagatsuka H, Shimozato K, Inagaki H. Clinicopathological significance of the CRTC3-MAML2 fusion

transcript in mucoepidermoid carcinoma. Mod Pathol. 2009;22(12):1575-81. Epub 2009/09/15. doi: 10.1038/modpathol.2009.126. PubMed PMID: 19749740.

131. Guerrero-Preston R, Godoy-Vitorino F, Jedlicka A, Rodriguez-Hilario A, Gonzalez H, Bondy J, Lawson F, Folawiyo O, Michailidi C, Dziedzic A, Thangavel R, Hadar T, Noordhuis MG, Westra W, Koch W, Sidransky D. 16S rRNA amplicon sequencing identifies microbiota associated with oral cancer, human papilloma virus infection and surgical treatment. Oncotarget. 2016;7(32):51320-34. Epub 2016/06/05. doi: 10.18632/oncotarget.9710. PubMed PMID: 27259999; PMCID: PMC5239478.

132. Coxon A, Rozenblum E, Park YS, Joshi N, Tsurutani J, Dennis PA, Kirsch IR, Kaye FJ. Mect1-Maml2 fusion oncogene linked to the aberrant activation of cyclic AMP/CREB regulated genes. Cancer Res. 2005;65(16):7137-44. doi: 10.1158/0008-5472.CAN-05-1125. PubMed PMID: 16103063.

133. Kar A, Adeniji A, Rao VUS, Ghosh M. Molecular landscape of salivary gland cancers. Oral oncology. 2020:104595. Epub 2020/02/23. doi: 10.1016/j.oraloncology.2020.104595. PubMed PMID: 32081526.

134. Brenner JC, Graham MP, Kumar B, Saunders LM, Kupfer R, Lyons RH, Bradford CR, Carey TE. Genotyping of 73 UM-SCC head and neck squamous cell carcinoma cell lines. Head Neck. 2010;32(4):417-26. doi: 10.1002/hed.21198. PubMed PMID: 19760794; PMCID: 3292176.

135. Takahashi Y, Lee J, Pickering C, Bell D, Jiffar TW, Myers JN, Hanna EY, Kupferman ME. Human epidermal growth factor receptor 2/neu as a novel therapeutic target in sinonasal undifferentiated carcinoma. Head & neck. 2016;38 Suppl 1:E1926-34. Epub 2016/01/12. doi: 10.1002/hed.24350. PubMed PMID: 26752332; PMCID: PMC6453572.

136. De Coster W, D'Hert S, Schultz DT, Cruts M, Van Broeckhoven C. NanoPack: visualizing and processing long-read sequencing data. Bioinformatics. 2018;34(15):2666-9. doi: 10.1093/bioinformatics/bty149. PubMed PMID: 29547981; PMCID: PMC6061794.

137. Ren J, Chaisson MJP. lra: A long read aligner for sequences and contigs. PLoS Comput Biol. 2021;17(6):e1009078. doi: 10.1371/journal.pcbi.1009078. PubMed PMID: 34153026; PMCID: PMC8248648.

138. Kent WJ. BLAT--the BLAST-like alignment tool. Genome Res. 2002;12(4):656-64. doi: 10.1101/gr.229202. PubMed PMID: 11932250; PMCID: PMC187518.

139. Li H. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics. 2018;34(18):3094-100. doi: 10.1093/bioinformatics/bty191. PubMed PMID: 29750242; PMCID: PMC6137996.

140. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Genome Project Data Processing S. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009;25(16):2078-9. doi: 10.1093/bioinformatics/btp352. PubMed PMID: 19505943; PMCID: PMC2723002.

141. Tham CY, Tirado-Magallanes R, Goh Y, Fullwood MJ, Koh BTH, Wang W, Ng CH, Chng WJ, Thiery A, Tenen DG, Benoukraf T. NanoVar: accurate characterization of patients' genomic structural variants using low-depth nanopore sequencing. Genome Biol. 2020;21(1):56. doi: 10.1186/s13059-020-01968-7. PubMed PMID: 32127024; PMCID: PMC7055087.

142. Heller D, Vingron M. SVIM: structural variant identification using mapped long reads. Bioinformatics. 2019;35(17):2907-15. doi: 10.1093/bioinformatics/btz041. PubMed PMID: 30668829; PMCID: PMC6735718.

143. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010;26(6):841-2. doi: 10.1093/bioinformatics/btq033. PubMed PMID: 20110278; PMCID: PMC2832824.

144. Murcia Pienkowski V, Kucharczyk M, Młynek M, Szczałuba K, Rydzanicz M, Poszewiecka B, Skórka A, Sykulski M, Biernacka A, Koppolu AA, Posmyk R, Walczak A, Kosińska J, Krajewski P, Castaneda J, Obersztyn E, Jurkiewicz E, Śmigiel R, Gambin A, Chrzanowska K, Krajewska-Walasek M, Płoski R. Mapping of breakpoints in balanced chromosomal translocations by shallow whole-genome sequencing points to EFNA5, BAHD1 and PPP2R5E as novel candidates for genes causing human Mendelian disorders. J Med Genet. 2019;56(2):104-12. Epub 2018/10/26. doi: 10.1136/jmedgenet-2018-105527. PubMed PMID: 30352868.

145. Pinatti L, Gu W, Wang Y, El Hossiny A, Bhangale A, Brummel C, Carey T, Mills R, Brenner J. SearcHPV: A Novel Approach to Identify and Assemble Human Papillomavirus-host Genomic Integration Events in Cancer. Cancer. 2021.

146. Heft Neal ME, Birkeland AC, Bhangale AD, Zhai J, Kulkarni A, Foltin SK, Jewell BM, Ludwig ML, Pinatti L, Jiang H, McHugh JB, Marentette L, McKean EL, Brenner JC. Genetic analysis of sinonasal undifferentiated carcinoma discovers recurrent SWI/SNF alterations and a novel PGAP3-SRPK1 fusion gene. BMC Cancer. 2021;21(1):636. doi: 10.1186/s12885-021-08370-x. PubMed PMID: 34051734; PMCID: PMC8164750.

147. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25(14):1754-60. doi: 10.1093/bioinformatics/btp324. PubMed PMID: 19451168; PMCID: PMC2705234.

148. Smith J, Kulkarni A, Birkeland AC, McHugh JB, Brenner JC. Whole-Exome Sequencing of Sinonasal Small Cell Carcinoma Arising within a Papillary Schneiderian Carcinoma In Situ. Otolaryngol Head Neck Surg. 2018;159(5):859-65. Epub 2018/05/08. doi: 10.1177/0194599818774004. PubMed PMID: 29734873; PMCID: PMC6212311.

149. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010;20(9):1297-303. doi: 10.1101/gr.107524.110. PubMed PMID: 20644199; PMCID: PMC2928508.

150. Masica DL, Douville C, Tokheim C, Bhattacharya R, Kim R, Moad K, Ryan MC, Karchin R. CRAVAT 4: Cancer-Related Analysis of Variants Toolkit. Cancer research.

2017;77(21):e35-e8. Epub 2017/11/03. doi: 10.1158/0008-5472.CAN-17-0338. PubMed PMID: 29092935; PMCID: PMC5850945.

151. Amarasinghe KC, Li J, Hunter SM, Ryland GL, Cowin PA, Campbell IG, Halgamuge SK. Inferring copy number and genotype in tumour exome data. BMC Genomics. 2014;15:732. doi: 10.1186/1471-2164-15-732. PubMed PMID: 25167919; PMCID: PMC4162913.

152. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA. Circos: an information aesthetic for comparative genomics. Genome Res. 2009;19(9):1639-45. doi: 10.1101/gr.092759.109. PubMed PMID: 19541911; PMCID: PMC2752132.

153. Wang X, Zhang B. customProDB: an R package to generate customized protein databases from RNA-Seq data for proteomics search. Bioinformatics. 2013;29(24):3235-7. doi: 10.1093/bioinformatics/btt543. PubMed PMID: 24058055; PMCID: PMC3842753.

154. Szolek A, Schubert B, Mohr C, Sturm M, Feldhahn M, Kohlbacher O. OptiType: precision HLA typing from next-generation sequencing data. Bioinformatics. 2014;30(23):3310-6. doi: 10.1093/bioinformatics/btu548. PubMed PMID: 25143287; PMCID: PMC4441069.

155. Jurtz V, Paul S, Andreatta M, Marcatili P, Peters B, Nielsen M. NetMHCpan-4.0: Improved Peptide-MHC Class I Interaction Predictions Integrating Eluted Ligand and Peptide Binding Affinity Data. J Immunol. 2017;199(9):3360-8. doi: 10.4049/jimmunol.1700893. PubMed PMID: 28978689; PMCID: PMC5679736.

156. Blokzijl F, Janssen R, van Boxtel R, Cuppen E. MutationalPatterns: comprehensive genome-wide analysis of mutational processes. Genome Med. 2018;10(1):33. doi: 10.1186/s13073-018-0539-0. PubMed PMID: 29695279; PMCID: PMC5922316.

157. Niu B, Ye K, Zhang Q, Lu C, Xie M, McLellan MD, Wendl MC, Ding L. MSIsensor: microsatellite instability detection using paired tumor-normal sequence data. Bioinformatics. 2014;30(7):1015-6. doi: 10.1093/bioinformatics/btt755. PubMed PMID: 24371154; PMCID: PMC3967115.

158. Liu T, Brown TC, Juhlin CC, Andreasson A, Wang N, Backdahl M, Healy JM, Prasad ML, Korah R, Carling T, Xu D, Larsson C. The activating TERT promoter mutation C228T is recurrent in subsets of adrenal tumors. Endocr Relat Cancer. 2014;21(3):427-34. Epub 2014/05/08. doi: 10.1530/ERC-14-0016. PubMed PMID: 24803525; PMCID: PMC4045219.

159. Uhrig S, Ellermann J, Walther T, Burkhardt P, Frohlich M, Hutter B, Toprak UH, Neumann O, Stenzinger A, Scholl C, Frohling S, Brors B. Accurate and efficient detection of gene fusions from RNA sequencing data. Genome Res. 2021;31(3):448-60. doi: 10.1101/gr.257246.119. PubMed PMID: 33441414; PMCID: PMC7919457.

160. Michmerhuizen NL, Leonard E, Matovina C, Harris M, Herbst G, Kulkarni A, Zhai J, Jiang H, Carey TE, Brenner JC. Rationale for Using Irreversible Epidermal Growth Factor Receptor Inhibitors in Combination with Phosphatidylinositol 3-Kinase Inhibitors for Advanced Head and Neck Squamous Cell Carcinoma. Mol Pharmacol. 2019;95(5):528-36. doi: 10.1124/mol.118.115162. PubMed PMID: 30858165; PMCID: PMC6442321.

161. Warner KA, Adams A, Bernardi L, Nor C, Finkel KA, Zhang Z, McLean SA, Helman J, Wolf GT, Divi V, Queimado L, Kaye FJ, Castilho RM, Nor JE. Characterization of tumorigenic cell lines from the recurrence and lymph node metastasis of a human salivary mucoepidermoid carcinoma. Oral Oncol. 2013;49(11):1059-66. doi: 10.1016/j.oraloncology.2013.08.004. PubMed PMID: 24035723; PMCID: PMC3821871.

162. Li Y, Roberts ND, Wala JA, Shapira O, Schumacher SE, Kumar K, Khurana E, Waszak S, Korbel JO, Haber JE, Imielinski M, Group PSVW, Weischenfeldt J, Beroukhim R, Campbell PJ, Consortium P. Patterns of somatic structural variation in human cancer genomes. Nature. 2020;578(7793):112-21. Epub 2020/02/07. doi: 10.1038/s41586-019-1913-9. PubMed PMID: 32025012; PMCID: PMC7025897.

163. Ordulu Z, Wong KE, Currall BB, Ivanov AR, Pereira S, Althari S, Gusella JF, Talkowski ME, Morton CC. Describing sequencing results of structural chromosome rearrangements with a suggested next-generation cytogenetic nomenclature. Am J Hum Genet. 2014;94(5):695-709. Epub 2014/04/22. doi: 10.1016/j.ajhg.2014.03.020. PubMed PMID: 24746958; PMCID: PMC4067557.

164. Qian W, Guo S. Abstract 3019: An ultra-accurate NGS assay reveals great variation in mouse ratios in xenograft tumors by mouse strain and cancer type. Cancer research. 2021;81(13_Supplement):3019-. doi: 10.1158/1538-7445.Am2021-3019.

165. Tate JG, Bamford S, Jubb HC, Sondka Z, Beare DM, Bindal N, Boutselakis H, Cole CG, Creatore C, Dawson E, Fish P, Harsha B, Hathaway C, Jupe SC, Kok CY, Noble K, Ponting L, Ramshaw CC, Rye CE, Speedy HE, Stefancsik R, Thompson SL, Wang S, Ward S, Campbell PJ, Forbes SA. COSMIC: the Catalogue Of Somatic Mutations In Cancer. Nucleic Acids Research. 2018;47(D1):D941-D7. doi: 10.1093/nar/gky1015.

166. Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SA, Behjati S, Biankin AV, Bignell GR, Bolli N, Borg A, Borresen-Dale AL, Boyault S, Burkhardt B, Butler AP, Caldas C, Davies HR, Desmedt C, Eils R, Eyfjord JE, Foekens JA, Greaves M, Hosoda F, Hutter B, Ilicic T, Imbeaud S, Imielinski M, Jager N, Jones DT, Jones D, Knappskog S, Kool M, Lakhani SR, Lopez-Otin C, Martin S, Munshi NC, Nakamura H, Northcott PA, Pajic M, Papaemmanuil E, Paradiso A, Pearson JV, Puente XS, Raine K, Ramakrishna M, Richardson AL, Richter J, Rosenstiel P, Schlesner M, Schumacher TN, Span PN, Teague JW, Totoki Y, Tutt AN, Valdes-Mas R, van Buuren MM, van 't Veer L, Vincent-Salomon A, Waddell N, Yates LR, Australian Pancreatic Cancer Genome I, Consortium IBC, Consortium IM-S, PedBrain I, Zucman-Rossi J, Futreal PA, McDermott U, Lichter P, Meyerson M, Grimmond SM, Siebert R, Campo E, Shibata T, Pfister SM, Campbell PJ, Stratton MR. Signatures of mutational processes in human cancer. Nature. 2013;500(7463):415-21. doi: 10.1038/nature12477. PubMed PMID: 23945592; PMCID: PMC3776390.

167. Jiang Z, Wang H, Li L, Hou Z, Liu W, Zhou T, Li Y, Chen S. Analysis of TGCA data reveals genetic and epigenetic changes and biological function of MUC family genes in colorectal cancer. Future oncology (London, England). 2019;15(35):4031-43. Epub 2019/11/28. doi: 10.2217/fon-2019-0363. PubMed PMID: 31773991.

168. Wang J, Zhou H, Wang Y, Huang H, Yang J, Gu W, Zhang X, Yang J. Serum mucin 3A as a potential biomarker for extrahepatic cholangiocarcinoma. Saudi J Gastroenterol. 2020;26(3):129-36. Epub 2020/04/10. doi: 10.4103/sjg.SJG_447_19. PubMed PMID: 32270773; PMCID: PMC7392288.

169. Honjo K, Hiraki T, Higashi M, Noguchi H, Nomoto M, Yoshimura T, Batra SK, Yonezawa S, Semba I, Nakamura N, Tanimoto A, Yamada S. Immunohistochemical expression profiles of mucin antigens in salivary gland mucoepidermoid carcinoma: MUC4- and MUC6-negative expression predicts a shortened survival in the early postoperative phase. Histol Histopathol. 2018;33(2):201-13. Epub 2017/06/27. doi: 10.14670/HH-11-913. PubMed PMID: 28649694.

170. Robinson L, van Heerden MB, Ker-Fox JG, Hunter KD, van Heerden WFP. Expression of Mucins in Salivary Gland Mucoepidermoid Carcinoma. Head and neck pathology. 2020. Epub 2020/09/23. doi: 10.1007/s12105-020-01226-z. PubMed PMID: 32959209.

171. Bell D, Hanna EY, Miele L, Roberts D, Weber RS, El-Naggar AK. Expression and significance of notch signaling pathway in salivary adenoid cystic carcinoma. Ann Diagn Pathol. 2014;18(1):10-3. Epub 2013/11/19. doi: 10.1016/j.anndiagpath.2013.10.001. PubMed PMID: 24238845; PMCID: PMC4807115.

172. Ding LC, She L, Zheng DL, Huang QL, Wang JF, Zheng FF, Lu YG. Notch-4 contributes to the metastasis of salivary adenoid cystic carcinoma. Oncol Rep. 2010;24(2):363-8. Epub 2010/07/03. doi: 10.3892/or_00000868. PubMed PMID: 20596622.

173. Meyers RM, Bryan JG, McFarland JM, Weir BA, Sizemore AE, Xu H, Dharia NV, Montgomery PG, Cowley GS, Pantel S, Goodale A, Lee Y, Ali LD, Jiang G, Lubonja R, Harrington WF, Strickland M, Wu T, Hawes DC, Zhivich VA, Wyatt MR, Kalani Z, Chang JJ, Okamoto M, Stegmaier K, Golub TR, Boehm JS, Vazquez F, Root DE, Hahn WC, Tsherniak A. Computational correction of copy number effect improves specificity of CRISPR-Cas9 essentiality screens in cancer cells. Nature genetics. 2017;49(12):1779-84. Epub 2017/10/31. doi: 10.1038/ng.3984. PubMed PMID: 29083409; PMCID: PMC5709193.

174. DepMap BDQPfDdmfv.

175. Chen Z, Lin S, Li JL, Ni W, Guo R, Lu J, Kaye FJ, Wu L. CRTC1-MAML2 fusion-induced lncRNA LINC00473 expression maintains the growth and survival of human mucoepidermoid carcinoma cells. Oncogene. 2018;37(14):1885-95. Epub 2018/01/23. doi: 10.1038/s41388-017-0104-0. PubMed PMID: 29353885; PMCID: PMC5889358.

176. Persson M, Andren Y, Mark J, Horlings HM, Persson F, Stenman G. Recurrent fusion of MYB and NFIB transcription factor genes in carcinomas of the breast and head and neck. Proceedings of the National Academy of Sciences of the United States of America. 2009;106(44):18740-4. Epub 2009/10/21. doi: 10.1073/pnas.0909114106. PubMed PMID: 19841262; PMCID: PMC2773970.

177. Cancer Genome Atlas N. Comprehensive genomic characterization of head and neck squamous cell carcinomas. Nature. 2015;517(7536):576-82. Epub 2015/01/30. doi: 10.1038/nature14129. PubMed PMID: 25631445; PMCID: PMC4311405.

178. Haberman AS, Isaac DD, Andrew DJ. Specification of cell fates within the salivary gland primordium. Dev Biol. 2003;258(2):443-53. Epub 2003/06/12. doi: 10.1016/s0012-1606(03)00140-4. PubMed PMID: 12798300.

179. Dang H, Lin AL, Zhang B, Zhang HM, Katz MS, Yeh CK. Role for Notch signaling in salivary acinar cell growth and differentiation. Dev Dyn. 2009;238(3):724-31. Epub 2009/02/25. doi: 10.1002/dvdy.21875. PubMed PMID: 19235730; PMCID: PMC2671016.

180. Morris LGT, Chandramohan R, West L, Zehir A, Chakravarty D, Pfister DG, Wong RJ, Lee NY, Sherman EJ, Baxi SS, Ganly I, Singh B, Shah JP, Shaha AR, Boyle JO, Patel SG, Roman BR, Barker CA, McBride SM, Chan TA, Dogan S, Hyman DM, Berger MF, Solit DB, Riaz N, Ho AL. The Molecular Landscape of Recurrent and Metastatic Head and Neck Cancers: Insights From a Precision Oncology Sequencing Platform. JAMA Oncol. 2017;3(2):244-55. doi: 10.1001/jamaoncol.2016.1790. PubMed PMID: 27442865; PMCID: PMC5253129.

181. Kim H, Ahn D, Sohn JH, Kim YH, Lee JH, Lee H. TERT Promoter Mutation and Telomere Length in Salivary Gland Tumors. Pathol Oncol Res. 2018;24(3):697-8. doi: 10.1007/s12253-017-0275-6. PubMed PMID: 28664476.

182. Nikolic N, Anicic B, Carkic J, Simonovic J, Toljic B, Tanic N, Tepavcevic Z, Vukadinovic M, Konstantinovic VS, Milasin J. High frequency of p16 and p14 promoter hypermethylation and marked telomere instability in salivary gland tumors. Archives of oral biology. 2015;60(11):1662-6. Epub 2015/09/10. doi: 10.1016/j.archoralbio.2015.08.011. PubMed PMID: 26351750.

183. Ghandi M, Huang FW, Jane-Valbuena J, Kryukov GV, Lo CC, McDonald ER, 3rd, Barretina J, Gelfand ET, Bielski CM, Li H, Hu K, Andreev-Drakhlin AY, Kim J, Hess JM, Haas BJ, Aguet F, Weir BA, Rothberg MV, Paolella BR, Lawrence MS, Akbani R, Lu Y, Tiv HL, Gokhale PC, de Weck A, Mansour AA, Oh C, Shih J, Hadi K, Rosen Y, Bistline J, Venkatesan K, Reddy A, Sonkin D, Liu M, Lehar J, Korn JM, Porter DA, Jones MD, Golji J, Caponigro G, Taylor JE, Dunning CM, Creech AL, Warren AC, McFarland JM, Zamanighomi M, Kauffmann A, Stransky N, Imielinski M, Maruvka YE, Cherniack AD, Tsherniak A, Vazquez F, Jaffe JD, Lane AA, Weinstock DM, Johannessen CM, Morrissey MP, Stegmeier F, Schlegel R, Hahn WC, Getz G, Mills GB, Boehm JS, Golub TR, Garraway LA, Sellers WR. Next-generation characterization of the Cancer Cell Line Encyclopedia. Nature. 2019;569(7757):503-8. doi: 10.1038/s41586-019-1186-3. PubMed PMID: 31068700; PMCID: PMC6697103.

184. Peraza A, Gomez R, Beltran J, Amarista FJ. Mucoepidermoid carcinoma. An update and review of the literature. J Stomatol Oral Maxillofac Surg. 2020;121(6):713-20. doi: 10.1016/j.jormas.2020.06.003. PubMed PMID: 32565266.

185. Perez-de-Oliveira ME, Wagner VP, Araujo ALD, Martins MD, Santos-Silva AR, Bingle L, Vargas PA. Prognostic value of CRTC1-MAML2 translocation in salivary mucoepidermoid

carcinoma: Systematic review and meta-analysis. J Oral Pathol Med. 2020;49(5):386-94. doi: 10.1111/jop.12970. PubMed PMID: 31661572.

186. Harada K, Ferdous T, Ueyama Y. PD-L1 expression in malignant salivary gland tumors. BMC Cancer. 2018;18(1):156. doi: 10.1186/s12885-018-4069-3. PubMed PMID: 29409471; PMCID: PMC5801834.

187. Witte HM, Gebauer N, Lappohn D, Umathum VG, Riecke A, Arndt A, Steinestel K. Prognostic Impact of PD-L1 Expression in Malignant Salivary Gland Tumors as Assessed by Established Scoring Criteria: Tumor Proportion Score (TPS), Combined Positivity Score (CPS), and Immune Cell (IC) Infiltrate. Cancers (Basel). 2020;12(4). doi: 10.3390/cancers12040873. PubMed PMID: 32260165; PMCID: PMC7226358.

188. Rodriguez CP, Wu QV, Voutsinas J, Fromm JR, Jiang X, Pillarisetty VG, Lee SM, Santana-Davila R, Goulart B, Baik CS, Chow LQM, Eaton K, Martins R. A Phase II Trial of Pembrolizumab and Vorinostat in Recurrent Metastatic Head and Neck Squamous Cell Carcinomas and Salivary Gland Cancer. Clin Cancer Res. 2020;26(4):837-45. doi: 10.1158/1078-0432.CCR-19-2214. PubMed PMID: 31796519.

189. Cohen RB, Delord JP, Doi T, Piha-Paul SA, Liu SV, Gilbert J, Algazi AP, Damian S, Hong RL, Le Tourneau C, Day D, Varga A, Elez E, Wallmark J, Saraf S, Thanigaimani P, Cheng J, Keam B. Pembrolizumab for the Treatment of Advanced Salivary Gland Carcinoma: Findings of the Phase 1b KEYNOTE-028 Study. Am J Clin Oncol. 2018;41(11):1083-8. doi: 10.1097/COC.0000000000000429. PubMed PMID: 29462123; PMCID: PMC6211783 Pennsylvania and was a member of an advisory board for Bristol-Myers Squibb. T.D. received research funding from Taiho, Novartis, Merck Serono, Astellas Pharma, Merck & Co. Inc., Janssen, Boehringer Ingelheim, Takeda, Pfizer, Eli Lilly Sumitomo Group, Chugai Pharma, Bayer, Kyowa Hakko Kirin, Daiichi Sankyo, Celgene, and Amgen. S.A.P. received research funding to the institution from the National Institutes of Health (grant number P30CA016672). S.V.L. was a member of an advisory board for Genentech, Boehringer Ingelheim, Celgene, Ariad, and Pfizer. J.G. received a research grant paid to the institution from Merck & Co. Inc., AstraZeneca, Novartis, Bristol-Myers Squibb, Karyopharm, and Pfizer. A.P.A. received a research grant from Merck & Co. Inc., paid to the institution. C.L. was a member of an advisory board for Merck & Co. Inc. S.S., J.C. are employees of Merck Sharp & Dohme Corp., a subsidiary of Merck & Co. Inc., Kenilworth, NJ. P.T. is currently an employee of Merck Sharp & Dohme Corp., a subsidiary of Merck & Co. Inc., Kenilworth, NJ, and was previously an employee of Boehringer Ingelheim, from May 2012 to August 2014, and owned stock in Gilead from June 2014 to June 2016. J.P.D., S.D., R-L.H., D.D., A.V., E.E., J.W., B.K. declare no conflicts of interest.

190. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proceedings of the National Academy of Sciences of the United States of America. 2005;102(43):15545-50. Epub 2005/10/04. doi: 10.1073/pnas.0506580102. PubMed PMID: 16199517; PMCID: PMC1239896.

191. Dohda T, Maljukova A, Liu L, Heyman M, Grander D, Brodin D, Sangfelt O, Lendahl U. Notch signaling induces SKP2 expression and promotes reduction of p27Kip1 in T-cell acute lymphoblastic leukemia cell lines. Exp Cell Res. 2007;313(14):3141-52. doi: 10.1016/j.yexcr.2007.04.027. PubMed PMID: 17560996.

192. Bild AH, Yao G, Chang JT, Wang Q, Potti A, Chasse D, Joshi MB, Harpole D, Lancaster JM, Berchuck A, Olson JA, Jr., Marks JR, Dressman HK, West M, Nevins JR. Oncogenic pathway signatures in human cancers as a guide to targeted therapies. Nature. 2006;439(7074):353-7. doi: 10.1038/nature04296. PubMed PMID: 16273092.

193. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics. 2008;9:559. doi: 10.1186/1471-2105-9-559. PubMed PMID: 19114008; PMCID: PMC2631488.

194. Nicorici D, Şatalan M, Edgren H, Kangaspeska S, Murumägi A, Kallioniemi O, Virtanen S, Kilkku O. FusionCatcher – a tool for finding somatic fusion genes in paired-end RNA-sequencing data. bioRxiv. 2014:011650. doi: 10.1101/011650.

195. Chen B, Khodadoust MS, Liu CL, Newman AM, Alizadeh AA. Profiling Tumor Infiltrating Immune Cells with CIBERSORT. Methods in molecular biology. 2018;1711:243-59. doi: 10.1007/978-1-4939-7493-1_12. PubMed PMID: 29344893; PMCID: 5895181.

196. Hao Y, Hao S, Andersen-Nissen E, Mauck WM, 3rd, Zheng S, Butler A, Lee MJ, Wilk AJ, Darby C, Zager M, Hoffman P, Stoeckius M, Papalexi E, Mimitou EP, Jain J, Srivastava A, Stuart T, Fleming LM, Yeung B, Rogers AJ, McElrath JM, Blish CA, Gottardo R, Smibert P, Satija R. Integrated analysis of multimodal single-cell data. Cell. 2021;184(13):3573-87 e29. doi: 10.1016/j.cell.2021.04.048. PubMed PMID: 34062119; PMCID: PMC8238499.

197. Aibar S, Gonzalez-Blas CB, Moerman T, Huynh-Thu VA, Imrichova H, Hulselmans G, Rambow F, Marine JC, Geurts P, Aerts J, van den Oord J, Atak ZK, Wouters J, Aerts S. SCENIC: single-cell regulatory network inference and clustering. Nat Methods. 2017;14(11):1083-6. doi: 10.1038/nmeth.4463. PubMed PMID: 28991892; PMCID: PMC5937676.

198. Danaher P, Kim Y, Nelson B, Griswold M, Yang Z, Piazza E, Beechem JM. Advances in mixed cell deconvolution enable quantification of cell types in spatial transcriptomic data. Nat Commun. 2022;13(1):385. doi: 10.1038/s41467-022-28020-5. PubMed PMID: 35046414; PMCID: PMC8770643 performing this work. No authors had nonfinancial competing interests. J.M.B. is listed on the relevant U.S. patent 10640816, "Simultaneous quantification of gene expression in a user-defined region of a cross-sectioned tissue".

199. Tabula Sapiens C, Jones RC, Karkanias J, Krasnow MA, Pisco AO, Quake SR, Salzman J, Yosef N, Bulthaup B, Brown P, Harper W, Hemenez M, Ponnusamy R, Salehi A, Sanagavarapu BA, Spallino E, Aaron KA, Concepcion W, Gardner JM, Kelly B, Neidlinger N, Wang Z, Crasta S, Kolluru S, Morri M, Pisco AO, Tan SY, Travaglini KJ, Xu C, Alcantara-Hernandez M, Almanzar N, Antony J, Beyersdorf B, Burhan D, Calcuttawala K, Carter MM, Chan CKF, Chang CA, Chang S, Colville A, Crasta S, Culver RN, Cvijovic I, D'Amato G, Ezran C, Galdos FX, Gillich A, Goodyer WR, Hang Y, Hayashi A, Houshdaran S, Huang X, Irwin JC,

Jang S, Juanico JV, Kershner AM, Kim S, Kiss B, Kolluru S, Kong W, Kumar ME, Kuo AH, Leylek R, Li B, Loeb GB, Lu WJ, Mantri S, Markovic M, McAlpine PL, de Morree A, Morri M, Mrouj K, Mukherjee S, Muser T, Neuhofer P, Nguyen TD, Perez K, Phansalkar R, Pisco AO, Puluca N, Qi Z, Rao P, Raquer-McKay H, Schaum N, Scott B, Seddighzadeh B, Segal J, Sen S, Sikandar S, Spencer SP, Steffes LC, Subramaniam VR, Swarup A, Swift M, Travaglini KJ, Van Treuren W, Trimm E, Veizades S, Vijayakumar S, Vo KC, Vorperian SK, Wang W, Weinstein HNW, Winkler J, Wu TTH, Xie J, Yung AR, Zhang Y, Detweiler AM, Mekonen H, Neff NF, Sit RV, Tan M, Yan J, Bean GR, Charu V, Forgo E, Martin BA, Ozawa MG, Silva O, Tan SY, Toland A, Vemuri VNP, Afik S, Awayan K, Botvinnik OB, Byrne A, Chen M, Dehghannasiri R, Detweiler AM, Gayoso A, Granados AA, Li Q, Mahmoudabadi G, McGeever A, de Morree A, Olivieri JE, Park M, Pisco AO, Ravikumar N, Salzman J, Stanley G, Swift M, Tan M, Tan W, Tarashansky AJ, Vanheusden R, Vorperian SK, Wang P, Wang S, Xing G, Xu C, Yosef N, Alcantara-Hernandez M, Antony J, Chan CKF, Chang CA, Colville A, Crasta S, Culver R, Dethlefsen L, Ezran C, Gillich A, Hang Y, Ho PY, Irwin JC, Jang S, Kershner AM, Kong W, Kumar ME, Kuo AH, Leylek R, Liu S, Loeb GB, Lu WJ, Maltzman JS, Metzger RJ, de Morree A, Neuhofer P, Perez K, Phansalkar R, Qi Z, Rao P, Raquer-McKay H, Sasagawa K, Scott B, Sinha R, Song H, Spencer SP, Swarup A, Swift M, Travaglini KJ, Trimm E, Veizades S, Vijayakumar S, Wang B, Wang W, Winkler J, Xie J, Yung AR, Artandi SE, Beachy PA, Clarke MF, Giudice LC, Huang FW, Huang KC, Idoyaga J, Kim SK, Krasnow M, Kuo CS, Nguyen P, Quake SR, Rando TA, Red-Horse K, Reiter J, Relman DA, Sonnenburg JL, Wang B, Wu A, Wu SM, Wyss-Coray T. The Tabula Sapiens: A multiple-organ, single-cell transcriptomic atlas of humans. Science. 2022;376(6594):eabl4896. doi: 10.1126/science.abl4896. PubMed PMID: 35549404; PMCID: PMC9812260.

200. Bankhead P, Loughrey MB, Fernandez JA, Dombrowski Y, McArt DG, Dunne PD, McQuaid S, Gray RT, Murray LJ, Coleman HG, James JA, Salto-Tellez M, Hamilton PW. QuPath: Open source software for digital pathology image analysis. Sci Rep. 2017;7(1):16878. doi: 10.1038/s41598-017-17204-5. PubMed PMID: 29203879; PMCID: PMC5715110.

201. Marin-Bejar O, Mas AM, Gonzalez J, Martinez D, Athie A, Morales X, Galduroz M, Raimondi I, Grossi E, Guo S, Rouzaut A, Ulitsky I, Huarte M. The human lncRNA LINC-PINT inhibits tumor cell invasion through a highly conserved sequence element. Genome Biol. 2017;18(1):202. doi: 10.1186/s13059-017-1331-y. PubMed PMID: 29078818; PMCID: 5660458.

202. Andratschke M, Hagedorn H, Luebbers CW, Schmitt B, Lang S, Zeidler R, Wollenberg B. Limited suitability of EpCAM for molecular staging of tumor borders in head and neck cancer. Anticancer Res. 2006;26(1A):153-8. PubMed PMID: 16475692.

203. Tsakiroglou AM, Fergie M, Oguejiofor K, Linton K, Thomson D, Stern PL, Astley S, Byers R, West CML. Spatial proximity between T and PD-L1 expressing cells as a prognostic biomarker for oropharyngeal squamous cell carcinoma. Br J Cancer. 2020;122(4):539-44. doi: 10.1038/s41416-019-0634-z. PubMed PMID: 31806878; PMCID: PMC7028988.

204. Matse JH, Bharos WK, Veerman ECI, Bloemena E, Bolscher JGM. Mucoepidermoid carcinoma-associated expression of MUC5AC, MUC5B and mucin-type carbohydrate antigen

sialyl-Tn in the parotid gland. Arch Oral Biol. 2017;82:121-6. Epub 2017/06/20. doi: 10.1016/j.archoralbio.2017.06.010. PubMed PMID: 28628803.

205. Lanzel EA, Pourian A, Sousa Melo SL, Brogden KA, Hellstein JW. Expression of Membrane-Bound Mucins and p63 in Distinguishing Mucoepidermoid Carcinoma from Papillary Cystadenoma. Head Neck Pathol. 2016;10(4):521-6. Epub 2016/10/28. doi: 10.1007/s12105-016-0735-4. PubMed PMID: 27278378; PMCID: PMC5082059.

206. Consortium EP. An integrated encyclopedia of DNA elements in the human genome. Nature. 2012;489(7414):57-74. doi: 10.1038/nature11247. PubMed PMID: 22955616; PMCID: PMC3439153.

207. Wouters MCA, Nelson BH. Prognostic Significance of Tumor-Infiltrating B Cells and Plasma Cells in Human Cancer. Clinical cancer research : an official journal of the American Association for Cancer Research. 2018;24(24):6125-35. Epub 2018/07/28. doi: 10.1158/1078-0432.CCR-18-1481. PubMed PMID: 30049748.

208. Kerche LE, de Sousa EA, Squarize CH, Oliveira KK, Marchi FA, Bettim BB, Kowalski LP, Soares FA, Lourenco SV, Coutinho-Camillo CM. EMT in salivary gland tumors: the expression of microRNAs miR-155 and miR-200c is associated with clinical-pathological parameters. Mol Biol Rep. 2022;49(3):2157-67. doi: 10.1007/s11033-021-07033-1. PubMed PMID: 34981333.

209. Shiratsuchi H, Nakashima T, Hirakawa N, Toh S, Nakagawa T, Saito T, Tsuneyoshi M, Komune S. beta-Catenin nuclear accumulation in head and neck mucoepidermoid carcinoma: its role in cyclin D1 overexpression and tumor progression. Head Neck. 2007;29(6):577-84. doi: 10.1002/hed.20583. PubMed PMID: 17315172.

210. Chandrashekar C, Angadi PV, Krishnapillai R. beta-Catenin expression in benign and malignant salivary gland tumors. Int J Surg Pathol. 2011;19(4):433-40. doi: 10.1177/1066896909346366. PubMed PMID: 20034989.

211. Schneider S, Thurnher D, Seemann R, Brunner M, Kadletz L, Ghanim B, Aumayr K, Heiduschka G, Lill C. The prognostic significance of beta-catenin, cyclin D1 and PIN1 in minor salivary gland carcinoma: beta-catenin predicts overall survival. Eur Arch Otorhinolaryngol. 2016;273(5):1283-92. doi: 10.1007/s00405-015-3609-6. PubMed PMID: 25801951.

212. Shieh YS, Hung YJ, Hsieh CB, Chen JS, Chou KC, Liu SY. Tumor-associated macrophage correlated with angiogenesis and progression of mucoepidermoid carcinoma of salivary glands. Ann Surg Oncol. 2009;16(3):751-60. doi: 10.1245/s10434-008-0259-6. PubMed PMID: 19116756.

213. Zhao SF, Yang XD, Lu MX, Sun GW, Wang YX, Zhang YK, Pu YM, Tang EY. Prognostic significance of VEGF immunohistochemical expression in oral cancer: a meta-analysis of the literature. Tumour Biol. 2013;34(5):3165-71. doi: 10.1007/s13277-013-0886-9. PubMed PMID: 23737289.

214. Schaaf MB, Garg AD, Agostinis P. Defining the role of the tumor vasculature in antitumor immunity and immunotherapy. Cell Death Dis. 2018;9(2):115. doi: 10.1038/s41419-017-0061-0. PubMed PMID: 29371595; PMCID: PMC5833710.

215. Bourhis M, Palle J, Galy-Fauroux I, Terme M. Direct and Indirect Modulation of T Cells by VEGF-A Counteracted by Anti-Angiogenic Treatment. Front Immunol. 2021;12:616837. doi: 10.3389/fimmu.2021.616837. PubMed PMID: 33854498; PMCID: PMC8039365.

216. McHugh JB, Visscher DW, Barnes EL. Update on selected salivary gland neoplasms. Arch Pathol Lab Med. 2009;133(11):1763-74. doi: 10.1043/1543-2165-133.11.176310.5858/133.11.1763. PubMed PMID: 19886710.

217. Goode RK, Auclair PL, Ellis GL. Mucoepidermoid carcinoma of the major salivary glands: clinical and histopathologic analysis of 234 cases with evaluation of grading criteria. Cancer. 1998;82(7):1217-24. doi: 10.1002/(sici)1097-0142(19980401)82:7<1217::aid-cncr2>3.0.co;2-c. PubMed PMID: 9529011.

218. Chen MM, Roman SA, Sosa JA, Judson BL. Histologic grade as prognostic indicator for mucoepidermoid carcinoma: a population-level analysis of 2400 patients. Head Neck. 2014;36(2):158-63. doi: 10.1002/hed.23256. PubMed PMID: 23765800.

219. Ali S, Sarhan M, Palmer FL, Whitcher M, Shah JP, Patel SG, Ganly I. Cause-specific mortality in patients with mucoepidermoid carcinoma of the major salivary glands. Ann Surg Oncol. 2013;20(7):2396-404. doi: 10.1245/s10434-013-2872-2. PubMed PMID: 23397152.

220. Haeggblom L, Ursu RG, Mirzaie L, Attoff T, Gahm C, Nordenvall LH, Nasman A. No evidence for human papillomavirus having a causal role in salivary gland tumors. Diagn Pathol. 2018;13(1):44. doi: 10.1186/s13000-018-0721-0. PubMed PMID: 30021645; PMCID: PMC6052678.

221. Sawabe M, Ito H, Takahara T, Oze I, Kawakita D, Yatabe Y, Hasegawa Y, Murakami S, Matsuo K. Heterogeneous impact of smoking on major salivary gland cancer according to histopathological subtype: A case-control study. Cancer. 2018;124(1):118-24. doi: 10.1002/cncr.30957. PubMed PMID: 28881386.

222. Saku T, Hayashi Y, Takahara O, Matsuura H, Tokunaga M, Tokunaga M, Tokuoka S, Soda M, Mabuchi K, Land CE. Salivary gland tumors among atomic bomb survivors, 1950-1987. Cancer. 1997;79(8):1465-75. PubMed PMID: 9118025.

223. Leopard D, El-Hitti E, Puttasiddaiah P, McLeod R, Owens D. Twenty-seven years of primary salivary gland carcinoma in Wales: an analysis of histological subtype and associated risk factors. J Laryngol Otol. 2022;136(2):167-72. doi: 10.1017/S002221512200007X. PubMed PMID: 35001862.

224. Sama S, Komiya T, Guddati AK. Advances in the Treatment of Mucoepidermoid Carcinoma. World J Oncol. 2022;13(1):1-7. doi: 10.14740/wjon1412. PubMed PMID: 35317327; PMCID: PMC8913015.

225. Stewart FW, Foote FW, Becker WF. Muco-Epidermoid Tumors of Salivary Glands. Ann Surg. 1945;122(5):820-44. doi: 10.1097/00000658-194511000-00005. PubMed PMID: 17858687; PMCID: PMC1618293.

226. Coca-Pelaz A, Rodrigo JP, Triantafyllou A, Hunt JL, Rinaldo A, Strojan P, Haigentz M, Jr., Mendenhall WM, Takes RP, Vander Poorten V, Ferlito A. Salivary mucoepidermoid carcinoma revisited. Eur Arch Otorhinolaryngol. 2015;272(4):799-819. doi: 10.1007/s00405-014-3053-z. PubMed PMID: 24771140.

227. Robinson L, van Heerden MB, Ker-Fox JG, Hunter KD, van Heerden WFP. Expression of Mucins in Salivary Gland Mucoepidermoid Carcinoma. Head Neck Pathol. 2021;15(2):491-502. doi: 10.1007/s12105-020-01226-z. PubMed PMID: 32959209; PMCID: PMC8134608.

228. Alos L, Lujan B, Castillo M, Nadal A, Carreras M, Caballero M, de Bolos C, Cardesa A. Expression of membrane-bound mucins (MUC1 and MUC4) and secreted mucins (MUC2, MUC5AC, MUC5B, MUC6 and MUC7) in mucoepidermoid carcinomas of salivary glands. Am J Surg Pathol. 2005;29(6):806-13. doi: 10.1097/01.pas.0000155856.84553.c9. PubMed PMID: 15897748.

229. Arruda CFJ, Coutinho-Camillo CM, Marques MM, Nagano CP, Bologna SB, Bettim BB, Germano JN, Pinto CAL, Hsieh R, Lourenco SV. Claudin expression is maintained in mucoepidermoid carcinoma of the salivary gland. Pathol Res Pract. 2020;216(11):153161. doi: 10.1016/j.prp.2020.153161. PubMed PMID: 32862070.

230. Adams A, Warner K, Nor JE. Salivary gland cancer stem cells. Oral Oncol. 2013;49(9):845-53. doi: 10.1016/j.oraloncology.2013.05.013. PubMed PMID: 23810400; PMCID: PMC3744607.

231. Adams A, Warner K, Pearson AT, Zhang Z, Kim HS, Mochizuki D, Basura G, Helman J, Mantesso A, Castilho RM, Wicha MS, Nor JE. ALDH/CD44 identifies uniquely tumorigenic cancer stem cells in salivary gland mucoepidermoid carcinomas. Oncotarget. 2015;6(29):26633-50. doi: 10.18632/oncotarget.5782. PubMed PMID: 26449187; PMCID: PMC4694941.

232. Binmadi N, Elsissi A, Elsissi N. Expression of cell adhesion molecule CD44 in mucoepidermoid carcinoma and its association with the tumor behavior. Head Face Med. 2016;12:8. doi: 10.1186/s13005-016-0102-4. PubMed PMID: 26821610; PMCID: PMC4731993.

233. Irani S, Jafari B. Expression of vimentin and CD44 in mucoepidermoid carcinoma: A role in tumor growth. Indian J Dent Res. 2018;29(3):333-40. doi: 10.4103/ijdr.IJDR_184_17. PubMed PMID: 29900918.

234. da Silva LP, Lopes M, Sarmento ASC, de Albuquerque Borges M, de Moura SRS, Sobral APV, de Souza LB. Increased expression of ALDH-1 is associated with clinical parameters of salivary glands neoplasms. Exp Mol Pathol. 2020;117:104552. doi: 10.1016/j.yexmp.2020.104552. PubMed PMID: 33053323.

235. Andrade NP, Warner KA, Zhang Z, Pearson AT, Mantesso A, Guimaraes DM, Altemani A, Mariano FV, Nunes FD, Nor JE. Survival of salivary gland cancer stem cells requires mTOR signaling. Cell Death Dis. 2021;12(1):108. doi: 10.1038/s41419-021-03391-7. PubMed PMID: 33479203; PMCID: PMC7820616.

236. Guimaraes DM, Almeida LO, Martins MD, Warner KA, Silva AR, Vargas PA, Nunes FD, Squarize CH, Nor JE, Castilho RM. Sensitizing mucoepidermoid carcinomas to chemotherapy by targeted disruption of cancer stem cells. Oncotarget. 2016;7(27):42447-60. doi: 10.18632/oncotarget.9884. PubMed PMID: 27285758; PMCID: PMC5173147.

237. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell. 2011;144(5):646-74. doi: 10.1016/j.cell.2011.02.013. PubMed PMID: 21376230.

238. Dahlenfors R, Lindahl L, Mark J. A fourth minor salivary gland mucoepidermoid carcinoma with 11q14-21 and 19p12 rearrangements. Hereditas. 1994;120(3):287-8. doi: 10.1111/j.1601-5223.1994.00287.x. PubMed PMID: 7928392.

239. Saade RE, Bell D, Garcia J, Roberts D, Weber R. Role of CRTC1/MAML2 Translocation in the Prognosis and Clinical Outcomes of Mucoepidermoid Carcinoma. JAMA Otolaryngol Head Neck Surg. 2016;142(3):234-40. doi: 10.1001/jamaoto.2015.3270. PubMed PMID: 26796488.

240. Luk PP, Wykes J, Selinger CI, Ekmejian R, Tay J, Gao K, Eviston TJ, Lum T, O'Toole SA, Clark JR, Gupta R. Diagnostic and prognostic utility of Mastermind-like 2 (MAML2) gene rearrangement detection by fluorescent in situ hybridization (FISH) in mucoepidermoid carcinoma of the salivary glands. Oral Surg Oral Med Oral Pathol Oral Radiol. 2016;121(5):530-41. doi: 10.1016/j.oooo.2016.01.003. PubMed PMID: 27068311.

241. Seethala RR, Dacic S, Cieply K, Kelly LM, Nikiforova MN. A reappraisal of the MECT1/MAML2 translocation in salivary mucoepidermoid carcinomas. Am J Surg Pathol. 2010;34(8):1106-21. doi: 10.1097/PAS.0b013e3181de3021. PubMed PMID: 20588178.

242. Okabe M, Miyabe S, Nagatsuka H, Terada A, Hanai N, Yokoi M, Shimozato K, Eimoto T, Nakamura S, Nagai N, Hasegawa Y, Inagaki H. MECT1-MAML2 fusion transcript defines a favorable subset of mucoepidermoid carcinoma. Clin Cancer Res. 2006;12(13):3902-7. doi: 10.1158/1078-0432.CCR-05-2376. PubMed PMID: 16818685.

243. Chen Z, Ni W, Li JL, Lin S, Zhou X, Sun Y, Li JW, Leon ME, Hurtado MD, Zolotukhin S, Liu C, Lu J, Griffin JD, Kaye FJ, Wu L. The CRTC1-MAML2 fusion is the major oncogenic driver in mucoepidermoid carcinoma. JCI Insight. 2021;6(7). doi: 10.1172/jci.insight.139497. PubMed PMID: 33830080; PMCID: PMC8119194.

244. Mootha VK, Lindgren CM, Eriksson KF, Subramanian A, Sihag S, Lehar J, Puigserver P, Carlsson E, Ridderstrale M, Laurila E, Houstis N, Daly MJ, Patterson N, Mesirov JP, Golub TR, Tamayo P, Spiegelman B, Lander ES, Hirschhorn JN, Altshuler D, Groop LC. PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. Nat Genet. 2003;34(3):267-73. doi: 10.1038/ng1180. PubMed PMID: 12808457.

245. Fehr A, Roser K, Heidorn K, Hallas C, Loning T, Bullerdiek J. A new type of MAML2 fusion in mucoepidermoid carcinoma. Genes Chromosomes Cancer. 2008;47(3):203-6. doi: 10.1002/gcc.20522. PubMed PMID: 18050304.

246. Zhang L, Mitani Y, Caulin C, Rao PH, Kies MS, Saintigny P, Zhang N, Weber RS, Lippman SM, El-Naggar AK. Detailed genome-wide SNP analysis of major salivary carcinomas localizes subtype-specific chromosome sites and oncogenes of potential clinical significance. Am J Pathol. 2013;182(6):2048-57. doi: 10.1016/j.ajpath.2013.02.020. PubMed PMID: 23583282; PMCID: PMC3668022.

247. Jee KJ, Persson M, Heikinheimo K, Passador-Santos F, Aro K, Knuutila S, Odell EW, Makitie A, Sundelin K, Stenman G, Leivo I. Genomic profiles and CRTC1-MAML2 fusion distinguish different subtypes of mucoepidermoid carcinoma. Mod Pathol. 2013;26(2):213-22. doi: 10.1038/modpathol.2012.154. PubMed PMID: 23018873.

248. Matse JH, Veerman ECI, Bolscher JGM, Leemans CR, Ylstra B, Bloemena E. High number of chromosomal copy number aberrations inversely relates to t(11;19)(q21;p13) translocation status in mucoepidermoid carcinoma of the salivary glands. Oncotarget. 2017;8(41):69456-64. doi: 10.18632/oncotarget.17282. PubMed PMID: 29050216; PMCID: PMC5642491.

249. Anzick SL, Chen WD, Park Y, Meltzer P, Bell D, El-Naggar AK, Kaye FJ. Unfavorable prognosis of CRTC1-MAML2 positive mucoepidermoid tumors with CDKN2A deletions. Genes Chromosomes Cancer. 2010;49(1):59-69. doi: 10.1002/gcc.20719. PubMed PMID: 19827123; PMCID: PMC2783528.

250. Suh JH, Johnson A, Albacker L, Wang K, Chmielecki J, Frampton G, Gay L, Elvin JA, Vergilio JA, Ali S, Miller VA, Stephens PJ, Ross JS. Comprehensive Genomic Profiling Facilitates Implementation of the National Comprehensive Cancer Network Guidelines for Lung Cancer Biomarker Testing and Identifies Patients Who May Benefit From Enrollment in Mechanism-Driven Clinical Trials. Oncologist. 2016;21(6):684-91. doi: 10.1634/theoncologist.2016-0030. PubMed PMID: 27151654; PMCID: PMC4912374.

251. Ross JS, Gay LM, Wang K, Vergilio JA, Suh J, Ramkissoon S, Somerset H, Johnson JM, Russell J, Ali S, Schrock AB, Fabrizio D, Frampton G, Miller V, Stephens PJ, Elvin JA, Bowles DW. Comprehensive genomic profiles of metastatic and relapsed salivary gland carcinomas are associated with tumor type and reveal new routes to targeted therapies. Ann Oncol. 2017;28(10):2539-46. doi: 10.1093/annonc/mdx399. PubMed PMID: 28961851; PMCID: PMC5834110.

252. Wang K, McDermott JD, Schrock AB, Elvin JA, Gay L, Karam SD, Raben D, Somerset H, Ali SM, Ross JS, Bowles DW. Comprehensive genomic profiling of salivary mucoepidermoid carcinomas reveals frequent BAP1, PIK3CA, and other actionable genomic alterations. Ann Oncol. 2017;28(4):748-53. doi: 10.1093/annonc/mdw689. PubMed PMID: 28327999.

253. Kato S, Elkin SK, Schwaederle M, Tomson BN, Helsten T, Carter JL, Kurzrock R. Genomic landscape of salivary gland tumors. Oncotarget. 2015;6(28):25631-45. doi: 10.18632/oncotarget.4554. PubMed PMID: 26247885; PMCID: PMC4694855.

254. Grunewald I, Vollbrecht C, Meinrath J, Meyer MF, Heukamp LC, Drebber U, Quaas A, Beutner D, Huttenbrink KB, Wardelmann E, Hartmann W, Buttner R, Odenthal M, Stenner M. Targeted next generation sequencing of parotid gland cancer uncovers genetic heterogeneity. Oncotarget. 2015;6(20):18224-37. doi: 10.18632/oncotarget.4015. PubMed PMID: 26053092; PMCID: PMC4627247.

255. Nakano T, Yamamoto H, Hashimoto K, Tamiya S, Shiratsuchi H, Nakashima T, Nishiyama K, Higaki Y, Komune S, Oda Y. HER2 and EGFR gene copy number alterations are predominant in high-grade salivary mucoepidermoid carcinoma irrespective of MAML2 fusion status. Histopathology. 2013;63(3):378-92. doi: 10.1111/his.12183. PubMed PMID: 23855785.

256. Cros J, Sbidian E, Hans S, Roussel H, Scotte F, Tartour E, Brasnu D, Laurent-Puig P, Bruneval P, Blons H, Badoual C. Expression and mutational status of treatment-relevant targets and key oncogenes in 123 malignant salivary gland tumours. Ann Oncol. 2013;24(10):2624-9. doi: 10.1093/annonc/mdt338. PubMed PMID: 23933559.

257. Karpinets TV, Mitani Y, Liu B, Zhang J, Pytynia KB, Sellen LD, Karagiannis DT, Ferrarotto R, Futreal AP, El-Naggar AK. Whole-Genome Sequencing of Common Salivary Gland Carcinomas: Subtype-Restricted and Shared Genetic Alterations. Clin Cancer Res. 2021;27(14):3960-9. doi: 10.1158/1078-0432.CCR-20-4071. PubMed PMID: 34011559; PMCID: PMC8598082.

258. Pendleton M, Sebra R, Pang AW, Ummat A, Franzen O, Rausch T, Stutz AM, Stedman W, Anantharaman T, Hastie A, Dai H, Fritz MH, Cao H, Cohain A, Deikus G, Durrett RE, Blanchard SC, Altman R, Chin CS, Guo Y, Paxinos EE, Korbel JO, Darnell RB, McCombie WR, Kwok PY, Mason CE, Schadt EE, Bashir A. Assembly and diploid architecture of an individual human genome via single-molecule technologies. Nat Methods. 2015;12(8):780-6. doi: 10.1038/nmeth.3454. PubMed PMID: 26121404; PMCID: PMC4646949.

259. Yin LX, Ha PK. Genetic alterations in salivary gland cancers. Cancer. 2016;122(12):1822-31. doi: 10.1002/cncr.29890. PubMed PMID: 26928905; PMCID: PMC5783696.